MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE REPUBLIC OF KAZAKHSTAN

ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ ҒЫЛЫМ ЖӘНЕ ЖОҒАРЫ БІЛІМ МИНИСТРЛІГІ

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РЕСПУБЛИКИ КАЗАХСТАН

KAZAKHSTAN

ХАЛЫҚАРАЛЫҚ УНИВЕРСИТЕТІ

INTERNATIONAL UNIVERSITY SINCE 2009

МЕЖДУНАРОДНЫЙ УНИВЕРСИТЕТ

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

Published since 2020.
Volume 7. 3 (27). 2026
July–September

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

2020 жылдан бері шығарылады
Том 7. 3 (27). 2026
Шілде-Қыркүйек

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

Издается с 2020 г.
Том 7. 3 (27). 2026
Июль-Сентябрь

Свидетельство о постановке на учет периодического печатного издания в Министерство информации и общественного развития Республики Казахстан № KZ82VPY00020475, выданное от 20.02.2020 г.

Зарегистрировано в Международном центре регистрации серийных изданий ISSN (ЮНЕСКО, Париж, Франция). ISSN 2708–2032 (print), ISSN 2708–2040 (online)

Журнал входит в Перечень научных изданий, рекомендуемых КОКНВО МНВО РК для публикации основных результатов научной деятельности.

EDITOR-IN-CHIEF:
Kateryna Kolesnikova — Doctor of Technical Sciences, professor, Vice-Rector for Research, International Information Technology University (Kazakhstan)

DEPUTY EDITOR-IN-CHIEF:
Madina Ipalakova — Candidate of Technical Sciences, associate professor, Director of the Research Department, International Information Technology University (Kazakhstan)

EDITORIAL BOARD:
Abdul Razak — PhD, professor, Department of Cybersecurity, International Information Technology University (Kazakhstan)
Lucio Tommaso De Paolis — Director of the R&D Department of the AVR Laboratory, Department of Engineering for Innovation, University of Salento (Italy)
Liz Bacon — Professor, Deputy Vice-Chancellor, Abertay University (United Kingdom)
Michele Pagano — PhD, Professor, University of Pisa (Italy)
Mukhtarbay Otelbayev — Doctor of Physical and Mathematical Sciences, professor, academician of the National Academy of Sciences of the Republic of Kazakhstan, professor of the Department of Mathematical and Computer Modeling, International Information Technology University (Kazakhstan)
Bolatbek Rysbaiuly — Doctor of Physical and Mathematical Sciences, professor, professor of the Department of Computing and Data Science, Astana IT University (Kazakhstan)
Yevgeniya Daineko — PhD, research professor, Department of Information Systems, International Information Technology University (Kazakhstan)
Nurzhan Duzbayev — PhD, associate professor, Vice-Rector for Digitalization and Innovation, International Information Technology University (Kazakhstan)
Bakhtgerei Sinchev — Doctor of Technical Sciences, professor, Department of Information Systems, International Information Technology University (Kazakhstan)
Nurgul Seilova — Candidate of Technical Sciences, Dean of the Faculty of Computer Technologies and Cybersecurity, International Information Technology University (Kazakhstan)
Ardak Mukhamediyeva — Candidate of Economic Sciences, Dean of the Faculty of Business, Media and Management, International Information Technology University (Kazakhstan)
Zamira Abdikalikova — PhD, associate professor, Head of the Department of Mathematical and Computer Modeling, International Information Technology University (Kazakhstan)
Yerlan Shildibekov — PhD, associate professor, Head of the Department of Economics and Business, International Information Technology University (Kazakhstan)
Damilya Yeskendirova — Candidate of Technical Sciences, associate professor, Head of the Department of Cybersecurity, International Information Technology University (Kazakhstan)
Aigul Niyazgulova — Candidate of Philological Sciences, Professor, Head of the Department of Media Communications and History of Kazakhstan, International Information Technology University (Kazakhstan)
Altai Aitmagambetov — Candidate of Technical Sciences, Professor, Department of Radio Engineering, Electronics and Telecommunications, International Information Technology University (Kazakhstan)
Yelena Bakhtiyarova — Candidate of Technical Sciences, associate professor, Head of the Department of Radio Engineering, Electronics and Telecommunications, International Information Technology University (Kazakhstan)
Kanibek Sansyzbay — PhD, research professor, Department of Cybersecurity, International Information Technology University (Kazakhstan)
Sakhybay Tynymbayev — Candidate of Technical Sciences, Professor, Research Professor, Department of Computer Engineering, International Information Technology University (Kazakhstan)
Ali Abd Almisreb — PhD, associate professor, Department of Cybersecurity, International Information Technology University (Kazakhstan)
Mohamed Ahmed Hamada — PhD, associate professor, Department of Information Systems, International Information Technology University (Kazakhstan)
Yang Im Chu — PhD, Professor, Gachon University (South Korea)
Tadeusz Wallas — PhD, Vice-Rector, Adam Mickiewicz University (Poland)
Orken Mamyrbayev — PhD, Deputy Director for Science, RSE Institute of Information and Computational Technologies, Committee for Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Kazakhstan)
Sergey Bushuyev — Doctor of Technical Sciences, professor, Director of the Ukrainian Project Management Association “UKRNET,” Head of the Department of Project Management, Kyiv National University of Construction and Architecture (Ukraine)
Svetlana Beloshitskaya — Doctor of Technical Sciences, professor, Department of Computing and Data Science, Astana IT University (Kazakhstan)

MANAGING EDITOR
Raushan Mrzabayeva — Master of Science, editor, International Information Technology University (Kazakhstan)

International Journal of Information and Communication Technologies
Periodicity: 4 times a year.
Languages: Kazakh, Russian, English
DOI prefix: 10.54309
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Thematic focus: "Information technology"; "Digital technologies in the development of socio-economic systems"; "Information security and communication technologies".
Distribution: Materials are distributed under the Creative Commons Attribution 4.0
Journal website: https://journal.iitu.edu.kz
Owner: International Information Technology University JSC (Almaty).
Copyright: © International Journal of Information and Communication Technologies, 2026

РЕДАКЦИЯ

БАС РЕДАКТОР:

Колесникова Катерина Викторовна — техника ғылымдарының докторы, профессор, Халықаралық ақпараттық технологиялар университетінің ғылыми-зерттеу қызметі жөніндегі проректор (Қазақстан)

БАС РЕДАКТОРДЫҢ ОРЫНБАСАРЫ:

Ипалакова Мадина Тулегеновна — техника ғылымдарының кандидаты, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университетінің ғылыми-зерттеу қызметі жөніндегі департамент директоры (Қазақстан)

РЕДАКЦИЯЛЫҚ АЛҚА:

Разак Абдул — PhD, Халықаралық ақпараттық технологиялар университеті киберқауіпсіздік кафедрасының профессоры (Қазақстан)

Лучио Томмазо де Паолис — Саленто Университеті (Италия) инновация және технологиялық инжиниринг департаменті AVR зертханасының зерттеу және әзірлеу бөлімінің директоры

Лиз Бэкон — профессор, Абертей Университеті (Ұлыбритания) вице-канцлерінің орынбасары

Микеле Пагано — PhD, Пиза Университетінің (Италия) профессоры

Өтелбаев Мухтарбай Өтелбайұлы — физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, Халықаралық ақпараттық технологиялар университеті математика және компьютерлік модельдеу кафедрасының профессоры (Қазақстан)

Рысбайулы Болатбек — физика-математика ғылымдарының докторы, профессор, Есептеу және деректер ғылымдары департаментінің профессоры, Astana IT University (Казахстан)

Дайнеко Евгения Александровна — PhD, Халықаралық ақпараттық технологиялар университеті ақпараттық жүйелер кафедрасының профессор-зерттеушісі (Қазақстан)

Дузбаев Нуржан Токкужаевич — PhD, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті цифрландыру және инновациялар жөніндегі проректор (Қазақстан)

Синчев Бахтгерей Куспанович — техника ғылымдарының докторы, профессор, Халықаралық ақпараттық технологиялар университеті ақпараттық жүйелер кафедрасының профессоры (Қазақстан)

Сейлова Нургуль Абадуллаевна — техника ғылымдарының докторы, Халықаралық ақпараттық технологиялар университеті компьютерлік технологиялар және киберқауіпсіздік факультетінің деканы (Қазақстан)

Мухамедиева Ардак Габитовна — экономика ғылымдарының кандидаты, Халықаралық ақпараттық технологиялар университеті бизнеса медиа және басқару факультетінің деканы (Қазақстан)

Абдикаликова Замира Турсынбаевна — PhD, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті математика және компьютерлік модельдеу кафедрасының меңгерушісі (Қазақстан)

Шильдибеков Ерлан Жаржанович — PhD, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті экономика және бизнес кафедрасының меңгерушісі (Қазақстан)

Дамеля Максутовна Ескендирова — техника ғылымдарының кандидаты, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті киберқауіпсіздік кафедрасының меңгерушісі (Қазақстан)

Ниязгулова Айгуль Аскарбековна — филология ғылымдарының кандидаты, доцент, профессор, Халықаралық ақпараттық технологиялар университеті медиакоммуникация және Қазақстан тарихы кафедрасының меңгерушісі (Қазақстан)

Айтмагамбетов Алтай Зуфарович — техника ғылымдарының кандидаты, Халықаралық ақпараттық технологиялар университеті радиотехника, электроника және телекоммуникация кафедрасының профессоры (Қазақстан)

Бахтиярова Елена Ажибековна — техника ғылымдарының кандидаты, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті радиотехника, электроника және телекоммуникация кафедрасының меңгерушісі (Қазақстан)

Канибек Сансызбай — PhD, қауымдастырылған профессор, Халықаралық ақпараттық технологиялар университеті киберқауіпсіздік кафедрасының профессор-зерттеушісі (Қазақстан)

Тынымбаев Сахибай — техника ғылымдарының кандидаты, профессор, Халықаралық ақпараттық технологиялар университеті компьютерлік инженерия кафедрасының профессор-зерттеушісі (Қазақстан)

Алмисреб Али Абд — PhD, Халықаралық ақпараттық технологиялар университеті киберқауіпсіздік кафедрасының қауымдастырылған профессоры (Қазақстан)

Мохамед Ахмед Хамада — PhD, Халықаралық ақпараттық технологиялар университеті ақпараттық жүйелер кафедрасының қауымдастырылған профессоры (Қазақстан)

Янг Им Чу — PhD, Гачон университетінің профессоры (Оңтүстік Корея)

Тадеуш Валлас – PhD, Адам Мицкевич атындағы (Польша) университеттің проректоры

Мамырбаев Оркен Жумажанович — PhD, ҚР ҒЖБМ Ғылым комитеті ақпараттық және есептеу технологиялары институты ӨМК директорының ғылым жөніндегі орынбасары (Қазақстан)

Бушуев Сергей Дмитриевич — техника ғылымдарының докторы, профессор, Украинаның "УКРНЕТ" жобаларды басқару қауымдастығының директоры, Киев ұлттық құрылыс және сәулет университеті жобаларды басқару кафедрасының меңгерушісі (Украина)

Белощицкая Светлана Васильевна — техника ғылымдарының докторы, доцент, Astana IT University есептеу және деректер ғылымы кафедрасының профессоры (Қазақстан)

ЖАУАПТЫ РЕДАКТОР:

Мрзабаева Раушан Жалиевна — магистр, Халықаралық ақпараттық технологиялар университетінің редакторы (Қазақстан)

Халықаралық ақпараттық және коммуникациялық технологиялар журналы
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Префикс DOI: 10.54309
Мерзімділігі: жылына 4 рет.
Басылым тілі: қазақ, орыс, ағылшын.
Тақырып бағыты: "Ақпараттық технологиялар"; "Ақпараттық қауіпсіздік және коммуникациялық технологиялар"; "Әлеуметтік-экономикалық жүйелерді дамытудағы цифрлық технология".
Журнал сайты: https://journal.iitu.edu.kz
Тарату: материалдар Creative Commons Attribution 4.0 лицензиясы бойынша таратылады
Меншік иесі: АҚ «Халықаралық ақпараттық технологиялар университеті» (Алматы қ.).
Авторлық құқық: © Халықаралық ақпараттық және коммуникациялық технологиялар журналы, 2026

РЕДАКЦИЯ

ГЛАВНЫЙ РЕДАКТОР:

Колесникова Катерина Викторовна — доктор технических наук, профессор, проректор по научно-исследовательской деятельности Международного университета информационных технологий (Казахстан)

ЗАМЕСТИТЕЛЬ ГЛАВНОГО РЕДАКТОРА:

Ипалакова Мадина Тулегеновна — кандидат технических наук, ассоциированный профессор, директор департамента по научно-исследовательской деятельности Международного университета информационных технологий (Казахстан)

РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

Разак Абдул — PhD, профессор кафедры кибербезопасности Международного университета информационных технологий (Казахстан)

Лучио Томмазо де Паолис — директор отдела исследований и разработок лаборатории AVR департамента инноваций и технологического инжиниринга Университета Саленто (Италия)

Лиз Бэкон — профессор, заместитель вице-канцлера Университета Абертей (Великобритания)

Микеле Пагано — PhD, профессор Университета Пизы (Италия)

Отелбаев Мухтарбай Отелбайулы — доктор физико-математических наук, профессор, академик НАН РК, профессор кафедры математического и компьютерного моделирования Международного университета информационных технологий (Казахстан)

Рысбайулы Болатбек — доктор физико-математических наук, профессор, профессор Astana IT University (Казахстан)

Дайнеко Евгения Александровна — PhD, профессор-исследователь кафедры информационных систем Международного университета информационных технологий (Казахстан)

Дузбаев Нуржан Токкужаевич — PhD, ассоциированный профессор, проректор по цифровизации и инновациям Международного университета информационных технологий (Казахстан)

Синчев Бахтгерей Куспанович — доктор технических наук, профессор, профессор кафедры информационных систем Международного университета информационных технологий (Казахстан)

Сейлова Нургуль Абадуллаевна — кандидат технических наук, декан факультета компьютерных технологий и кибербезопасности Международного университета информационных технологий (Казахстан)

Мухамедиева Ардак Габитовна — кандидат экономических наук, декан факультета бизнеса медиа и управления Международного университета информационных технологий (Казахстан)

Абдикаликова Замира Турсынбаевна — PhD, ассоциированный профессор, заведующая кафедрой математического и компьютерного моделирования Международного университета информационных технологий (Казахстан)

Шильдибеков Ерлан Жаржанович — PhD, ассоциированный профессор, заведующий кафедрой экономики и бизнеса Международного университета информационных технологий (Казахстан)

Дамеля Максутовна Ескендирова — кандидат технических наук, ассоциированный профессор, заведующая кафедрой кибербезопасности Международного университета информационных технологий (Казахстан)

Ниязгулова Айгуль Аскарбековна — кандидат филологических наук, доцент, профессор, заведующая кафедрой медиакоммуникаций и истории Казахстана Международного университета информационных технологий (Казахстан)

Айтмагамбетов Алтай Зуфарович — кандидат технических наук, профессор кафедры радиотехники, электроники и телекоммуникаций Международного университета информационных технологий (Казахстан)

Бахтиярова Елена Ажибековна — кандидат технических наук, ассоциированный профессор, заведующая кафедрой радиотехники, электроники и телекоммуникаций Международного университета информационных технологий (Казахстан)

Канибек Сансызбай – PhD, ассоциированный профессор, профессор-исследователь кафедры кибербезопасности, Международного университета информационных технологий (Казахстан)

Тынымбаев Сахибай – кандидат технических наук, профессор, профессор-исследователь кафедры компьютерной инженерии, Международного университета информационных технологий (Казахстан)

Алмисреб Али Абд — PhD, ассоциированный профессор кафедры кибербезопасности Международного университета информационных технологий (Казахстан)

Мохамед Ахмед Хамада — PhD, ассоциированный профессор кафедры информационных систем Международного университета информационных технологий (Казахстан)

Янг Им Чу — PhD, профессор университета Гачон (Южная Корея)

Тадеуш Валлас – PhD, проректор университета имен Адама Мицкевича (Польша)

Мамырбаев Оркен Жумажанович — PhD, заместитель директора по науке РГП Института информационных и вычислительных технологий Комитета науки МНВО РК (Казахстан)

Бушуев Сергей Дмитриевич — доктор технических наук, профессор, директор Украинской ассоциации управления проектами «УКРНЕТ», заведующий кафедрой управления проектами Киевского национального университета строительства и архитектуры (Украина)

Белощицкая Светлана Васильевна — доктор технических наук, доцент, профессор кафедры вычислений и науки о данных Astana IT University (Казахстан)

ОТВЕТСТВЕННЫЙ РЕДАКТОР:

Мрзабаева Раушан Жалиевна — магистр, редактор Международного университета информационных технологий (Казахстан)

Международный журнал информационных и коммуникационных технологий
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Префикс DOI: 10.54309
Периодичность: 4 выпусков в год.
Язык издания: казахский, русский, английский.
Тематическая направленность: "Информационные технологии"; "Информационная безопасность и коммуникационные технологии"; "Цифровые технологии в развитии социо-экономических систем".
Сайт журнала: https://journal.iitu.edu.kz
Распространение: материалы распространяются по лицензии Creative Commons Attribution 4.0
Собственник: АО «Международный университет информационных технологий» (г. Алматы).
Авторские права: © Международный журнал информационных и коммуникационных технологий, 2026

CONTENTS

DIGITAL TECHNOLOGIES IN THE DEVELOPMENT OF SOCIO-ECONOMIC SYSTEMS

A Adalbek, S.Ye. Kerimkhulle, K.D. Kuliev, O. Mekesh, Zh.Zh. Khamitova
DEVELOPMENT OF RESULT-ORIENTED DATA SMOOTHING MODELS AND ANALYSIS OF THEIR EFFECTIVENESS 11
A.T. Iskakova, G.E. Kurmankulova, K. Gromaszek, K.O. Togzhanova, R.S. Algozhayeva
APPLICATION OF MACHINE LEARNING AND DEEP LEARNING FOR PREDICTIVE PLANNING AND RESOURCE EFFICIENCY IN PRODUCTION SYSTEMS 30
A.B. Kassymova, R.K. Uskenbayeva, Zh.A. Sydykov, D. Omirzak, S. Kalibekov
ARCHITECTURE OF AN INTELLIGENT CHATBOT FOR SMART CITY: AI-BASED RECEIVING, ANALYSIS AND ROUTING OF CITIZEN REQUESTS 46
O.M. Olzhayev, B.S. Omarov, A. Suliman
YOLO-CBAM: A LIGHTWEIGHT NEURAL NETWORK MODEL WITH INTEGRATED ATTENTION MECHANISM FOR ROAD DAMAGE ANALYSIS 66
K. Yermagambetova, A. Sarsekeyeva, Y. Akhlassov
OVERCOMING FRAGMENTATION IN THE CLASSIFICATION OF DIGITAL CULTURAL HERITAGE: AN INTEGRATIVE MULTIDIMENSIONAL MODEL 78

INFORMATION TECHNOLOGY

D. Abzhanova, S. Sharipova
ARCHITECTURE AND PRACTICAL IMPLEMENTATION OF A DIGITAL PLATFORM FOR MONITORING AND DATA EXCHANGE ON TRANSBOUNDARY WATER RESOURCES IN THE ZHAMBYL REGION 101
A. Aitanuly, O.A. Ussatova
ENHANCING TWO-FACTOR AUTHENTICATION FOR CLOUD HEALTHCARE SECURITY: THE ROLE OF DEEP LEARNING 122
O. Alseitov, O. Kuchanskyi
PHYSICS-INFORMED HYBRID LEARNING FOR HOURLY POWER-PLANT EMISSIONS: A MULTI-POLLUTANT BENCHMARK WITH UNCERTAINTY AND TRANSFER 141
A.M. Akhmetova, D.O. Baizhanova, Zh.N. Alisheva, B.A. Shaimov
ADAPTIVE OUTPUT ROUTER FOR NATURAL LANGUAGE MODELS CONSIDERING KEY PERFORMANCE INDICATORS 160
A.K. Aitim
AUTOMATIC AND SEMI-AUTOMATIC ANNOTATION OF KAZAKH TEXT CORPORA 180
Z.E. Balabayev, S.V. Biloshchytska
HANDWRITTEN TEXT RECOGNITION FOR CHAGATAI MANUSCRIPTS: COMPARING CRNN TRAINING FROM SCRATCH AND TRANSFORMER-BASED TRANSFER LEARNING 198
D.B. Baumuratova, T.K. Zhukabayeva, K.A. Abdrakhmanov, Z. Ahmad, A.A. Seitov
A NEURO-DIGITAL LEARNING MODEL FOR AI DRIVEN DIGITAL TRANSFORMATION IN HIGHER EDUCATION 219
T. Daoyun, G.A. Amirkhanova, Fu Yanwei, Y. Faruzkyzy
MULTIMODAL SEMANTIC MANIPULATION IN FAKE VIDEOS AND ITS INFLUENCE MECHANISM ON USERS' PERCEIVED TRUST 241
Zh. Doumchariyev, J. Tussupov, A. Akynbekova, M. Khassanova, M. Kassim
ENSEMBLE AND HYBRID LEARNING FOR DIGITAL TWIN MILK CLASSIFICATION 258
K.Z. Igibayev, A.Kh. Kozbakova, O. Auelbekov , T. Merembayev, K. Yessentayev
STUDY OF A DYNAMIC MIXING DEVICE IN A BIOREACTOR USING IOT TECHNOLOGIES 279
A.N. Ismukanova, G. Kochshanova, D.V. Pleskachev, Zh. Sarsenbayeva, A.A. Daurenova
AUTOMATED CLASSIFICATION OF SCIENTIFIC DATA USING ARTIFICIAL INTELLIGENCE: A COMPARATIVE ANALYSIS OF TRANSFORMER ARCHITECTURES AND HYBRID APPROACHES 306
T. Zh. Kamalov, Ali Abd. Almisreb, Mohammed A. Saleh
INVESTIGATION OF HYBRID QUANTUM-CLASSICAL ALGORITHMS FOR MACHINE LEARNING 323
B. Karimkyzy, K.G. Karasenko, N.A. Alisherov, N.M. Muratbek, A. Karim

KEY PUBLICATIONS OF 2021–2025 ON ARTIFICIAL INTELLIGENCE METHODS IN THE MANUFACTURING AND DIAGNOSTICS OF MEDICAL DEVICES FOR GLYCEMIC MONITORING: A THEMATIC ANALYTICAL REVIEW 338
S.K. Kumyzbaeva, K.E. Baglanbaev, K.S. Sarkyt, M.Zh. Konyrova
THREE-PHASE TRAINING OF INTELLIGENT ENERGY MANAGEMENT FOR A HYBRID ENERGY SYSTEM WITH A CLIMATE-AWARE STATE REPRESENTATION 357
N.E. Karabayev, S.K. Serikbayeva, E. Benkhelif, B. Tassuov
RESEARCH ON THE ROBUSTNESS OF ENSEMBLE MACHINE LEARNING MODELS TO ADVERSARIAL ATTACKS IN INTRUSION DETECTION SYSTEMS FOR IOT NETWORKS 369
M. Nurtas, G. Kapyshev
A METHODOLOGICAL FRAMEWORK FOR THE PREPROCESSING AND INTEGRATION OF GEOSPATIAL AND GEOPHYSICAL DATA FOR MINERAL PROSPECTIVITY PREDICTION USING MACHINE LEARNING 392
A. Neftissov, A. Biloshchytskyi, I. Kazambayev, A. Medetkhan, L. Kirichenko
ENERGY EFFICIENT MICROCLIMATE STABILIZATION FOR SEALED ARCHIVAL STORAGE ENCLOSURES 405
A.Zh. Omurtayeva, U.T. Makhazhanova, A.K. Adilova, Sh.K. Yelezhanova, J. Kultan
AN ADAPTIVE NEURAL-FUZZY MODEL FOR ASSESSING INVESTMENT RISKS IN THE AGRI-INDUSTRIAL COMPLEX: ODERNISATION OF THE DECISION-SUPPORT ARCHITECTURE THROUGH THE INTEGRATION OF MACHINE LEARNING 421
N. Shazhdekeyeva, A. Mukhanbetkaliyeva, A. Abdigaliyeva, A. Zhanekeshova, L. Nurgaliyeva
COMPARATIVE ANALYSIS OF MATHEMATICAL MODELS OF COLLABORATIVE FILTERING IN THE DEVELOPMENT OF A MOVIE RECOMMENDER SYSTEM 445
S. Sharmukhanbet, G. Turmukhanova, O. Findik, V. Makhatova, M. Rakhmetov
AN INTELLIGENT PIPELINE DEFECT DETECTION SYSTEM USING AN IN-PIPE ROBOT 461

INFORMATION SECURITY AND COMMUNICATIONTECHNOLOGIES

S.Z. Daruish, A.U. Utegenova, A.U. Kalizhanova, Zh.Zh. Zhumabekova
MACHINE LEARNING FOR POLLUTION CLASSIFICATION IN ROCKET BOOSTER IMPACT AREAS 480
S.N. Issabayeva, A.K. Rysbayeva, Zh.B. Ibrayeva, N.R. Mussabekov, N.Sh. Akimzhan
MACHINE LEARNING-BASED CYBERATTACK DETECTION IN OT-NETWORKS: A COMPARATIVE ANALYSIS OF NETWORK ANOMALY CLASSIFICATION METHODS 495
G. Shangytbayeva, A.N. Amanov, M.K. Issenova, G.E. Berikkhanova, K.O. Shakerkhan
INFORMATION SECURITY IN GOVERNMENT CLOUD SERVICES: RISK ASSESSMENT AND PROTECTION STRATEGIES 512
Ye. N. Tursynbek, N. Albanbay, D. Djenouri
OPTIMISATION OF COMMUNICATION COSTS IN TWO-TIER FEDERATED LEARNING FOR CYBERATTACK DETECTION IN IOT NETWORKS 529
A. Tastemirov, S.B. Mukhanov, S.T. Amanzholova, B. Aldosh
A COMPARATIVE EVALUATION OF MACHINE LEARNING MODELS FOR ACTIVE DIRECTORY THREAT DETECTION: FEATURE-LEAKAGE CONTROL AND CROSS-SCENARIO VALIDATION ON WINDOWS SECURITY EVENT LOGS 545

МАЗМҰНЫ
ӘЛЕУМЕТТІК-ЭКОНОМИКАЛЫҚ ЖҮЙЕЛЕРДІ ДАМЫТУДАҒЫ ЦИФРЛЫҚ ТЕХНОЛОГИЯЛАР

А Адалбек, С.Е. Керімқұл, К.Д. Кулиев, О. Мекеш, Ж.Ж. Хамитова
ДЕРЕКТЕРДІҢ НӘТИЖЕ БАҒДАРЛАНҒАН ТЕГІСТЕУ МОДЕЛЬДЕРІН ӘЗІРЛЕУ ЖӘНЕ ОЛАРДЫҢ ТИІМДІЛІГІН ТАЛДАУ 11
А.Т. Искакова, Г.Е. Курманкулова, Конрад Громашек, К.О. Тогжанова, Р.С. Алгожаева
ӨНДІРІС ЖҮЙЕЛЕРІНДЕ БОЛЖАМДЫ ЖОСПАРЛАУ ЖӘНЕ РЕСУРСТАРДЫ ҮНЕМДЕУ ҮШІН МАШИНАЛЫҚ ЖӘНЕ ТЕРЕҢ ОҚЫТУДЫ ҚОЛДАНУ 30
А.Б. Касымова, Р.К. Ускенбаева, Ж.А. Сыдыков, Д. Өмірзақ, С. Калибеков
SMART CITY ҮШІН ИНТЕЛЛЕКТУАЛДЫ ЧАТ-БОТ АРХИТЕКТУРАСЫ: АЗАМАТТАРДЫҢ

ӨТІНІШТЕРІН ЖАСАНДЫ ИНТЕЛЛЕКТ НЕГІЗІНДЕ ҚАБЫЛДАУ, ТАЛДАУ ЖӘНЕ БАҒЫТТАУ 46
О.М. Олжаев, Б.С. Омаров, A.Suliman
YOLO-CBAM: ЖОЛДЫҢ ЗАҚЫМДАНУЫН ТАЛДАУҒА АРНАЛҒАН ИНТЕГРАЦИЯЛАНҒАН ЗЕЙІН МЕХАНИЗМІ БАР ЖЕҢІЛ САЛМАҚТЫ НЕЙРОНДЫҚ ЖЕЛІ МОДЕЛІ 66
К. Ермагамбетова, А. Сарсекеева, Е. Ахласов
ЦИФРЛЫҚ МӘДЕНИ МҰРАНЫ ЖІКТЕУДЕГІ ФРАГМЕНТАЦИЯНЫ ЕҢСЕРУ: ИНТЕГРАТИВТІ КӨПӨЛШЕМДІ МОДЕЛЬ 78

АҚПАРАТТЫҚ ТЕХНОЛОГИЯЛАР

Д.Е. Абжанова, С.Е. Шарипова
ЖАМБЫЛ ӨҢІРІНІҢ ТРАНСШЕКАРАЛЫҚ СУ РЕСУРСТАРЫН МОНИТОРИНГТЕУ ЖӘНЕ ДЕРЕКТЕР АЛМАСУҒА АРНАЛҒАН ЦИФРЛЫҚ ПЛАТФОРМАНЫҢ АРХИТЕКТУРАСЫ МЕН ПРАКТИКАЛЫҚ ІСКЕ АСЫРЫЛУЫ 101
А.Айтанұлы, О.А. Усатова
БҰЛТТЫ МЕДИЦИНАЛЫҚ ЖҮЙЕЛЕРДЕ ЕКІ ФАКТОРЛЫ АУТЕНТИФИКАЦИЯНЫ ЖАҚСАРТУ: ТЕРЕҢ ОҚЫТУ ӘДІСТЕРІНІҢ ҚОЛДАНУ МҮМКІНДІКТЕРІ 122
О. Альсеитов, О. Кучанский
ЭЛЕКТР СТАНЦИЯЛАРЫНЫҢ САҒАТТЫҚ ШЫҒАРЫНДЫЛАРЫН ФИЗИКАЛЫҚ ЗАҢДЫЛЫҚТАРДЫ ЕСКЕРЕТІН ГИБРИДТІ ОҚЫТУ: БЕЛГІСІЗДІК ПЕН ТАСЫМАЛДАУДЫ ЕСКЕРЕТІН КӨПЛАСТАУЫШТЫ БЕНЧМАРК 141
А.М. Ахметова, Д.О. Байжанова, Ж.Н. Алишева, Б.А. Шаймов
КРІ КӨРСЕТКІШТЕРІН ЕСКЕРЕ ОТЫРЫП, ТАБИҒИ ТІЛ МОДЕЛЬДЕРІНЕ АРНАЛҒАН ШЫҒЫС МАРШРУТТАУЫНЫҢ АДАПТИВТІ ЖҮЙЕСІ 160
Ә.Қ. Әйтім
ҚАЗАҚ МӘТІНДІК КОРПУСТАРЫН АВТОМАТТЫ ЖӘНЕ ЖАРТЫЛАЙ АВТОМАТТЫ АННОТАЦИЯЛАУ 180
Ж.Е. Балабаев, С.В. Белощицкая
ШАҒАТАЙ ҚОЛЖАЗБАЛАРЫНЫҢ МӘТІНІН ТАНУ: CRNN-ДІ НӨЛДЕН БАСТАП ОҚЫТУ МЕН ТРАНСФОРМЕР НЕГІЗІНДЕГІ ТРАНСФЕРЛІК ОҚЫТУДЫ САЛЫСТЫРУ 198
Д.Б. Баумуратова, Т.К. Жукабаева, К.А. Абдрахманов, З. Ахмад, А.А. Сеитов
ЖОҒАРЫ БІЛІМ БЕРУДЕ AI НЕГІЗДЕЛГЕН ЦИФРЛЫҚ ТРАНСФОРМАЦИЯҒА АРНАЛҒАН НЕЙРО-ЦИФРЛЫҚ ОҚЫТУ МОДЕЛІ 219
Т. Даоюн, Г. Амирханова, Фу Яньвэй, Е. Фарузқызы
ФЕЙК ВИДЕОЛАРДАҒЫ МУЛЬТИМОДАЛЬДЫ СЕМАНТИКАЛЫҚ МАНИПУЛЯЦИЯНЫҢ ПАЙДАЛАНУШЫЛАРДЫҢ ҚАБЫЛДАЙТЫН СЕНІМІНЕ ӘСЕР ЕТУ МЕХАНИЗМІ 241
Ж. Доумчариева, Д. Тусупов, А. Акынбекова, М. Хасанова, М. Кассим
ЦИФРЛЫҚ ЕГІЗДЕГІ СҮТ САПАСЫН АНСАМБЛЬДІК ЖӘНЕ ГИБРИДТІ ОҚЫТУ АРҚЫЛЫ ЖІКТЕУ 258
Қ.З. Игибаев, А.Х. Козбакова, О. Әуелбеков, Т. Мерембаев, Қ. Есентаев
БИОРЕАКТОРДА IoT ТЕХНОЛОГИЯЛАРЫН ПАЙДАЛАНА ОТЫРЫП ДИНАМИКАЛЫҚ АРАЛАСТЫРУ ҚҰРЫЛҒЫСЫН ЗЕРТТЕУ 279
А.Н. Исмуканова, Г.Кощанова, Д.В. Плескачев, Ж.Сарсенбаева, А.А. Дауренова
ЖАСАНДЫ ИНТЕЛЛЕКТ НЕГІЗІНДЕ ҒЫЛЫМИ ДЕРЕКТЕРДІ АВТОМАТТЫ ТҮРДЕ ЖІКТЕУ: ТРАНСФОРМЕР АРХИТЕКТУРАЛАРЫ МЕН ГИБРИДТІ ТӘСІЛДЕРДІҢ САЛЫСТЫРМАЛЫ ТАЛДАУЫ 306
Т.Ж. Камалов, Али Абд. Альмисреб, Мохаммед А. Салех
МАШИНАЛЫҚ ОҚЫТУҒА АРНАЛҒАН ГИБРИДТІ КВАНТТЫҚ-КЛАССИКАЛЫҚ АЛГОРИТМДЕРІН ЗЕРТТЕУ 323
Б. Кәрімқызы, К.Г. Карасенко, Н.А. Әлішеров, Н.М. Мұратбек, А. Кәрім
2021–2025 ЖЫЛДАР АРАЛЫҒЫНДАҒЫ ГЛИКЕМИЯЛЫҚ МОНИТОРИНГКЕ АРНАЛҒАН МЕДИЦИНАЛЫҚ ҚҰРЫЛҒЫЛАРДЫ ӨНДІРУ ЖӘНЕ ДИАГНОСТИКАЛАУДАҒЫ ЖАСАНДЫ ИНТЕЛЛЕКТ ӘДІСТЕРІ БОЙЫНША НЕГІЗГІ ЖАРИЯЛАНЫМДАР 338
С. К. Кумызбаева, К.Е. Багланбаев, К.С. Саркыт, М. Конырова
КҮЙ-ЖАЙДЫ КЛИМАТҚА СЕЗІМТАЛ ҰСЫНУМЕН ГИБРИДТІ ЭНЕРГОЖҮЙЕНІҢ ИНТЕЛЛЕКТУАЛДЫ ЭНЕРГОМЕНЕДЖМЕНТІН ҮШФАЗАЛЫ ОҚЫТУ 357
Н.Е. Қарабаев, С.К. Серикбаева, Е. Бенкхелифа, Б. Тасуов

IOT-ЖЕЛІЛЕРІНДЕГІ БАСЫП КІРУДІ АНЫҚТАУЖҮЙЕЛЕРІНДЕ МАШИНАЛЫҚ ОҚЫТУДЫҢ АНСАМБЛЬДІ КҮЛГІЛЕРІНІҢ БӘСЕКЕЛЕС ШАБУЫЛДАРҒА ТӨЗІМДІЛІГІН ЗЕРТТЕУ 369
М. Нұртас, Ғ. Капышев
МАШИНАЛЫҚ ОҚЫТУДЫ ҚОЛДАНА ОТЫРЫП, ПАЙДАЛЫ ҚАЗБАЛАРДЫҢ ПЕРСПЕКТИВАЛЫЛЫҒЫН БОЛЖАУ ҮШІН ГЕОКЕҢІСТІКТІК ЖӘНЕ ГЕОФИЗИКАЛЫҚ ДЕРЕКТЕРДІ АЛДЫН АЛА ӨҢДЕУ ЖӘНЕ БІРІКТІРУДІҢ ӘДІСТЕМЕЛІК НЕГІЗІ 392
А. Нефтиссов, А. Билощицкий, И. Казамбаев, А. Медетхан, Л. Кириченко
ГЕРМЕТИКАЛЫҚ АРХИВТІК САҚТАУ ҚАПТАМАЛАРЫНДА МИКРОКЛИМАТТЫ ЭНЕРГИЯ ТИІМДІЛІГІН ЕСКЕРЕ ОТЫРЫП ТҰРАҚТАНДЫРУ 405
Ә.Ж. Омуртаева, У.Т. Махажанова, А.К. Адилова, Ш.К. Ележанова, Я. Култан
АГРОӨНЕРКӘСІПТІК КЕШЕНДЕГІ ИНВЕСТИЦИЯЛЫҚ ТӘУЕКЕЛДЕРДІ БАҒАЛАУҒА АРНАЛҒАН БЕЙІМДЕЛГІШ НЕЙРОНДЫҚ-БҰЛДЫР МОДЕЛЬ: МАШИНАЛЫҚ ОҚЫТУДЫ ИНТЕГРАЦИЯЛАУ АРҚЫЛЫ ШЕШІМ ҚАБЫЛДАУДЫ ҚОЛДАУ СӘУЛЕТІН ЖАҢҒЫРТУ 421
Н. Шаждекеева, А. Мұханбетқалиева, А. Әбдіғалиева, А. Жанекешова, Л. Нұрғалиева
ФИЛЬМДЕРДІ ҰСЫНУ ЖҮЙЕСІН ҚҰРУ КЕЗІНДЕГІ КОЛЛАБОРАТИВТІК СҮЗГІЛЕУДІҢ МАТЕМАТИКАЛЫҚ МОДЕЛЬДЕРІН САЛЫСТЫРМАЛЫ ТАЛДАУ 445
С. Шармұханбет, Г. Тұрмұханова, О. Финдик, В. Махатова, М. Рахметов
ҚҰБЫРІШІЛІК РОБОТТЫ ПАЙДАЛАНУ АРҚЫЛЫ ҚҰБЫР АҚАУЛАРЫН АНЫҚТАУДЫҢ ЗИЯТКЕРЛІК ЖҮЙЕСІ 461

АҚПАРАТТЫҚ ҚАУІПСІЗДІК ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАРҒА АРНАЛҒАН

С.З. Дәруіш, А.У. Утегенова, А.У. Калижанова, Ж.Ж. Жумабекова
ЗЫМЫРАН ТАСЫМАЛДАУШЫЛАРДЫҢ ҚҰЛАУ АЙМАҚТАРЫНЫҢ ЛАСТАНУЫН МАШИНАЛЫҚ ОҚЫТУМЕН ЖІКТЕУ 480
С.Н. Исабаева, А.К. Рысбаева, Ж.Б. Ибраева, Н.Р. Муссабеков, Н.Ш. Акимжан
МАШИНАЛЫҚ ОҚЫТУ НЕГІЗІНДЕ ОТ-ЖЕЛІЛЕРІНДЕГІ КИБЕРШАБУЫЛДАРДЫ АНЫҚТАУ: ЖЕЛІЛІК АНОМАЛИЯЛАРДЫ ЖІКТЕУ ӘДІСТЕРІНІҢ САЛЫСТЫРМАЛЫ ТАЛДАУЫ 495
Г.А. Шангытбаева, А.Н. Аманов, М.К. Исенова, Г.Е. Берикханова, Қ.О. Шакерхан
МЕМЛЕКЕТТІК ОРГАНДАРДЫҢ БҰЛТТЫ СЕРВИСТЕРІНДЕГІ АҚПАРАТТЫҚ ҚАУІПСІЗДІК: ТӘУЕКЕЛДЕРДІ БАҒАЛАУ ЖӘНЕ ҚОРҒАУ СТРАТЕГИЯЛАРЫ 512
Е.Н. Турсынбек, Н. Албанбай, Д. Дженоури
IOT ЖЕЛІЛЕРІНДЕ КИБЕРШАБУЫЛДАРДЫ АНЫҚТАУ ҮШІН ЕКІ ДЕҢГЕЙЛІ ФЕДЕРАЦИЯЛАНҒАН ОҚЫТУДАҒЫ БАЙЛАНЫС ШЫҒЫНДАРЫН ОҢТАЙЛАНДЫРУ 529
А. Тастемиров, С.Т. Аманжолова, Б. Алдош, С.Б. Муханов
ACTIVE DIRECTORY ЖҮЙЕСІНДЕ ҚАУІПТЕРДІ АНЫҚТАУ ҮШІН МАШИНАЛЫҚ ОҚЫТУ МОДЕЛЬДЕРІН САЛЫСТЫРМАЛЫ БАҒАЛАУ: WINDOWS ҚАУІПСІЗДІК ОҚИҒАЛАРЫ ЖУРНАЛДАРЫНДА БЕЛГІ ЖЫЛЫСТАУЫН БАҚЫЛАУ ЖӘНЕ КРОСС-СЦЕНАРЛЫҚ ВАЛИДАЦИЯ 545

СОДЕРЖАНИЕ
ЦИФРОВЫЕ ТЕХНОЛОГИИ В РАЗВИТИИСОЦИО-ЭКОНОМИЧЕСКИХ СИСТЕМ

А Адалбек, С.Е. Керимкулов, К.Д. Кулиев, О. Мекеш, Ж.Ж. Хамитова
РАЗРАБОТКА РЕЗУЛЬТАТИВНО-ОРИЕНТИРОВАННЫХ МОДЕЛЕЙ СГЛАЖИВАНИЯ ДАННЫХ И АНАЛИЗ ИХ ЭФФЕКТИВНОСТИ 11
А.Т. Искакова, Г.Е. Курманкулова, К. Громашек, К.О. Тогжанова, Р.С. Алгожаева
ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО И ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗНОГО ПЛАНИРОВАНИЯ И РЕСУРСОСБЕРЕЖЕНИЯ В ПРОИЗВОДСТВЕННЫХ СИСТЕМАХ 30
А.Б. Касымова, Р.К. Ускенбаева, Ж.А.Сыдыков, Д. Омирзак, С. Калибеков
АРХИТЕКТУРА ИНТЕЛЛЕКТУАЛЬНОГО ЧАТ-БОТА ДЛЯ SMART CITY: ПРИЁМ, АНАЛИЗ И МАРШРУТИЗАЦИЯ ОБРАЩЕНИЙ ГРАЖДАН НА ОСНОВЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА 46
О.М. Олжаев, Б.С. Омаров, A.Suliman
YOLO-CBAM: ЛЕГКОВЕСНАЯ НЕЙРОСЕТЕВАЯ МОДЕЛЬ С ИНТЕГРИРОВАННЫМ МЕХАНИЗМОМ ВНИМАНИЯ ДЛЯ АНАЛИЗА ДОРОЖНЫХ ПОВРЕЖДЕНИЙ 66
К. Ермагамбетова, А. Сарсекеева, Е. Ахласов
ПРЕОДОЛЕНИЕ ФРАГМЕНТАЦИИ В КЛАССИФИКАЦИИ ЦИФРОВОГО КУЛЬТУРНОГО НАСЛЕДИЯ:

ИНТЕГРАТИВНАЯ МНОГОМЕРНАЯ МОДЕЛЬ 78

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Д.Е. Абжанова, С.Е. Шарипова
АРХИТЕКТУРА И ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ ЦИФРОВОЙ ПЛАТФОРМЫ МОНИТОРИНГА И ОБМЕНА ДАННЫМИ О ТРАНСГРАНИЧНЫХ ВОДНЫХ РЕСУРСАХ ЖАМБЫЛСКОГО РЕГИОНА 101
А. Айтанулы, О.А. Усатова
СОВЕРШЕНСТВОВАНИЕ ДВУХФАКТОРНОЙ АУТЕНТИФИКАЦИИ ДЛЯ ОБЕСПЕЧЕНИЯ БЕЗОПАСНОСТИ ОБЛАЧНЫХ МЕДИЦИНСКИХ СИСТЕМ: ВОЗМОЖНОСТИ ПРИМЕНЕНИЯ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ 122
О. Альсеитов, О. Кучанский
ФИЗИЧЕСКИ ИНФОРМИРОВАННОЕ ГИБРИДНОЕ ОБУЧЕНИЕ ДЛЯ ПОЧАСОВОГО ПРОГНОЗИРОВАНИЯ ВЫБРОСОВ ЭЛЕКТРОСТАНЦИЙ: МНОГОКОМПОНЕНТНЫЙ БЕНЧМАРК С ОЦЕНКОЙ НЕОПРЕДЕЛЁННОСТИ И ПЕРЕНОСИМОСТЬЮ МОДЕЛЕЙ 141
А.М. Ахметова, Д.О. Байжанова, Ж.Н. Алишева, Б.А. Шаймов
АДАПТИВНЫЙ МАРШРУТИЗАТОР ВЫВОДА С УЧЕТОМ КЛЮЧЕВЫХ ПОКАЗАТЕЛЕЙ ЭФФЕКТИВНОСТИ ДЛЯ МОДЕЛЕЙ ЕСТЕСТВЕННОГО ЯЗЫКА 160
Ә.Қ. Әйтім
АВТОМАТИЧЕСКАЯ И ПОЛУАВТОМАТИЧЕСКАЯ АННОТАЦИЯ КАЗАХСКИХ ТЕКСТОВЫХ КОРПУСОВ 180
Ж.Е. Балабаев, С.В. Белощицкая
РАСПОЗНАВАНИЕ РУКОПИСНОГО ТЕКСТА ШАГАТАЙСКИХ РУКОПИСЕЙ: СРАВНЕНИЕ ОБУЧЕНИЯ CRNN С НУЛЯ И ТРАНСФЕРНОГО ОБУЧЕНИЯ НА БАЗЕ ТРАНСФОРМЕРОВ 198
Д.Б. Баумуратова, Т.К. Жукабаева, К.А. Абдрахманов, З. Ахмад, А.А. Сеитов
НЕЙРО-ЦИФРОВАЯ МОДЕЛЬ ОБУЧЕНИЯ ДЛЯ ЦИФРОВОЙ ТРАНСФОРМАЦИИ ВЫСШЕГО ОБРАЗОВАНИЯ С ПОМОЩЬЮ AI 219
Т. Даоюн, Г. Амирханова, Фу Яньвэй, Е. Фарузкызы
МЕХАНИЗМ ВЛИЯНИЯ МУЛЬТИМОДАЛЬНОЙ СЕМАНТИЧЕСКОЙ МАНИПУЛЯЦИИ В ФЕЙКОВЫХ ВИДЕО НА ВОСПРИНИМАЕМОЕ ДОВЕРИЕ ПОЛЬЗОВАТЕЛЕЙ 241
Ж. Доумчариева, Д. Тусупов, А. Акынбекова, М. Хасанова, М. Кассим
КЛАССИФИКАЦИЯ КАЧЕСТВА МОЛОКА В ЦИФРОВОМ ДВОЙНИКЕ АНСАМБЛЕВЫМИ И ГИБРИДНЫМИ МЕТОДАМИ 258
К.З. Игибаев, А.Х. Козбакова, О. Ауелбеков, Т. Мерембаев, К. Есентаев
ИССЛЕДОВАНИЕ УСТРОЙСТВА ДИНАМИЧЕСКОГО ПЕРЕМЕШИВАНИЯ В БИОРЕАКТОРЕ С ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИЙ IoT 279
А. Н. Исмуканова, Г.Кощанова, Д.В. Плескачев, Ж.Сарсенбаева, А.А. Дауренова
АВТОМАТИЗИРОВАННАЯ КЛАССИФИКАЦИЯ НАУЧНЫХ ДАННЫХ НА ОСНОВЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: СРАВНИТЕЛЬНЫЙ АНАЛИЗ ТРАНСФОРМЕРНЫХ АРХИТЕКТУР И ГИБРИДНЫХ МЕТОДОВ 306
Т.Ж. Камалов, Али Абд. Альмисреб, Мохаммед А. Салех
ИССЛЕДОВАНИЕ ГИБРИДНЫХ КВАНТОВО-КЛАССИЧЕСКИХ АЛГОРИТМОВ ДЛЯ МАШИННОГО ОБУЧЕНИЯ 323
Б. Каримкызы, К.Г. Карасенко, Н.А. Алишеров, Н.М. Муратбек, А. Карим
КЛЮЧЕВЫЕ ПУБЛИКАЦИИ 2021–2025 ГОДОВ ПО МЕТОДАМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ПРОИЗВОДСТВЕ И ДИАГНОСТИКЕ МЕДИЦИНСКИХ ИЗДЕЛИЙ ДЛЯ ГЛИКЕМИЧЕСКОГО МОНИТОРИНГА 338
С.К. Кумызбаева, К.Е. Багланбаев, К.С. Сарқыт, М.Ж. Конырова
ТРЁХФАЗОВОЕ ОБУЧЕНИЕ ИНТЕЛЛЕКТУАЛЬНОГО ЭНЕРГОМЕНЕДЖМЕНТА ГИБРИДНОЙ ЭНЕРГОСИСТЕМЫ С КЛИМАТО-ОСВЕДОМЛЁННЫМ ПРЕДСТАВЛЕНИЕМ СОСТОЯНИЯ 357
Н.Е. Қарабаев, С.К. Серикбаева, Е. Бенкхелифа, Б. Тасуов
ИССЛЕДОВАНИЕ УСТОЙЧИВОСТИ АНСАМБЛЕВЫХ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ К СОСТЯЗАТЕЛЬНЫМ АТАКАМ В СИСТЕМАХ ОБНАРУЖЕНИЯ ВТОРЖЕНИЙ IOT-СЕТЕЙ 369
М. Нуртас, Г. Капышев
МЕТОДОЛОГИЧЕСКАЯ ОСНОВА ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ И ИНТЕГРАЦИИ ГЕОПРОСТРАНСТВЕННЫХ И ГЕОФИЗИЧЕСКИХ ДАННЫХ ДЛЯ ПРОГНОЗИРОВАНИЯ ПЕРСПЕКТИВНО-

СТИ МИНЕРАЛЬНЫХ РЕСУРСОВ С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ 392
А. Нефтиссов, А. Билощицкий, И. Казамбаев, А. Медетхан, Л. Кириченко
ЭНЕРГОЭФФЕКТИВНАЯ СТАБИЛИЗАЦИЯ МИКРОКЛИМАТА В ГЕРМЕТИЧНЫХ АРХИВНЫХ ХРАНИЛИЩАХ 405
А.Ж. Омуртаева, У.Т. Махажанова, А.К. Адилова, Ш.К. Ележанова, Я.Култан
АДАПТИВНАЯ НЕЙРО-НЕЧЕТКАЯ МОДЕЛЬ ДЛЯ ОЦЕНКИ ИНВЕСТИЦИОННЫХ РИСКОВ В АГРОПРОМЫШЛЕННОМ КОМПЛЕКСЕ: МОДЕРНИЗАЦИЯ АРХИТЕКТУРЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПОСРЕДСТВОМ ИНТЕГРАЦИИ МАШИННОГО ОБУЧЕНИЯ 421
Н. Шаждекеева, А. Муханбеткалиева, А. Абдигалиева, А. Жанекешова, Л. Нургалиева
СРАВНИТЕЛЬНЫЙ АНАЛИЗ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ КОЛЛАБОРАТИВНОЙ ФИЛЬТРАЦИИ ПРИ ПОСТРОЕНИИ РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ ФИЛЬМОВ 445
С. Шармуханбет, Г. Турмуханова, О.Финдик, В.Махатова, М. Рахметов
ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА ОБНАРУЖЕНИЯ ДЕФЕКТОВ ТРУБОПРОВОДОВ С ИСПОЛЬЗОВАНИЕМ ВНУТРИТРУБНОГО РОБОТА 461

ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ И КОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ

С.З. Дәруіш, А.У. Утегенова, А.У. Калижанова, Ж.Ж. Жумабекова
МАШИННОЕ ОБУЧЕНИЕ ДЛЯ КЛАССИФИКАЦИИ ЗАГРЯЗНЕНИЯ РАЙОНОВ ПАДЕНИЯ РАКЕТ-НОСИТЕЛЕЙ 480
С.Н. Исабаева, А.К. Рысбаева, Ж.Б. Ибраева, Н.Р. Муссабеков, Н.Ш. Акимжан
ОБНАРУЖЕНИЕ КИБЕРАТАК В ОТ-СЕТЯХ НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ: СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ КЛАССИФИКАЦИИ СЕТЕВЫХ АНОМАЛИЙ 495
Г.А. Шангытбаева, А.Н. Аманов, М.К. Исенова, Г.Е. Берикханова, Қ.О. Шакерхан
ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ В ОБЛАЧНЫХ СЕРВИСАХ ГОСУДАРСТВЕННЫХ ОРГАНОВ: ОЦЕНКА РИСКОВ И СТРАТЕГИИ ЗАЩИТЫ 512
Е.Н. Турсынбек, Н. Албанбай, Д. Дженоури
ОПТИМИЗАЦИЯ КОММУНИКАЦИОННЫХ ЗАТРАТ В ДВУХУРОВНЕВОМ ФЕДЕРАТИВНОМ ОБУЧЕНИИ ДЛЯ ОБНАРУЖЕНИЯ КИБЕРАТАК В IOT-СЕТЯХ 529
А. Тастемиров, С.Т. Аманжолова, Б. Алдош, С.Б. Муханов
СРАВНИТЕЛЬНАЯ ОЦЕНКА МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОБНАРУЖЕНИЯ УГРОЗ В ACTIVE DIRECTORY: КОНТРОЛЬ УТЕЧКИ ПРИЗНАКОВ И КРОСС-СЦЕНАРНАЯ ВАЛИДАЦИЯ НА ЖУРНАЛАХ БЕЗОПАСНОСТИ WINDOWS 545

DIGITAL TECHNOLOGIES IN THE DEVELOPMENT OF SOCIO-ECONOMIC SYSTEMS
ӘЛЕУМЕТТІК-ЭКОНОМИКАЛЫҚ ЖҮЙЕЛЕРДІ ДАМЫТУДАҒЫ ЦИФРЛЫҚ ТЕХНОЛОГИЯЛАР
ЦИФРОВЫЕ ТЕХНОЛОГИИ В РАЗВИТИИСОЦИО-ЭКОНОМИЧЕСКИХ СИСТЕМ

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 11–29
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.001

DEVELOPMENT OF RESULT-ORIENTED DATA SMOOTHING MODELS AND ANALYSIS OF THEIR EFFECTIVENESS

A Adalbek[1]*, S.Ye. Kerimkhulle[1], K.D. Kuliev[2], O. Mekesh[3], Zh.Zh. Khamitova[1]
[1]L.N. Gumilyov Eurasian National University, Astana, Kazakhstan;
[2]Samarkand State University named after Sh.Rashidov, Samarkand, Uzbekistan;
[3]S. Seifullin Kazakh Agrotechnical Research University, Astana, Kazakhstan.
E–mail: alibek.adalbek@gmail.com

Alibek Adalbek — PhD candidate of the Department of Information Systems, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
E–mail: alibek.adalbek@gmail.com, https://orcid.org/0000–0001–8543–2192;
Seyit Ye. Kerimkhulle — Professor of the Department of Information Systems, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
https://orcid.org/0000–0002–5886–6064;
Komil D. Kuliev — Professor, of the Department «Mathematical physics and functional analysis», Samarkand State University named after Sh.Rashidov, Samarkand, Uzbekistan
https://orcid.org/0000–0002–1111–456X;
Oralbek Mekesh — Senior Lecturer, Department of Higher Mathematics, S. Seifullin Kazakh Agrotechnical Research University, Astana, Kazakhstan
https://orcid.org/0000-0001-7816-8856;
Zhainagul Zh. Khamitova — PhD candidate of the Department of Information Systems, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
https://orcid.org/0009-0005-3351-7449.

© A. Adalbek, S.Y. Kerimkhulle, K.D. Kuliev, O. Mekesh, Zh.Zh. Khamitova

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Abstract. This study examines four widely used algorithms for smoothing fluctuating data: Simple Exponential Smoothing, Moving Average, Random Walk, and Savitzky–Golay. Their effectiveness was evaluated using root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and signal-to-noise ratio (SNR). The evaluation was conducted using Kazakhstan's gross domestic product (GDP) data for 1990–2024 obtained from official World Bank sources. The developed software module enables users to smooth time series and discrete measurements using four methods and adjust estimation parameters to identify optimal settings. It also supports data visualization and export in CSV, XLSX, JSON, and PDF formats. The Savitzky–Golay filter achieved the lowest RMSE (5.27) and the highest SNR (28.49 dB). The scientific novelty lies in integrating four smoothing algorithms into a unified web-based architecture with flexible parameter control, support for multiple input formats, and comparative evaluation using RMSE, MAE, MAPE, and SNR. The module can be applied to time-series preprocessing and economic forecasting.

Keywords: time series, Savitzky–Golay, exponential smoothing, Moving Average, GDP, estimation parameters, intelligent systems

For citation: Adalbek A., Kerimkhulle S.Y., Kuliev K.D., Mekesh O., Khamitova Zh.Zh. (2026). Development of result-oriented data smoothing models and analysis of their effectiveness // International journal of information and communication technologies. Vol. 7. No. 27. Pp. 11–29. https://doi.org/10.54309/IJICT.2026.27.3.001. (In Kaz.).

Conflict of interest: The authors declare that there is no conflict of interest.

Acknowledgements. *This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP26196337).*

ДЕРЕКТЕРДІҢ НӘТИЖЕ БАҒДАРЛАНҒАН ТЕГІСТЕУ МОДЕЛЬДЕРІН ӘЗІРЛЕУ ЖӘНЕ ОЛАРДЫҢ ТИІМДІЛІГІН ТАЛДАУ

А Адалбек[1]*, С.Е. Керімқұл[1], К.Д. Кулиев[2], О. Мекеш[3], Ж.Ж. Хамитова[1]

[1]Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан;

[2]Ш. Рашидов атындағы Самарқан мемлекеттік университеті, Самарқан, Өзбекстан;

[3]С. Сейфуллин атындағы Қазақ агротехникалық зерттеу университеті, Астана, Қазақстан.

E–mail: alibek.adalbek@gmail.com

Адалбек Алибек — «Ақпараттық жүйелер» кафедрасының PhD докторанты, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан
E–mail: alibek.adalbek@gmail.com, https://orcid.org/0000–0001–8543–2192;

Керімқұл Сейіт Есілбайұлы — «Ақпараттық жүйелер» кафедрасының профессоры, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

https://orcid.org/0000–0002–5886–6064;
Кулиев Комил Данабаевич — «Математикалық физика және функционалдық талдау» кафедрасының профессоры, Ш. Рашидов атындағы Самарқан мемлекеттік университеті, Самарқан, Өзбекстан
https://orcid.org/0000–0002–1111–456X;
Мекеш Оралбек — «Жоғары математика» кафедрасының аға оқытушы, С. Сейфуллин атындағы Қазақ агротехникалық зерттеу университеті, Астана, Қазақстан
https://orcid.org/0000-0001-7816-8856;
Хамитова Жайнагуль Жанаткызы — «Ақпараттық жүйелер» кафедрасының PhD докторанты, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан
https://orcid.org/0009-0005-3351-7449.

© А. Адалбек, С.Е. Керімқұл, К.Д. Кулиев, О. Мекеш, Ж.Ж. Хамитова

Аннотация. Зерттеуде ауытқымалы деректерді тегістеуге арналған кеңінен қолданылатын төрт алгоритм қарастырылады: қарапайым экспоненциалды тегістеу, жылжымалы орташа, кездейсоқ жүріс және Савицкий–Голей сүзгісі. Әдістердің тиімділігі орташа квадраттық қате (RMSE), орташа абсолюттік қате (MAE), орташа абсолюттік пайыздық қате (MAPE) және сигнал/шу қатынасы (SNR) көрсеткіштері арқылы бағаланды. Бағалау үшін Дүниежүзілік банктің ресми дереккөздерінен алынған Қазақстанның 1990–2024 жылдардағы жалпы ішкі өнімінің (ЖІӨ) деректері пайдаланылды. Әзірленген бағдарламалық модуль уақыттық қатарлар мен дискретті өлшемдерді төрт әдіс арқылы тегістеуге және оңтайлы мәндерді анықтау үшін бағалау параметрлерін реттеуге мүмкіндік береді. Сонымен қатар, модуль нәтижелерді визуализациялауды және деректерді CSV, XLSX, JSON және PDF форматтарында экспорттауды қолдайды. Савицкий–Голей сүзгісі орташа квадраттық қатенің ең төменгі мәнін (RMSE = 5,27) және сигнал/шу қатынасының ең жоғары көрсеткішін (SNR = 28,49 дБ) көрсетті. Зерттеудің ғылыми жаңалығы төрт тегістеу алгоритмін параметрлерді икемді реттеуді, әртүрлі кіріс деректер форматтарын қолдауды және RMSE, MAE, MAPE және SNR көрсеткіштері бойынша әдістерді салыстырмалы бағалауды қамтамасыз ететін бірыңғай веб-архитектураға біріктірумен сипатталады. Әзірленген модуль уақыттық қатарларды алдын ала өңдеуде және экономикалық болжау жүйелерінде қолданылуы мүмкін.

Түйінді сөздер: уақыт қатарлары, Савицки–Голай сүзгісі, қарапайым экспоненциалды тегістеу, жылжымалы орта, ЖІӨ, бағалау параметрлері, интеллектуалды жүйелер

Дәйексөздер үшін: Адалбек А., Керімқұл С.Е., Кулиев К.Д., Мекеш О., Хамитова Ж.Ж. (2026). Деректердің нәтиже бағдарланған тегістеу модельдерін әзірлеу және олардың тиімділігін талдау // Халықаралық ақпараттық және коммуникациялық технологиялар журналы. Т. 7. No 27. Б. 11–29. https://doi.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

org/10.54309/IJICT.2026.27.3.001. (Қаз. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

Алғыс. *Бұл зерттеу Қазақстан Республикасының Ғылым және жоғары білім министрлігінің Ғылым комитеті тарапынан қаржыландырылды (Грант № AP26196337).*

РАЗРАБОТКА РЕЗУЛЬТАТИВНО-ОРИЕНТИРОВАННЫХ МОДЕЛЕЙ СГЛАЖИВАНИЯ ДАННЫХ И АНАЛИЗ ИХ ЭФФЕКТИВНОСТИ

А Адалбек[1]*, С.Е. Керимкулов[1], К.Д. Кулиев[2], О. Мекеш[3], Ж.Ж. Хамитова[1]

[1]Евразийский национальный университет им. Л.Н. Гумилёва, Астана, Казахстан;

[2]Самаркандский государственный университет им. Ш.Рашидова, Самарканд, Узбекистан;

[3]Казахский агротехнический научно-исследовательский университет им. С. Сейфуллина, Астана, Казахстан.

E-mail: alibek.adalbek@gmail.com

Адалбек Алибек — докторант кафедры информационных систем Евразийского национального университета им. Л.Н. Гумилёва, Астана, Казахстан
E-mail: alibek.adalbek@gmail.com, https://orcid.org/0000–0001–8543–2192;

Керимкулов Сейт Есилбаевич — профессор кафедры информационных систем Евразийского национального университета им. Л.Н. Гумилёва, Астана, Казахстан
https://orcid.org/0000–0002–5886–6064;

Кулиев Комил Данабаевич — профессор кафедры «Математическая физика и функциональный анализ», Самаркандский государственный университет им. Ш. Рашидова, Самарканд, Узбекистан;

Мекеш Оралбек — старший преподаватель кафедры высшей математики Казахского научно-исследовательского агротехнического университета им. С. Сейфуллина, Астана, Казахстан
https://orcid.org/0000-0001-7816-8856;

Хамитова Жайнагуль Жанаткызы — докторант кафедры информационных систем Евразийского национального университета им. Л.Н. Гумилёва, Астана, Казахстан
https://orcid.org/0009-0005-3351-7449.

© А. Адалбек, С.Е. Керимкулов, К.Д. Кулиев, О. Мекеш, Ж.Ж. Хамитова

Аннотация. В исследовании рассматриваются четыре широко используемых алгоритма сглаживания флуктуирующих данных: простое экспоненциальное сглаживание, скользящее среднее, случайное блуждание и фильтр

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Савицкого–Голея. Эффективность методов оценивалась с использованием среднеквадратичной ошибки (RMSE), средней абсолютной ошибки (MAE), средней абсолютной процентной ошибки (MAPE) и отношения сигнал/шум (SNR). Для оценки использовались данные о валовом внутреннем продукте (ВВП) Казахстана за 1990–2024 годы, полученные из официальных источников Всемирного банка. Разработанный программный модуль позволяет сглаживать временные ряды и дискретные измерения с помощью четырёх методов, а также настраивать параметры оценки для определения оптимальных значений. Модуль поддерживает визуализацию результатов и экспорт данных в форматах CSV, XLSX, JSON и PDF. Фильтр Савицкого–Голея продемонстрировал наименьшую среднеквадратичную ошибку (RMSE = 5,27) и наибольшее отношение сигнал/шум (SNR = 28,49 дБ). Научная новизна исследования заключается в интеграции четырёх алгоритмов сглаживания в единую веб-архитектуру, обеспечивающую гибкую настройку параметров, поддержку различных форматов входных данных и сравнительную оценку методов по показателям RMSE, MAE, MAPE и SNR. Разработанный модуль может применяться для предварительной обработки временных рядов и в системах экономического прогнозирования.

Для цитирования: Адалбек А., Керимкулов С.Е., Кулиев К.Д., Мекеш О., Хамитова Ж.Ж. (2026). Разработка результативно-ориентированных моделей сглаживания данных и анализ их эффективности // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 27. Стр. 11–29. https://doi.org/10.54309/IJICT.2026.27.3.001. (На каз.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Благодарности. *Данное исследование профинансировано Комитетом науки Министерства науки и высшего образования Республики Казахстан (грант № AP26196337).*

Кіріспе.

Ұлттық экономикалық көрсеткіштердің уақыт қатарлары көбіне кездейсоқ тербелістер мен маусымдық ауытқуларға (Savitzky et al., 1964) жиі ие болады. Мұндай деректерді талдау барысында бастапқы деректердің тұрақсыздығы болжам дәлдігіне теріс әсер етеді. Сондықтан да бастапқы деректерге тегістеу әдістерін қолдану ғылыми зерттеу нәтижелерінің сенімділігін барынша арттырады. Соңғы кездері жылжымалы орташа, экспоненциалды тегістеу және Савицки–Голей сияқты тегістеу әдістері статистикалық талдау мен машиналық оқыту жүйелерінде кеңінен қолданыла бастады (Savitzky et al., 1964; Hening et al., 2022).

Дегенмен, қолданыстағы зерттеулер көбіне бір ғана тегістеу тәсілімен (Önder et al., 2013) шектеледі, сонымен осы тәсілдерді салыстырған зерттеулерде бар (Kaneko et al., 2015), бірақ оларда параметрлерін икемді түрде басқару мүмкіндігі аз. Осыған байланысты бұл зерттеу төрт классикалық әдісті біріктіретін, бірнеше деректер форматтарын қолдайтын, параметрлерді автоматты басқаруды, визуализацияны, экспорттау функцияларын және бірыңғай статистикалық баға-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

лауды қамтамасыз ететін бірыңғай тегістеу модулін әзірлеу арқылы үлес қосады (Petropoulos et al., 2022; Önder et al., 2013).

Ғылыми зерттеулерде (Тураров т.б., 2023)қаржылық болжам жасау кезінде уақыт қатарларын тегістеу және трендтерді анықтау үшін жылжымалы орташа сияқты әдістер қолданылады, ал (Адамбекова т.б., 2019) зерттеуде қор индексі деректері де айнымалы және шуылы көп болғандықтан, авторлар орташа мәндер арқылы тегістеу, яғни жылжымалы орта элементтерін қолданады. Сонымен бірге (Утибаев т.б., 2019)зерттеуінде ауылшаруашылық өндірісінің динамикасын зерттеген, яғни маусымдық және жылдық өзгерістерді тегістеу үшін жылжымалы орташа және корреляциялық әдістер пайдаланылған. Демек, бұл мысалдар экономикалық және өндірістік статистикада тегістеу әдістерінің қолданылуын нақтылайды.

Бұл ғылыми зерттеуде ұсынылып отырған модуль – уақыт қатарлары мен дискретті өлшеулерден келетін деректерді бағалау параметрлеріне сай тегістеу әдісін таңдауға арналған бағдарламалық компонент. Модуль қолданушыға төрт тегістеу әдісінің бірін таңдау арқылы деректердегі шу мен кездейсоқ ауытқуларды азайтып, негізгі ақпараттарды жоғалтпай маңызды құрылымды анықтауға және экономикалық болжам жүйелерін оңтайлы жетілдіруде қолдануға мүмкіндік береді. Мақала құрылымы келесідей: 2–бөлімде әдістемелік негіздер, 3–бөлімде эксперименттік нәтижелер мен бағалау нәтижелері берілген, 4–бөлімде қорытынды жасалған.

Әдістер мен материалдар.

Бұл бөлімде қолданушы деректерінің құрылымы мен типіне сәйкес келетін тегістеудің төрт негізгі әдісіне қысқаша теориялық сипаттама берілген, сонымен қатар әдістерді пайдалану қадамдары мен қандай деректерге қолдануға болатыны туралы қысқаша нұсқаулық келтірілген. Енгізілген деректерді тегістеу модулінің архитектурасы 1–суретте сипатталған:

Енді осы архитектураның деректермен жұмыс істеу принциптерін қадам бойынша қарастырайық.

1. Қолданушы деректерді енгізу қабаты (User Input Layer). Бұл қабат қолданушымен жүйе арасындағы интерфейс қызметін атқарады, яғни жүйеге деректерді тізбек түрінде енгізуге болады немесе *.csv, *.xlsx, *.json форматындағы файл ретінде жүктеуге болады. Содан кейін алдын ала көрсетілген форматтарға сәйкестігін тексеру мен валидация жасалады. Енгізілген деректерде мынадай міндетті өрістер болу керек: уақыттық индекс (timestamp немесе integer), өлшеу мәні (float).

Жүйеге енгізілетін деректердің жалпы түрі келесідей:

– *деректер* (data): тізбектелген нүктелер $\{(t_i, x_i)\}_{i=1}^{N}$.

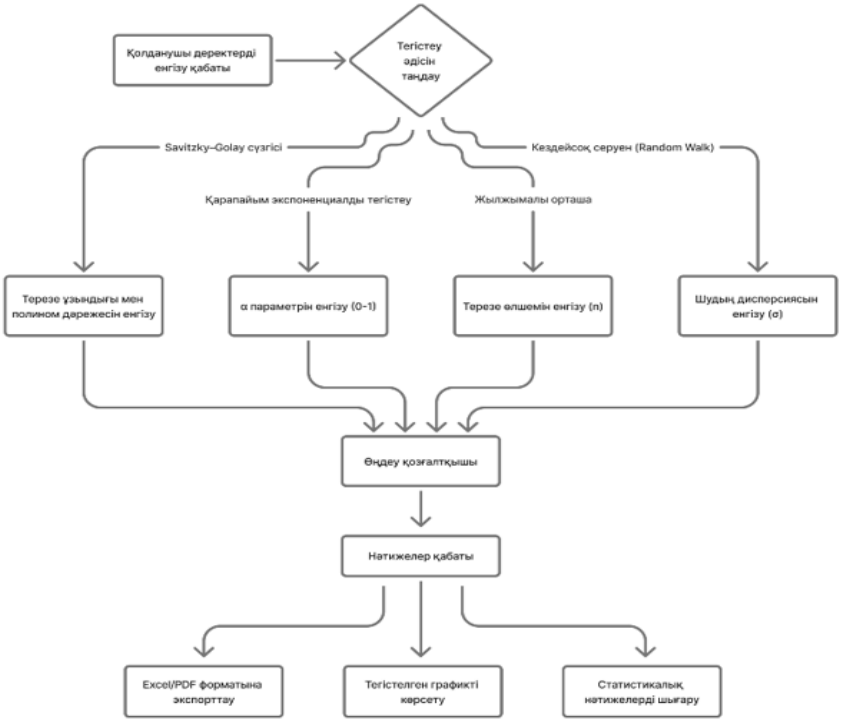
– *деректер метадеректері* (metadata): уақыттық қадам: 1 жыл, өлшем бірлігі: млрд АҚШ доллары, бастапқы статистика: (орташа, медиана, стандарттық ауытқу).

2. Тегістеу әдісін таңдау (Method Selection)

Бұл қабатта қолданушыға ұсынылған төрт әдістің бірін таңдау мүмкін-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

дігін береді: қарапайым экспоненциалды тегістеу, жылжымалы орташа, кездейсоқ жүріс (модель/салыстырмалы базис ретінде), Савицки–Голай сүзгісі. Әдісті таңдау интерфейсі пайдаланушыға әр әдістің қысқаша сипаттамасын және қолдану шарттарын көрсетеді (trend/seasonality болуы, сигналдың спектрлік қасиеттері және т.б.).



Сур. 1. Ұсынылған деректерді тегістеу модулінің жалпы архитектурасы және ақпараттық ағыны

Қолданушыға әдіс таңдауында есептеу мақсатына және бастапқы деректердің статистикалық қасиеттеріне қарай нақты нұсқаулар беріледі. Мысалы:

– Егер есептің мақсаты – соңғы мәнге жанасатын тегістеу болса, яғни соңғы мәнге көбірек салмақ беру керек болса, онда қарапайым экспоненциалды тегістеу ұсынылады.

– Егер есептің мақсаты – берілген сигналдың орташа қозғалысын қарапайым түрде сипаттау болса, онда Moving Average тегістеуі ұсынылады.

– Егер есептің мақсаты – берілген сигнал деңгейіндегі нүктелер арасындағы құрылымды сақтау және сигналдың максимум/минимум нүктелерін бұзбау қажет болса, онда Savitzky–Golay тегістеуі ұсынылады.

– Егер есептің мақсаты – болжанбайтын, кездейсоқ сипаттағы уақыт қатарларын модельдеу болса, онда Random Walk моделі тегістеуі ұсынылады.

3. Әр әдіске тән параметрлер мен олардың семантикасы

3.1 Қарапайым экспоненциалды тегістеу (Simple Exponential Smoothing)

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Қолданатын параметрі $\alpha, 0 < \alpha \leq 1$ болып табылады (Woo et al., 2022; Önder et al., 2013; Swari et al., 2021).

Математикалық өрнектелуі:

$$S_t = \alpha x_t + (1-\alpha)S_{t-1}, S_1 = x_1. \quad (1)$$

мұндағы $x_t - t$ – шы нүктедегі бастапқы дерек, $S_t - t$ – шы нүктедегі тегістелген мән, $S_{t-1} - t - 1$- ші нүктедегі тегістелген мән.

Мұнда α мәні – соңғы мәндерге берілетін салмақты басқарады. $\alpha \to 1$ болса соңғы мәндерге тәуелділігі артып, тегістеу «жеңілдейді»; $\alpha \to 0$ болса, тегістеу көп жылдамдықты тербелістерді басады.

Әдетте бастапқы баптаулар: 0.1–0.3 – қатты тегістеу, 0.6–0.9 – жеңіл тегістеу болып табылады.

3.2 Жылжымалы орташа (Moving Average)

Қолданатын параметрі: – терезе өлшемі (оң, бүтін, тақ сан).

Математикалық өрнектелуі (Petropoulos et al., 2022; Hewamalage et al., 2021; Swari et al., 2021):

$$MA_t = \frac{1}{n}\sum_{i=0}^{n-1} x_{t-i}. \quad (2)$$

мұндағы $x_t - t$ – шы нүктедегі бастапқы дерек, $MA_t - t$ – шы нүктедегі тегістелген мән.

Мұнда *n* үлкен болса, тегістеу күшті, әрі кешігу (lag) артады; *n* кішкентай болса сигналдың бастапқы құрылымы сақталады.

3.3 Кездейсоқ жүріс (Random Walk)

Қолданатын параметрі (Önder et al., 2013): σ – қадамдық шуылдың стандарттық ауытқуы (немесе дисперсия σ^2).

Модельдің формуласы:

$$x_t = x_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2). \quad (3)$$

мұндағы x_t– бастапқы позиция, $\varepsilon_t - t$– шы қадамды сипаттайтын кездейсоқ шама.

Мұнда кездейсоқ жүріс – деректердің эволюциясын модельдеу үшін қолданылатын стохастикалық процесс болып табылады және ол өзіндік «тегістеу» емес, керісінше, ол көбіне деректің кездейсоқ прогрессиясын сипаттауға қолданылады.

3.4 Савицки–Голай сүзгісі

Қолданатын параметрі (Savitzky et al., 1964; Schmid et al., 2022; Liang et al., 2022): терезе өлшемі $(window_length)$ оң, бүтін, тақ сан, полином дәрежесі

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

$(polyorder)$ $(0 \leq polyorder < window_length)$ көп жағдайда $polyorder = 2$ немесе 3 деп жиі қолданылады.

Математикалық өрнектелуі:

$$y_t = \sum_{j=-m}^{m} c_j x_{t-j}. \tag{4}$$

мұндағы $y_t - t-$ нүктесіндегі тегістелген мән, $x_t - t -$ нүктесіндегі бастапқы мән, c_j – конволюция коэффициенті, $m-$ тегістеу терезесінің өлшемінің жартысы, яғни отралық нүктеден $(m-1)/2$ нүкте сол және оң жағынан алынады.

Бұл әдіс бастапқы сигналдың локальды құрылымын сақтап, жиілік компоненттерін тиімді үлестіреді, әсіресе спектрлік өлшеулер, сигналдағы максимум/минимум нүктелерін сақтау маңызды болған кезде өте қолайлы.

3.5 Қалдықтарды талдау мен есептеу тиімділігі

Әрбір тегістеу әдісі үшін қалдықтар (Residual = Original − Smoothed) толық есептелді. Қалдықтардың орташа мәні, стандарттық ауытқуы, Дарбин–Уотсон статистикасы және Льюнг–Бокс критерийі арқылы қалдықтардың кездейсоқтығы нақты бағаланды (Smyl et al., 2024; Makridakis et al., 2021).

Әрбір тегістеу әдістерінің алгоритмнің орындалу уақыты Python time. perf_counter() функциясы арқылы миллисекундпен өлшенді (Liang et al., 2022; Makridakis, 2024).

4. Өңдеу қозғалтқышы (Processing Engine)

Бұл блок – қолданушының таңдаған әдісіне сәйкес сандық есептеулерді атқаратын орталық компонент болып табылады. Оның қызметтеріне бастапқы деректерді алдын ала масштабтау/нормалау, таңдалған алгоритмнің айрықша имплементациясы, параметрлік тексеру және есептеу нәтижелерінің сақталуы мен тестілеу жатады.

Компонент келесі құрамнан тұрады:

– Препроцессинг модулі: уақыттық интерполяция, масштабтау – таңдалған әдіске қатысты.

– Валидация модулі: параметр бағалаулар, мысалы Савицки–Голай әдісінде терезе өлшемі – тақ сан болуын, қарапайым экспоненциалды тегістеуде диапазоны бірден аспауын қадағалау.

– Алгоритм модульдері: әр әдіс үшін жоғарыда келтірілген формулаларға сәйкес жеке функция немесе класс құрылады.

– Тестілеу және нәтижелердің сапасын бақылау. Бағалау критерийлері RMSE, MAE, MAPE бойынша тегістелген деректердің метрикаларын салыстыру. Мысалы, RMSE критерийлері

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(x_t - \hat{x}_t)^2} \tag{5}$$

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

формуласы бойынша анықталады. Сонымен қатар, Фридман критерийі арқылы барлық тегістеу әдістері арасындағы айырмашылықтың нақты статистикалық маңыздылығы бағаланды (Demšar, 2006).

5. Нәтижелер қабаты (Output Layer)

Осылайша, тегістеу әдістерін қолданып алынған нәтижелерді қолданушыға визуализация, статистикалық метрикалар және экспорт опциялары форматында ұсынады. Нақтырақ айтқанда:

1. Бастапқы деректерді мен төрт әдіс бойынша алынған барлық мәндерді бір графикте көрнекілейді. Графиктерді салыстырып талдау жасау ыңғайлы.

2. Бастапқы мәндер, тегістелген мәндер, айырмашылық (residuals) бағандарын кесте түрінде ұсынады.

3. Статистикалық қорытындылар жасалады. Мысалы, сигнал–шу қатынасы (SNR)

$$SNR = 10\log_{10}\left(\frac{\sum \hat{x}_t^2}{\sum (x_t - \hat{x}_t)^2}\right) \quad (6)$$

есептеледі.

4. Алынған нәтижелерді *.xlsx, *.csv, *.json және *.pdf форматында тікелей экспорттауға және графиктерді *.png/svg форматында сақтауға мүмкіндік береді.

Нәтижелер және оларды талқылау.

Алдыңғы бөлімде келтірген теориялық материалдарды Python ортасында нақты практикалық және имплементация түрде жүзеге асырылды. Ол үшін бастапқы дерек ретінде Дүниежүзілік банктің ресми дерекқорынан ағымдағы баға бойынша Қазақстанның 1990–2024 жылдар арасындағы ЖІӨ–нің (жалпы ішкі өнім) статистикасы қолданылды (1–кесте).

1 кесте. Дүниежүзілік банктің ағымдағы баға бойынша Қазақстанның ЖІӨ–нің бастапқы деректер статистикасы, млрд АҚШ доллары, 1990–2024 жж.

(i)	(ii)	(i)	(ii)	(i)	(ii)	(i)	(ii)
1990	26.932	1999	16.871	2008	133.442	2017	166.806
1991	24.923	2000	18.292	2009	115.309	2018	179.340
1992	24.917	2001	22.153	2010	148.047	2019	181.667
1993	23.409	2002	24.637	2011	192.626	2020	171.082
1994	21.251	2003	30.834	2012	207.999	2021	197.112
1995	20.374	2004	43.152	2013	236.635	2022	225.496
1996	21.035	2005	57.124	2014	221.416	2023	262.642
1997	22.166	2006	81.004	2015	184.388	2024	288.406
1998	22.135	2007	104.850	2016	137.278		

(i) – жылдар, (ii) – ЖІӨ, млрд АҚШ доллары.

Ендігі ретте, 1–кестедегі бастапқы деректерді берілген 4 әдіспен тегістеу үшін 2–кестеде көрсетілген бағалау парамертлері арқылы сынақ жүргіземіз (Makridakis et al., 2022; Адамбекова т.б., 2019; Утибаев т.б., 2019).

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

2 кесте. Әдістерді бағалау параметрлері

№	Әдіс	Сыналған параметрлер	Түсіндірме
1	Қарапайым экспоненциалды тегістеу	$\alpha \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$	10% жаңа мән, 90% алдыңғы тренд ескерілген
2	Жылжымалы орташа	$window \in \{3,5,7,9\}$	Әрбір мән әртүрлі тақ жылдық орташа бойынша есептелген
3	Кездейсоқ жүріс	$\sigma \in \{0.5, 1.0, 1.5, 2.0\}$	Тегістелген, бірақ кездейсоқ қозғалыс әсері бар модель
4	Савицки–Голай	$window \in \{3,5,7,9\}$, $polyorder \in \{1,2,3\}$	Полиномдық тегістеу, терезе ұзындығы әртүрлі тақ жыл

Осылайша, тегістеу әдістерінің параметрлері алдын ала жүргізілген зертханалық эксперименттер негізінде таңдалды. Әрбір әдіс үшін бірнеше параметр комбинациялары толық қарастырылып, олардың нақты тиімділігі RMSE, MAE, MAPE және SNR көрсеткіштері (5)–(6) бойынша толық бағаланды (3–кесте).

3 кесте. Зертханалық эксперимент нәтижесі.

Әдіс	Параметрлер	RMSE	MAE	MAPE (%)	SNR (дБ)
(i)	α = 0.1	64,288	45,635	33,606	6,757
(i)	α = 0.25	40,953	27,625	22,730	10,674
(i)	α = 0.5	29,288	20,274	17,354	13,586
(i)	α = 0.75	23,973	17,213	14,843	15,325
(i)	α = 0.9	21,961	16,147	14,039	16,087
(ii)	w = 3	18,596	13,292	11,469	17,778
(ii)	w = 5	30,695	22,819	18,648	13,689
(ii)	w = 7	40,145	30,316	23,143	11,642
(ii)	w = 9	49,124	37,291	27,679	10,193
(iii)	σ = 0.5	115,970	84,279	58,923	1,633
(iii)	σ = 1.0	112,102	82,062	61,715	1,927
(iii)	σ = 1.5	108,312	79,844	64,507	2,226
(iii)	σ = 2.0	104,607	77,827	67,763	2,528
(iv)	w = 3, p = 1	7,287	4,388	4,183	25,669
(iv)	w = 3, p = 2	0,000	0,000	0,000	307,554
(iv)	w = 5, p = 1	12,022	7,611	7,616	21,320
(iv)	w = 5, p = 2	5,267	3,371	2,969	28,489
(iv)	w = 5, p = 3	5,214	3,230	2,837	28,577
(iv)	w = 7, p = 1	16,207	10,780	10,801	18,725
(iv)	w = 7, p = 2	7,996	5,133	4,715	24,862
(iv)	w = 7, p = 3	7,799	4,795	4,612	25,078
(iv)	w = 9, p = 1	19,154	13,114	13,825	17,274
(iv)	w = 9, p = 2	10,811	6,805	6,247	22,242
(iv)	w = 9, p = 3	10,925	6,960	6,212	22,151

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

(i) – қарапайым экспоненциалды тегістеу, (ii) – жылжымалы орташа, (iii) – кездейсоқ жүріс, (iv) – Савицки–Голай сүзгісі. p – полином дәрежесі, w – терезе өлшемі.

Зерттеу нәтижесін алу үшін Python ортасының келесі кітапханалар мен функциялар қолданылды:

1. NumPy – кітапханасы көпөлшемді массивтермен (array), векторлық есептеулермен, математикалық операциялармен жұмыс істеу үшін қолданылды.

2. Pandas – деректерді кестелік (DataFrame) форматта сақтау мен өңдеу үшін пайдаланылды. Бұл кітапхана арқылы бастапқы экономикалық деректер жүктеліп, индекстер, бағандар және тегістеу нәтижелері өңделді.

3. Statsmodels.tsa.holtwinters – модулінен алынған SimpleExpSmoothing класы жәй экспоненциалды тегістеу (Simple Exponential Smoothing) әдісін жүзеге асыру үшін қолданылды.

4. Scipy.signal – модулі сигналдарды өңдеуге арналған функцияларды қамтиды. Зерттеу барысында бұл модульдің savgol_filter функциясы Savitzky–Golay фильтрациясын жүзеге асыру үшін пайдаланылды.

5. Sklearn.metrics – модулінен алынған функциялар (мысалы, mean_squared_error, mean_absolute_error) модельдердің дәлдігін бағалау үшін қолданылды. Осы метрикалар арқылы RMSE, MAE, және MAPE сияқты бағалау көрсеткіштері есептелді.

6. Matplotlib.pyplot – кітапханасы тегістеу нәтижелерін визуализациялау үшін пайдаланылды.

Әрі қарай салыстыру үшін RMSE көрсеткіші ең төмен параметрлері 3–кестеден таңдалды. Зертханалық эксперименттің нәтижесінде келесі параметрлер таңдалды (4–кесте):

4 кесте. Зертханалық эксперименттің нәтижесінде таңдап алынған параметрлер

Әдіс	Параметрлер	RMSE	MAE	MAPE (%)	SNR (дБ)
(i)	α = 0.90	21,961	16,147	14,039	16,087
(ii)	w = 3	18,596	13,292	11,469	17,778
(iii)	σ = 2.0	104,607	77,827	67,763	2,528
(iv)	w = 3, p = 2	0,000	0,000	0,000	307,554

(i) – қарапайым экспоненциалды тегістеу, (ii) – жылжымалы орташа, (iii) – кездейсоқ жүріс, (iv) – Савицки–Голай сүзгісі. p – полином дәрежесі, w – терезе өлшемі.

Енді Python ортасында осы ғылыми зерттеу арналып жазылған кодтан үзінді келтірейік:

1. Қарапайым экспоненциалды тегістеу әдісінің қысқаша коды (1):

```
alpha = 0.90
ses_model = SimpleExpSmoothing(data[«ЖІӨ (млрд $)»]).fit(smoothing_level=alpha, optimized=False)
```

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

```
data[«SimpleExponential»] = ses_model.fittedvalues
```

2. Жылжымалы орташа тегістеу әдісінің қысқаша коды (2):

```
window = 3
data[«MovingAverage»] = data[«ЖІӨ (млрд $)»].rolling(window=window).mean()
```

3. Кездейсоқ жүріс тегістеу әдісінің қысқаша коды (3):

```
np.random.seed(0)
sigma = 2.0
random_walk = [data[«ЖІӨ (млрд $)»][0]]
for i in range(1, len(data)):
    random_walk.append(random_walk[–1] + np.random.normal(0, sigma))
data[«RandomWalk»] = random_walk
```

4. Савицки–Голай тегістеу әдісінің қысқаша коды (4):

```
data[«SavitzkyGolay»] = savgol_filter(data[«ЖІӨ (млрд $)»], window_length=3, polyorder=2)
```

Осы Python ортасын қолданып бастапқы деректерге әртүрлі бағалау параметрлері бойынша тестілеу мен эксперименттер жүргізілді. Бұл зертеуде тек 4–кестегі бағалау параметрлерінің нәтижелері көрсетілді. Осылайша, 4–кестеде көрсетілген бағалау парамертлерін қолданып бастапқы деректерді тегістеу нәтижесі 5–кестеде көрсетілген.

5 кесте. Бағалау параметрлерін қолданып деректерді тегістеу нәтижелері (1990–2024).

Жыл	ЖІӨ (млрд $)	Қарапайым экспоненциалды тегістеу	Жылжымалы орташа	Кездейсоқ жүріс	Савицки–Голай
1990	26.93	26,93	NaN	26,93	26,93
1991	24.92	26,93	NaN	30,46	24,92
1992	24.92	25,12	25,59	31,26	24,92
1993	23.41	24,94	24,42	33,22	23,41
1994	21.25	23,56	23,19	37,70	21,25
1995	20.37	21,48	21,68	41,43	20,37
1996	21.04	20,49	20,89	39,48	21,04
1997	22.17	20,98	21,19	41,38	22,17
1998	22.14	22,05	21,78	41,08	22,14
1999	16.87	22,13	20,39	40,87	16,87
2000	18.29	17,40	19,10	41,69	18,29
2001	22.15	18,20	19,11	41,98	22,15
2002	24.64	21,76	21,69	44,89	24,64
2003	30.83	24,35	25,87	46,41	30,83
2004	43.15	30,19	32,87	46,65	43,15
2005	57.12	41,86	43,70	47,54	57,12

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

2006	81.0	55,60	60,43	48,21	81,00
2007	104.85	78,46	80,99	51,20	104,85
2008	133.44	102,21	106,43	50,79	133,44
2009	115.31	130,32	117,87	51,41	115,31
2010	148.05	116,81	132,27	49,71	148,05
2011	192.63	144,92	151,99	44,60	192,63
2012	208.0	187,86	182,89	45,91	208,00
2013	236.63	205,98	212,42	47,64	236,63
2014	221.42	233,57	222,02	46,15	221,42
2015	184.39	222,63	214,15	50,69	184,39
2016	137.28	188,21	181,03	47,78	137,28
2017	166.81	142,37	162,82	47,87	166,81
2018	179.34	164,36	161,14	47,50	179,34
2019	181.67	177,84	175,94	50,56	181,67
2020	171.08	181,28	177,36	53,50	171,08
2021	197.11	172,10	183,29	53,81	197,11
2022	225.5	194,61	197,90	54,57	225,50
2023	262.64	222,41	228,42	52,79	262,64
2024	288.41	258,62	258,85	48,83	288,41

Әрі қарай 5–кестедегі тегістелген деректердің бағалау метрикасын жүргізілді. Тегістеу әдістерінің алынған тиімділігін бағалау үшін бірнеше стандартты статистикалық метрикалар қолданылды. Бұл метрикалар қолданылған модельдің бастапқы деректерге қаншалықты жуық нәтижелер беретінін және оның шуылға (noise) қаншалықты төзімді екенін дәлірек анықтауға мүмкіндік береді. Содан кейін орташа квадраттық қате (5) (RMSE), орташа абсолюттік қате (MAE), орташа пайыздық қате (MAPE), сигнал–шу қатынасы (SNR) параметрлері (6) бойынша бағаланды. Осы қолданған параметрлер негізінде әрбір тегістеу әдісінің тиімділігі салыстырылды, ал нәтижелері 6–кестеде берілді.

Енді 5–кестенің нәтижелері бойынша қарапайым экспоненциалды тегістеу әдісі $(DW = 0{,}944;\ p = 0{,}170)$ қалдықтарда статистикалық маңызды автокорреляцияның болмауымен ең жоғары тұрақтылық көрсетті. Ал Савицки–Голай және өзге әдістердегі $p < 0{,}05$ көрсеткіші модельдердің бастапқы деректерді жоғары дәлдікпен жуықтауына қарамастан, олардың қалдықтарында жүйелі тәуелділіктің сақталғанын айқындайды.

6–кестеде көрініп тұрғандай Савицки–Голай–дәлдік көрсеткіштері (RMSE, MAE, MAPE, SNR) бойынша ең жақсы болып табылады.

Тегістеу әдістерінің сапасын кешенді түрде саралау мақсатында қалдықтардың (residuals) статистикалық сипаттамалары Дарбин–Уотсон және Льюнг–Бокс критерийлері арқылы автокорреляция мен кездейсоқтық тұрғы-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

сынан талданды. 7–кестеде жинақталған бұл көрсеткіштер қалдықтардың статистикалық сенімділігін негіздейді.

6 кесте. Бағалау параметрлері.

Әдіс	RMSE	MAE	MAPE (%)	SNR (dB)
Жылжымалы орташа	18,596	13,292	11,469	17,778
Кездейсоқ жүріс	104,607	77,827	67,763	2,528
Савицки–Голай	0,000	0,000	0,000	307,554
Қарапайым экспоненциалды тегістеу	21,961	16,147	14,039	16,087

7 кесте. Тегістеу әдістерінің қалдықтарын статистикалық талдау нәтижелері.

Әдіс	Қалдықтардың орташа мәні	Қалдықтардың стандарттық ауытқуы	Дарбин–Уотсон статистикасы	Льюнг–Бокс р-мәні
Жылжымалы орташа	7,704	17,188	0,642	0,022
Кездейсоқ жүріс	65,121	83,060	0,042	0,000
Савицки–Голай	0,000	0,000	0,190	0,000
Қарапайым экспоненциалды тегістеу	8,206	20,667	0,944	0,170

Қарастырылған тегістеу әдістерінің есептеу жылдамдығын салыстыру мақсатында әрбір алгоритмнің орындалу уақыты Python бағдарламалау ортасында time.perf_counter() функциясы арқылы нақты өлшенді. Осылайша, есептеу уақыты миллисекундпен анықталды және әр әдіс үшін алдыңғы бөлімде таңдалған ең тиімді параметрлер қолданылды. Есептеу тиімділігінің нәтижелері 8–кестеде келтірілген.

8 кесте. Тегістеу әдістерінің есептеу тиімділігін салыстыру.

Әдіс	Ең тиімді параметрлер	Есептеу уақыты (мс)
Жылжымалы орташа	$w = 3$	1,139
Кездейсоқ жүріс	$\sigma = 2.0$	0,204
Савицки–Голай	$w = 3$, $p = 2$	1,155
Қарапайым экспоненциалды тегістеу	$\alpha = 0.90$	1,496

p – полином дәрежесі, w – терезе өлшемі.

8–кестеден байқағанымыздай жылжымалы орташа, Савицки–Голай сүзгісі және қарапайым экспоненциалды тегістеу әдістерінің орындалу уақыттары сәйкесінше 1,139 мс, 1,155 мс және 1,496 мс болды. Барлық әдістердің есептеу уақыты 2 мс-тан төмен болғандықтан, олардың есептеу күрделілігі қарастырылған деректер көлемі үшін өте төмен деп бағаланды. Сондықтан әдісті таңда-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

уда есептеу уақыты емес, тегістеу сапасын сипаттайтын RMSE, MAE, MAPE және SNR көрсеткіштері негізгі критерий ретінде пайдаланылды.

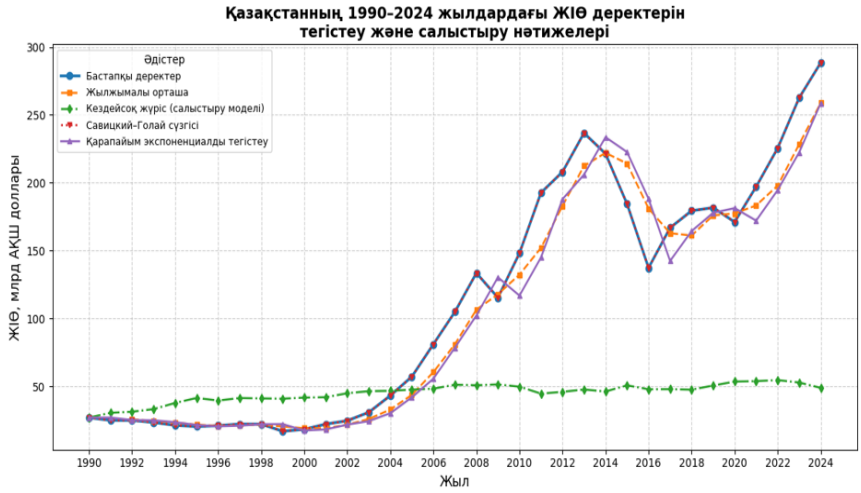
Әдістер арасындағы айырмашылықтың статистикалық маңыздылығын бағалау үшін Фридман критерийі қолданылды. Бұл критерий бірнеше әдістің нәтижелерін бірдей деректер жиынында салыстыруға мүмкіндік беретін параметрлік емес статистикалық тәсіл болып табылады (Demšar, 2006). Фридман тестінің нәтижелері 9–кестеде берілген.

9 кесте. Фридман критерийі бойынша әдістерді статистикалық салыстыру нәтижелері.

Көрсеткіш	Мәні
Фридман статистикасы	85,6182
p-мәні	0
Қорытынды	Әдістер арасындағы айырмашылық статистикалық маңызды

9–кестеден көрініп тұрғандай, Фридман статистикасының мәні 85,6182, ал p-мәні 0 болды. Яғни, бұл әдістер арасындағы айырмашылықтың статистикалық тұрғыдан маңызды екенін көрсетеді. Демек, қарастырылған тегістеу әдістерінің нәтижелері бірдей емес.

Келесі де Python ортасын қолданып бастапқы деректер мен тегістелген деректерді көрнекілейік (визуализация), ал нәтижесі график түрінде 2–суретте көрсетілген.



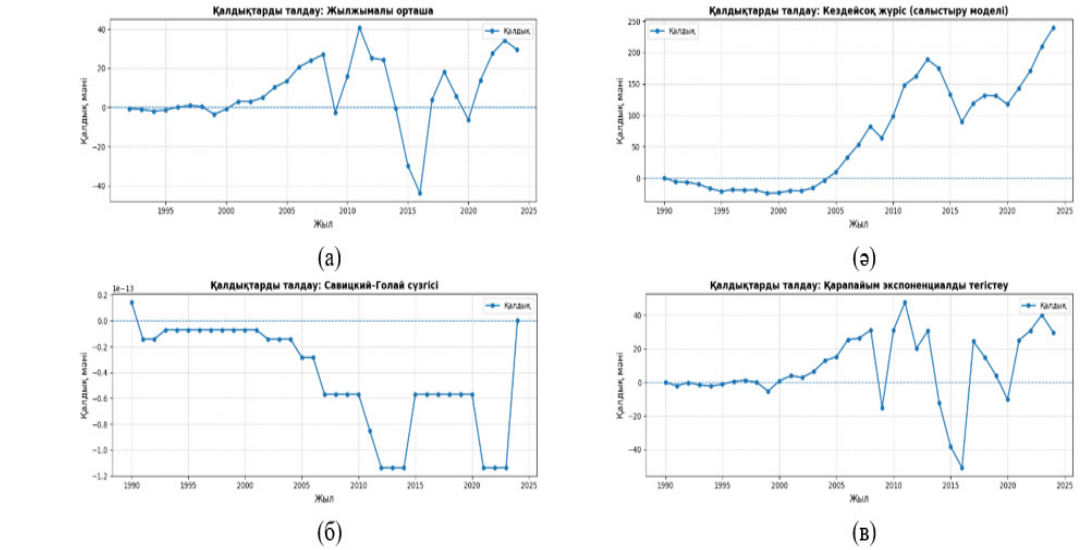
Сур. 2. ЖІӨ уақыт қатарын тегістеу әдістерінің графиктері

Бұл модулде тегістелген деректердің графигімен қатар соңғы нәтижені *.csv, *.xlsx, *.JSON, *.pdf – форматтарында көшіріп алуға мүмкіндігі бар.

Төмендегі 3–суретте жылжымалы орташа, кездейсоқ жүріс, Савицки–Голай сүзгісі және қарапайым экспоненциалды тегістеу әдістері бойынша бастапқы және тегістелген мәндер арасындағы қалдықтардың жылдар бойынша өз-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

геріс динамикасы көрсетілген.



Сур. 3. Қарастырылған тегістеу әдістері бойынша қалдықтардың динамикасы:
(а) – Жылжымалы орташа әдісі бойынша қалдықтар;
(ә) – Кездейсоқ жүріс негізіндегі салыстыру моделі бойынша қалдықтар;
(б) – Савицкий–Голай сүзгісі бойынша қалдықтар;
(в) – Қарапайым экспоненциалды тегістеу әдісі бойынша қалдықтар.

Қорытынды.

Жүргізілген ғылыми зерттеу нәтижесінде Қазақстанның 1990–2024 жылдар аралығындағы ЖІӨ бастапқы деректерін тегістеу үшін қарапайым экспоненциалды тегістеу, жылжымалы орташа, Савицки–Голай сүзгісі және кездейсоқ жүріс негізіндегі базалық салыстыру моделі қолданылды. Әдістердің тиімділігі RMSE, MAE, MAPE, SNR көрсеткіштері мен қалдықтарды талдау және есептеу уақыты арқылы нақты бағаланды.

Қолданылған параметрлерді салыстыру нәтижесінде қарапайым экспоненциалды тегістеу үшін $\alpha = 0.9$, жылжымалы орташа үшін терезе = 3, кездейсоқ жүріс моделі үшін $\sigma = 2.0$, ал Савицки–Голай сүзгісі үшін терезе = 3, полином дәрежесі = 2 ең тиімді параметрлер ретінде зертханалық эксперимент арқылы анықталды. Сонымен қатар, дәлдік метрикалары бойынша Савицки–Голай сүзгісі ең төмен қателікті көрсетті (RMSE ≈ 0, SNR = 307.55 дБ). Алайда бұл нәтиже таңдалған параметрлердің бастапқы деректерді толық дерлік қайта қалпына келтіруімен байланысты болғандықтан, оны нақты тегістеу әсері тұрғысынан сақтықпен түсіндіру қажет деп санаймыз.

Ендігі ретте, жылжымалы орташа әдісі тұрақты және түсіндіруге ыңғайлы нәтиже көрсетті (RMSE = 18.60, MAE = 13.29, MAPE = 11.47%, SNR = 17.78 дБ). Ал қарапайым экспоненциалды тегістеу әдісі (RMSE = 21.96) қалдықтардағы автокорреляция тұрғысынан салыстырмалы түрде тұрақты болып шықты,

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

себебі Льюнг–Бокс тестінің p-мәні 0.170 болып, 0.05 деңгейінен жоғары шықты. Кездейсоқ жүріс моделі жоғары қателік көрсеткендіктен, ал бұл ғылыми жұмыста тек базалық салыстыру моделі ретінде қарастырылды.

Алайда есептеу тиімділігі бойынша барлық әдістер дерлік 2 мс-тан төмен уақытта орындалды, сондықтан қарастырылған бастапқы деректер көлемі үшін олардың есептеу күрделілігі төмен деп бағаланды. Жалпы алғанда, ұсынылған модуль бірнеше тегістеу әдісін бір архитектурада біріктіріп, параметрлерді икемді таңдауға, нәтижелерді салыстыруға, қалдықтарды талдауға және визуализациялауға мүмкіндік береді (Demšar, 2006; Ljung et al., 1978). Болашақта бұл архитектураны интеллектуалды ақпараттық жүйенің веб-модулі ретінде енгізу және әртүрлі экономикалық уақыт қатарларында қосымша сынақтан өткізу жоспарланады.

REFERENCES

Adambekova, A. A., & Omir, A. Zh. (2019). Stock Market Indices: Analysis and Features of Application // Central Asian Economic Review. No. 5. — Pp. 54–66. [In Kaz.].

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets // Journal of Machine Learning Research. — Vol. 7. — Pp. 1–30. [In Eng.].

Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions // International Journal of Forecasting. — Vol. 37. — No. 1. — Pp. 388–427. [In Eng.].

Hening, A., Tran, K. Q., & Ungureanu, S. C. (2022). The effects of random and seasonal environmental fluctuations on optimal harvesting and stocking // Journal of Mathematical Biology. — Vol. 84. — No. 6. — Pp. 41. [In Eng.].

Kaneko, H., & Funatsu, K. (2015). Smoothing–combined soft sensors for noise reduction and improvement of predictive ability // Industrial & Engineering Chemistry Research. — Vol. 54. — No. 50. — Pp. 12630–12638. [In Eng.].

Liang, H., Zhao, X., & Guo, Y. (2022). Noise Reduction Method for the Ring LaserGyro Signal Based on Ceemdan and the Savitzky–Golay Algorithm // Fluctuation and Noise Letters. — Vol. 21. — No. 01. Article 2250005. [In Eng.].

Long, X., Schmidt, D. F., Bergmeir, C., & Smyl, S. (2024). Fast Gibbs sampling for the local and global trend Bayesian exponential smoothing model // International Journal of Forecasting / preprint version. [In Eng.].

Liang, H., Zhao, X., & Guo, Y. (2022). Noise reduction method for the ring laser gyro signal based on CEEMDAN and the Savitzky–Golay algorithm // Fluctuation and Noise Letters. — Vol. 21. — No. 01. Article 2250005. [In Eng.].

Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. Biometrika. — Vol. 65. — No. 2. — Pp. 297–303. [In Eng.].

Makridakis, S., Fry, C., Petropoulos, F., & Spiliotis, E. (2021). The future of forecasting competitions: Design attributes and principles // International Journal of Forecasting. [In Eng.].

Makridakis, S., Spiliotis, E., Assimakopoulos, V., Chen, Z., Gaba, A., Tsetlin, I., & Winkler, R. L. (2022). The M5 uncertainty competition: Results, findings and conclusions // International Journal of Forecasting. — Vol. 38. — No. 4. — Pp. 1365–1385. [In Eng.].

Makridakis, S. (2024). The M6 forecasting competition: Bridging the gap between forecasting and investment decisions // International Journal of Forecasting. [In Eng.].

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). The M5 accuracy competition: Results, findings and conclusions // International Journal of Forecasting. — Vol. 38. — No. 4. — Pp. 1346–1364. [In Eng.].

Önder, E., Bayır, F., & Hepsen, A. (2013). Forecasting macroeconomic variables using artificial neural network and traditional smoothing techniques // Journal of Applied Finance and Banking. — Vol. 3. — No. 4. — Pp. 73–104. [In Eng.].

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Ben Taieb, S., et al. (2022). Forecasting: theory and practice // International Journal of Forecasting. — Vol. 38. — No. 3. — Pp. 705–871. [In Eng.].

Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

procedures // Analytical Chemistry. — Vol. 36. — No. 8. — Pp. 1627–1639. [In Eng.].

Schmid, M., Rath, D., & Diebold, U. (2022). Why and how Savitzky–Golay filters should be replaced // ACS Measurement Science Au. — Vol. 2. — No. 2. — Pp. 185–196. [In Eng.].

Swari, M. H. P., Handika, I. P. S., & Satwika, I. K. S. (2021, October). Comparison of simple moving average, single and modified single exponential smoothing. In 2021 IEEE 7th Information Technology International Seminar (ITIS) (pp. 1-5). IEEE. [In Eng.].

Smyl, S., Bergmeir, C., Dokumentov, A., Long, X., Wibowo, E., & Schmidt, D. F. (2024). Local and global trend Bayesian exponential smoothing models. International Journal of Forecasting. [In Eng.].

Turarov, D. R., & Issayeva, A. A. (2023). Financial Forecasting Using Stochastic Models: Evidence from the Kazakhstan Commodity Exchange. Journal of Economic Research & Business Administration. — Vol. 146. — No. 4. [In Kaz.].

Utibayev, B. S., & Utibayeva, G. B. (2019). Analysis of the Sustainability of Grain Production Using Statistical Methods. Problems of AgriMarket. No. 1. Pp. 129–135. [In Russ.].

Woo, G., Liu, C., Sahoo, D., Kumar, A., & Hoi, S. C. H. (2022). ETSformer: Exponential smoothing transformers for time-series forecasting. Proceedings of the 39th International Conference on Machine Learning (ICML). [In Eng.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 30–45
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.002
УДК 519.7.004

APPLICATION OF MACHINE LEARNING AND DEEP LEARNING FOR PREDICTIVE PLANNING AND RESOURCE EFFICIENCY IN PRODUCTION SYSTEMS

A.T. Iskakova[1]*, G.E. Kurmankulova[1], K. Gromaszek[2], K.O. Togzhanova[3], R.S. Algozhayeva[3]

[1]Al-Farabi Kazakh National University, Almaty, Kazakhstan;
[2]Lublin University of Technology, Lublin, Poland;
[3]Almaty Technological University, Almaty, Kazakhstan.
E-mail: iskakova.1977@mail.ru

Ainur Iskakova — 2nd-year doctoral student of the Department of Artificial Intelligence and Big Data, Al-Farabi Kazakh National University, Almaty, Kazakhstan
E-mail: iskakova.1977@mail.ru, https://orcid.org/0009-0008-4138-2240;
Gulzhan Kurmankulova — Candidate of Pedagogical Sciences, Acting Associate Professor of the Department of Artificial Intelligence and Big Data, Al-Farabi Kazakh National University, Almaty, Kazakhstan
https://orcid.org/0000-0003-1172-2394;
Konrad Gromaszek — Doctor of Technical Sciences, Professor, Lublin University of Technology, Lublin, Poland
https://orcid.org/0000-0002-3265-3714;
Kulzhan Togzhanova — PhD, Associate Professor, Head of the Department of Information Systems, Almaty Technological University, Almaty, Kazakhstan
https://orcid.org/0000-0003-1962-844X;
Raikhan Algozhayeva — Senior Lecturer of the Department of Information Systems, Almaty Technological University, Almaty, Kazakhstan
https://orcid.org/0000-0003-1575-5425.

© A.T. Iskakova, G.E. Kurmankulova, K. Gromaszek, K.O. Togzhanova, R.S. Algozhayeva

Abstract. The digitalization of manufacturing enterprises, combined with the integration of intelligent shop-floor devices and software systems into manufacturing execution systems (MES), has led to a sharp increase in available data. The extent to which enterprises can extract valuable information from big data and gain actionable insights is a decisive factor in developing management systems that optimize production and conserve resources. Machine learning and big data technologies are becoming increasingly widespread, finding applications in several critical areas of

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

planning and management. Cloud manufacturing enables the real-time deployment of these technologies, reducing implementation and deployment costs. In this regard, this paper considers the application of a machine learning approach to ensure information transparency and optimization of an object within a cloud environment. The primary objective of the work is to optimize energy consumption in production systems indirectly through efficient resource management and predictive modeling of equipment technological parameters. In the course of the study, hybrid models based on LSTM and CNN+LSTM are proposed for time-series forecasting. The proposed approach allows for the prevention of emergency shutdowns by predicting equipment overheating in advance, thereby reducing energy costs and increasing planning accuracy. Experimental results demonstrate the efficiency of the model and prove the prospects of its application in a real production environment.

Keywords: Cloud Manufacturing (CMfg), cloud computing, production processes, IoT (Internet of Things), Big Data, Machine Learning, Deep Learning, LSTM, CNN+LSTM, predictive maintenance, intelligent systems, data analytics, real-time monitoring, Digital Twin, production optimization

For citation: Iskakova A.T., Kurmankulova G.E., Gromaszek K., Togzhanova K.O., Algozhayeva R.S. (2026). Application of machine learning and deep learning for predictive planning and resource efficiency in production systems // International journal of information and communication technologies. Vol. 7. No. 27. Pp. 30–45. https://doi.org/10.54309/IJICT.2026.27.3.002. (In Kaz.).

Conflict of interest: The authors declare that there is no conflict of interest.

ӨНДІРІС ЖҮЙЕЛЕРІНДЕ БОЛЖАМДЫ ЖОСПАРЛАУ ЖӘНЕ РЕСУРСТАРДЫ ҮНЕМДЕУ ҮШІН МАШИНАЛЫҚ ЖӘНЕ ТЕРЕҢ ОҚЫТУДЫ ҚОЛДАНУ

А.Т. Искакова[1]*, Г.Е. Курманкулова[1], К. Громашек[2], К.О. Тогжанова[3], Р.С. Алгожаева[3]

[1]Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан;
[2]Люблин технологиялық университеті, Люблин, Польша;
[3]Алматы технологиялық университеті, Алматы, Қазақстан.
E-mail: iskakova.1977@mail.ru

Искакова Айнур — Әл-Фараби атындағы Қазақ ұлттық университеті, «Жасанды интеллект және Big Data» кафедрасының 2-курс докторанты, Алматы, Қазақстан
E-mail: iskakova.1977@mail.ru, https://orcid.org/0009-0008-4138-2240;
Курманкулова Гульжан — п.ғ.к., доцент м.а., Әл-Фараби атындағы Қазақ ұлттық университеті, «Жасанды интеллект және Big Data» кафедрасы, Алматы, Қазақстан
https://orcid.org/0000-0003-1172-2394;
Громашек Конрад — т.ғ.д., профессор, Люблин технологиялық университеті, Люблин, Польша

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

https://orcid.org/0000-0002-3265-3714;
Тогжанова Кульжан — PhD, қауымдастырылған профессор, «Ақпараттық жүйелер» кафедрасының меңгерушісі, Алматы технологиялық университеті, Алматы, Қазақстан
https://orcid.org/0000-0003-1962-844X;
Алғожаева Райхан — «Ақпараттық жүйелер» кафедрасының сеньор-лекторы, Алматы технологиялық университеті, Алматы, Қазақстан
https://orcid.org/0000-0003-1575-5425.

© А.Т. Искакова, Г.Е. Курманкулова, К. Громашек, К.О. Тогжанова, Р.С. Алгожаева

Аннотация. Өндіріс орындарын цифрландыру және интеллектуалды цех құрылғылары мен бағдарламалық жүйелерді өндірісті басқару жүйелерінде (ӨБЖ) бірлесе қолдану, қолжетімді деректердің күрт өсуіне әкелді. Кәсіпорындардың үлкен деректерден құнды ақпараттарды алу және пайдалы түсініктеге қол жеткізу дәрежесі өндірісті оңтайландыратын және ресурстарды қорғайтын басқару жүйелерін құруда шешуші фактор болып табылады. Машиналық оқыту және үлкен деректер технологиялары барған сайын кең таралуда және жоспарлау мен басқарудың бірнеше маңызды салаларында қолданылуда. Бұлтты өндіріс бұл технологияларды нақты уақыт режимінде қолдануға мүмкіндік береді, енгізу және орналастыру шығындарын азайтады. Осыған байланысты, бұл мақалада бұлтты ортада объект туралы ақпараттарды және оны оңтайландыруды қамтамасыз ету үшін машиналық оқыту тәсілін қолдану қарастырылады. Жұмыстың негізгі мақсаты - өндірістік орталерде ресурстарды тиімді басқару және жабдықтардың технологиялық параметрлерін предиктивті модельдеу арқылы жаңама түрде энергия тұтынуды оңтайландыру болып табылады. Зерттеу барысында уақыттық қатарларды болжауға арналған LSTM және CNN+LSTM негізіндегі гибридті модельдер ұсынылады. Ұсынылған тәсіл жабдықтың қызып кетуін алдын ала болжай отырып, апаттық тоқтауларды болдырмауға, энергия шығынын азайтуға және жоспарлау дәлдігін арттыруға мүмкіндік береді. Эксперименттік нәтижелер модельдің тиімділігін көрсетеді және оны нақты өндірістік ортада қолданудың перспективаларын дәлелдейді.

Түйінді сөздер: Cloud Manufacturing (CMfg), бұлтты есептеу, өндірістік процестер, IoT (Internet of Things), үлкен деректер (Big Data), машиналық оқыту (Machine Learning), терең оқыту (Deep Learning), LSTM, CNN+LSTM, predictive maintenance, интеллектуалды жүйелер, деректерді талдау, нақты уақыттағы мониторинг, цифрлық егіз (Digital Twin), өндірісті оңтайландыру

Дәйексөздер үшін: Искакова А.Т., Курманкулова Г.Е., Громашек К., Тогжанова К.О., Алгожаева Р.С. (2026). Өндіріс жүйелерінде болжамды жоспарлау және ресурстарды үнемдеу үшін машиналық және терең оқытуды қолдану // Халықаралық ақпараттық және коммуникациялық технологиялар журналы. Т. 7. No. 27. Б. 30–45. https://doi.org/10.54309/IJICT.2026.27.3.002. (Қаз. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

жоқ деп мәлімдейді.

ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО И ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗНОГО ПЛАНИРОВАНИЯ И РЕСУРСОСБЕРЕЖЕНИЯ В ПРОИЗВОДСТВЕННЫХ СИСТЕМАХ

А.Т. Искакова[1]*, Г.Е. Курманкулова[1], К. Громашек[2], К.О. Тогжанова[3], Р.С. Алгожаева[3]

[1]Казахский национальный университет имени аль-Фараби, Алматы, Казахстан;
[2]Люблинский технологический университет, Люблин, Польша;
[3]Алматинский технологический университет, Алматы, Казахстан.
E-mail: iskakova.1977@mail.ru

Искакова Айнур — докторант 2 курса кафедры «Искусственный интеллект и Big Data», Казахского национального университета имени аль-Фараби, Алматы, Казахстан
E-mail: iskakova.1977@mail.ru, https://orcid.org/0009-0008-4138-2240;
Курманкулова Гульжан — к.п.н., и.о. доцента кафедры «Искусственный интеллект и Big Data», Казахского национального университета имени аль-Фараби, Алматы, Казахстан
https://orcid.org/0000-0003-1172-2394;
Громашек Конрад — д.т.н, профессор, Люблинский технологический университет, Люблин, Польша
https://orcid.org/0000-0002-3265-3714;
Тогжанова Кульжан — PhD, ассоциированный профессор, заведующая кафедрой информационных систем, Алматинский технологический университет, Алматы, Казахстан
https://orcid.org/0000-0003-1962-844X;
Алгожаева Райхан — сеньор-лектор кафедры «Информационные системы», Алматинский технологический университет, Алматы, Казахстан
https://orcid.org/0000-0003-1575-5425.

© А.Т. Искакова, Г.Е. Курманкулова, К. Громашек, К.О. Тогжанова, Р.С. Алгожаева

Аннотация. Цифровизация производственных предприятий и совместное использование интеллектуальных цеховых устройств и программных систем в системах управления производством (СУП) привели к резкому росту объёма доступных данных. Степень способности предприятий извлекать ценные знания из больших данных и получать полезные инсайты является ключевым фактором при построении систем управления, направленных на оптимизацию производства и эффективное использование ресурсов. Технологии машинного обучения и больших данных находят всё более широкое применение в различных областях планирования и управления. Облачное производство (Cloud Manufacturing)

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

позволяет использовать эти технологии в режиме реального времени, сокращая затраты на внедрение и развертывание. В связи с этим в данной статье рассматривается применение подхода машинного обучения для обеспечения информационной прозрачности объекта и его оптимизации в облачной среде. Основная цель работы заключается в оптимизации энергопотребления в производственных системах косвенным путем – посредством эффективного управления ресурсами и предиктивного моделирования технологических параметров оборудования. В ходе исследования для прогнозирования временных рядов предложены гибридные модели на основе LSTM и CNN+LSTM. Предложенный подход позволяет предотвратить аварийные остановки за счет заблаговременного прогнозирования перегрева оборудования, снизить затраты энергии и повысить точность планирования. Экспериментальные результаты демонстрируют эффективность модели и доказывают перспективы ее применения в реальной производственной среде.

Ключевые слова: Cloud Manufacturing (CMfg), облачные вычисления, производственные процессы, IoT, большие данные, машинное обучение, глубокое обучение, LSTM, CNN+LSTM, predictive maintenance, интеллектуальные системы, анализ данных, мониторинг в режиме реального времени, цифровой двойник, оптимизация производства

Для цитирования: Искакова А.Т., Курманкулова Г.Е., Громашек К., Тогжанова К.О., Алгожаева Р.С. (2026). Применение методов машинного и глубокого обучения для прогнозного планирования и ресурсосбережения в производственных системах // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 27. Стр. 30–45. https://doi.org/10.54309/IJICT.2026.27.3.002. (На каз.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Кіріспе.

Қазіргі заманғы өнеркәсіптік жүйелердің күрделене түсуі өндірістік ресурстарды басқарудың интеллектуалды тәсілдерін енгізуді талап етеді. Әсіресе энергия тұтынуды тиімді жоспарлау өндірістік шығындарды азайтып, технологиялық жабдықтардың сенімді жұмысын қамтамасыз етудің негізгі факторларының бірі болып табылады. Жабдықтардың электр энергиясын артық тұтынуы көбіне олардың механикалық тозуына, өндірістік жүктеменің теңгерімсіз бөлінуіне және технологиялық режимдердің бұзылуына байланысты болады. Осыған байланысты үлкен деректерді талдау және машиналық оқыту әдістерін қолдану өндірістік процестердегі жасырын заңдылықтарды анықтап, ресурстарды басқару тиімділігін арттыруға мүмкіндік береді.

Соңғы жылдары өндірістік процестерді интеллектуалды басқаруда Machine Learning және Deep Learning әдістері кеңінен қолданылып келеді. Әсіресе уақыттық қатарларды өңдеуге негізделген LSTM және CNN+LSTM модельдері өндірістік жабдықтардың күйін болжау, энергия тұтынуды талдау және ақауларды ерте анықтау міндеттерінде жоғары тиімділік көрсетуде. Сонымен қатар

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

бұлттық есептеулердің (Cloud Manufacturing) және IoT технологияларының дамуы өндірістік жабдықтардан нақты уақыт режимінде үлкен көлемдегі деректерді жинауға және өңдеуге мүмкіндік береді. Дегенмен қолданыстағы зерттеулердің басым бөлігі жеке міндеттерді ғана қарастырып, ресурстарды басқару, энергия тиімділігі және жабдық күйін болжауды бірыңғай интеллектуалды жүйеде біріктіруге жеткілікті көңіл бөлмейді.

Осы зерттеудің мақсаты–өндірістік ортадағы технологиялық жабдықтардың температуралық режимін болжау негізінде ресурстарды тиімді басқаруға арналған интеллектуалды тәсілді әзірлеу. Ұсынылған әдіс IoT сенсорларынан алынатын үлкен деректерді, Cloud Manufacturing архитектурасын және LSTM/CNN+LSTM негізіндегі терең оқыту модельдерін біріктіреді. Бұл өндірістік ресурстарды тиімді жоспарлауға, энергия тұтынуды оңтайландыруға, жабдықтардың жұмыс режимін бақылауға және ықтимал ақауларды алдын ала анықтауға мүмкіндік береді.

Жұмыстың ғылыми жаңалығы Cloud Manufacturing, IoT және Big Data технологияларын терең оқыту модельдерімен кешенді біріктіру арқылы өндірістік ресурстарды интеллектуалды басқарудың жаңа тәсілін ұсынуында жатыр. Ұсынылған тәсіл нақты уақыт режимінде өндірістік деректерді өңдеп, жабдықтардың температуралық режимін жоғары дәлдікпен болжайды, ресурстарды адаптивті түрде қайта жоспарлауға және болжамды техникалық қызмет көрсету (Predictive Maintenance) механизмдерін жүзеге асыруға мүмкіндік береді. LSTM моделі уақыттық қатарлардағы ұзақ мерзімді тәуелділіктерді тиімді анықтайтындықтан негізгі модель ретінде таңдалды, ал GRU, ARIMA, XGBoost және CNN+LSTM модельдері салыстырмалы бағалау үшін пайдаланылды.

Зерттеу барысында өнеркәсіптік жабдықтарға орнатылған IoT сенсорлары мен SCADA/MES өндірістік жүйелерінен жиналған нақты өндірістік деректер пайдаланылды. Эксперименттер жүргізу үшін осы деректер негізінде арнайы деректер жиынтығы қалыптастырылды. Жиналған ақпарат өндірістік процестердің нақты жұмыс режимдерін сипаттайды және ұсынылған модельдерді оқыту мен олардың тиімділігін бағалау үшін қолданылды.

Зерттеудің белгілі бір шектеулері бар. Ұсынылған модель нақты өндірістік деректер негізінде әзірленіп, тексерілгенімен, алынған нәтижелер қолданылған жабдықтың техникалық ерекшеліктері мен өндірістік процестің сипатына байланысты өзгеруі мүмкін. Сондықтан бұл тәсілді басқа өндіріс салаларында қолдану кезінде модельді қайта оқыту немесе жаңа деректерге бейімдеу қажет болуы ықтимал. Сонымен қатар болжау дәлдігі оқыту үшін пайдаланылған деректердің көлеміне, сапасына және толықтығына тікелей тәуелді. Сенсорлардан келетін деректердегі шу, жетіспейтін мәндер немесе өндірістік процестегі күтпеген өзгерістер модельдің жұмыс тиімділігіне белгілі бір деңгейде әсер етуі мүмкін. Осыған байланысты алдағы зерттеулерде әртүрлі өндірістік кәсіпорындардың деректерін пайдалану, модельдің тұрақтылығы мен сенімділігін арттыру, сондай-ақ оның әртүрлі өндірістік жағдайларда жалпылау мүмкіндігін жан-жақты бағалау жоспарланып отыр.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Зерттеудің негізгі үлесі нақты уақыт режимінде өндірістік деректерді өңдеуге негізделген және үлкен деректер мен машиналық оқыту алгоритмдерін біріктіретін гибридті басқару тәсілін әзірлеу болып табылады.

Әдістер мен материалдар.

Зерттеу барысында өндірістік ортадағы ресурстарды ұтымды пайдалану және энергия тиімділігін арттыру мақсатында технологиялық жабдықтардың температурасын болжау міндеті қарастырылды. Кіріс деректері ретінде келесі параметрлер қолданылды: ағымдағы температура, өндірістік жүктеме, уақыт факторы және өндіріс көлемі.

Уақыттық қатарларды өңдеу үшін рекуррентті нейрондық желі – LSTM және гибридті модельдер өндірістік орталерде тиімді қолданылуда (Wang et al., 2020; Han et al., 2020). Бұл модельдер ұзақ мерзімді тәуелділіктерді тиімді анықтауға мүмкіндік береді.

Модельдің жалпы математикалық көрінісі төмендегідей:

$$y_t = f(x_t, x_{t-1}), \dots, x_{t-n}$$

мұндағы $\mathbf{x_t}$ – кіріс параметрлері, ал y_t – модельдің шығысы, яғни алдағы уақыт мезетіндегі болжанатын жабдық температурасы (°C).

Эксперименттік зерттеу жүргізу үшін өндірістік ортадағы технологиялық жабдықтардың SCADA/MES жүйелерінен және IoT сенсорларынан алынған мәліметтер негізінде арнайы деректер жиынтығы қалыптастырылды. Оның құрамына температура, вибрация көрсеткіштері, өндірістік тораптардың жүктемесі, энергия тұтыну деңгейі және уақыттық параметрлер енгізілді. Жалпы алғанда, деректер жиынтығы 12 000–нан астам уақыттық жазбаларды қамтыды.

Мәліметтерді жинау процесі алты ай бойы бес минуттық интервалмен жүргізілді. Ақпарат нақты уақыт режимінде өндірістік ортада орнатылған интеллектуалды сенсорлық модульдер мен бұлттық мониторинг платформасы арқылы алынды. Мониторинг жүргізу үшін DS18B20 және PT100 RTD температура сенсорлары, BMP280 қысым датчигі, ADXL345 акселерометрі, DHT22 ылғалдылық сенсоры және PZEM–004T энергия тұтынуын бақылау модулі пайдаланылды.

Сенсорлардың өлшеу диапазоны, дәлдігі және деректерді беру жиілігі зерттеу талаптарына сәйкес таңдалып, өндірістік жағдайды жоғары дәлдікпен сипаттауға мүмкіндік берді.

Өндірістік ортаның ерекшеліктері де ескерілді: үздіксіз жұмыс режимі, жабдық жүктемесінің өзгермелілігі, көпфакторлы технологиялық процесс және сыртқы орта әсерлері. Осы факторлар алынған деректердің күрделілігін арттырып, модельдің нақты өндірістік жағдайға бейімделуін қамтамасыз етті.

Алдын ала өңдеу кезеңінде деректер сапасын арттыруға бағытталған бірнеше қадам жүзеге асырылды. Атап айтқанда, жетіспейтін мәндер жойылып, кездейсоқ шуылдар сүзгіден өткізілді. Сонымен қатар Min-Max scaling әдісі арқылы нормализация жүргізіліп, LSTM моделіне қажетті уақыттық терезелер қалып-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

тастырылды. Ағындық деректерді агрегаттау да осы кезеңде орындалды. Бұл қадамдар модельдің тұрақты жұмысын қамтамасыз етіп, болжау дәлдігін арттыруға ықпал етті.

Зерттеудің сенімділігі мен қайта өндірілуін қамтамасыз ету үшін деректер 80% оқыту және 20% тестілеу жиынтықтарына бөлінді, бастапқы кездейсоқ мән (random seed) бастапқы мәндері бекітіліп, оқыту және тестілеу жиынтықтарына бөлу (train/test split) схемасы тұрақты түрде қолданылды және 5-fold cross-validation әдісі пайдаланылды. Барлық модельдер бірдей аппараттық ресурстар негізінде оқытылып, параметрлердің біріздігі сақталды.

LSTM моделінің параметрлері келесідей таңдалды: екі жасырын қабат, сәйкесінше 64 және 32 нейрон, белсендіру функциясы ретінде ReLU, оңтайландыру алгоритмі ретінде Adam қолданылды. Оқыту жылдамдығы (learning rate) мәні 0,001-ге, пакет өлшемі (batch size) 32–ге, ал оқыту эпохаларының саны 100-ге тең етіп орнатылды. Модельді оқыту барысында шығын функциясы ретінде орташа квадраттық қате (Mean Squared Error, MSE) қолданылды.

CNN+LSTM моделінде Conv1D сверткіш қабаты мен MaxPooling1D қабаты қолданылды. LSTM қабатындағы нейрондар саны 64–ке тең болды, dropout коэффициенті 0,2 ретінде таңдалды. Бұл модельде де оңтайландыру үшін Adam алгоритмі пайдаланылды.

Ұсынылған модельдердің тиімділігін бағалау үшін салыстырмалы талдау жүргізілді. Зерттеу барысында ұсынылған тәсіл ARIMA, XGBoost, GRU және CNN+LSTM сияқты кең таралған базалық алгоритмдермен салыстырылды. Салыстыру нәтижелері модельдердің болжау дәлдігін, тұрақтылығын және өндірістік процестерді талдау тиімділігін анықтауға мүмкіндік берді. Нәтижелер төмендегі кестеде көрсетілген.

Кесте 1. Модельдердің жабдық температурасын болжау дәлдігін салыстыру нәтижелері

Модель	MAE	RMSE	R^2
ARIMA	3.84	4.91	0.79
XGBoost	2.17	2.94	0.88
GRU	1.26	1.73	0.93
CNN+LSTM	0.97	1.51	0.95
Ұсынылған LSTM	0.84	1.42	0.96

Алынған нәтижелер ұсынылған модельдің дәстүрлі статистикалық әдістермен және басқа ML/DL модельдерімен салыстырғанда болжау дәлдігінің жоғары екенін көрсетті. Бұл оның өндірістік ортадағы технологиялық жабдықтардың температуралық режимін сенімді болжауға жарамды екенін дәлелдейді.

Нәтижелердің статистикалық тұрақтылығы модельді бірнеше рет қайта оқыту және кросс-валидация әдістерін қолдану арқылы тексерілді. Эксперимент қорытындылары бойынша нәтижелердің орташа ауытқуы 3–5% аралығынан аспады. Бұл ұсынылған тәсілдің сенімділігін, тұрақтылығын және практикалық

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

қолдануға жарамдылығын көрсетеді.

Болжамды техникалық қызмет көрсету міндетін жүзеге асыру мақсатында технологиялық жабдық температурасының алдын ала белгіленген шекті мәннен (осы зерттеуде – 80°C) жоғары көтерілуі аномалиялық жағдай ретінде қабылданды. Ұсынылған модель мұндай аномалияларды алдын ала анықтауға және техникалық қызмет көрсету қажеттілігі туралы ескерту қалыптастыруға мүмкіндік береді.

Нәтижелер және оларды талқылау.

Жүргізілген эксперименттер нәтижесінде ұсынылған модель өндірістік орталардағы жабдық күйін жоғары дәлдікпен болжай алатыны анықталды. Осы технологиялық болжамға сүйене отырып, өндірісті басқаруға қатысты бірнеше тиімді шешімдер қабылданды. Атап айтқанда, жабдықтың қызып кету қаупі алдын ала анықталып, өндірістік жүктеме уақыт бойынша қайта бөлініп, осылайша энергияны көп тұтынатын жабдықтар оңтайлы жұмыс режиміне көшірілді.

Энергияны үнемдеу көрсеткіші болжамды басқару жүйесі қолданылған және қолданылмаған жағдайдағы жабдықтың жалпы энергия тұтынуын салыстыру арқылы анықталды. Ұсынылған LSTM моделі жабдықтың қызып кету ықтималдығын алдын ала болжауға мүмкіндік берді. Соның нәтижесінде өндірістік жүктеме тиімді жоспарланып, технологиялық операциялар оңтайлы бөлінді, шектен тыс жүктемелер азайып, апатты тоқтаулардың алдын алу мүмкін болды. Алты айлық тәжірибе барысында энергия тұтынудың жабдықтың жұмыс режиміне қарай 10–25 % төмендейтіні байқалды.

Зерттеу нәтижелері көрсеткендей, ұсынылған болжамды басқару тәсілі өндірістік жабдықтың жылулық режимін тиімді бақылауға және өндірістік жүктемені оңтайландыру арқылы энергия шығынын 10–25 % аралығында азайтуға мүмкіндік береді.

Ұсынылған архитектура (Сурет 1) интеллектуалды жүйені жобалау мен енгізудің толық функционалдық архитектурасын сипаттайды. Архитектурада деректерді жинау, алдын ала өңдеу, модельді оқыту, жүйеге енгізу және нақты уақыт режимінде нәтижелерді түзету кезеңдері біртұтас көпсатылы процесс ретінде көрсетілген.

Жүйені әзірлеудің бастапқы кезеңінде пәндік облысқа талдау жүргізіліп, шешілетін мәселенің негізгі шекаралары, функционалдық талаптары және қолдану сценарийлері анықталады. Осыдан кейін әртүрлі ақпарат көздерінен құрылымдалған және құрылымдалмаған деректер жиналып, олар кейінгі талдау мен модель құру процесінің бастапқы деректер қоры ретінде пайдаланылады.

Келесі кезеңде интеллектуалды басқару жүйесінің архитектурасы жобаланып, оның негізгі модульдері мен компоненттері арасындағы өзара байланыс қалыптастырылады. Осы архитектура негізінде CNN, LSTM сияқты машиналық оқыту мен нейрондық желілерге сүйенген модельдер таңдалып, жүйенің математикалық моделі жасалады.

Әзірленген модельдер бағдарламалық ортада іске асырылып, оқыту және тестілеу кезеңдерінен өтеді. Бұл кезеңде модельдердің дәлдігі, тұрақтылығы

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

және жұмыс тиімділігі бағаланып, параметрлері оңтайландырылады. Кейін жүйе нақты өндірістік ортаға енгізіліп, деректер ағынын өңдеу және шешім қабылдау модульдерімен біріктіріледі.

Жүйенің келесі деңгейінде бейне немесе сигналдарды өңдеу жүзеге асырылады. Бұл интеллектуалды талдау процесінің маңызды бөлігі болып саналады. Алынған нәтижелер негізінде критикалық нүктелер анықталып, олардың маңыздылық деңгейі бағаланады. Егер қабылданған шешімдер қойылған талаптарға сәйкес келмесе, жүйе автоматты түрде қайта есептеу немесе параметрлерді түзету механизмін іске қосады.

Жалпы алғанда, ұсынылған архитектура деректерді көпсатылы өңдеуге, интеллектуалды талдауға және адаптивті басқаруға негізделген кешенді жүйенің құрылымын сипаттайды. Мұндай тәсіл қазіргі заманғы интеллектуалды басқару жүйелерінде, әсіресе өндірістік процестерді автоматтандыру мен цифрлық модельдеу саласында кеңінен қолданылады.

Берілген архитектура өндірістік жабдықтардың жағдайын бақылауға және predictive maintenance қағидалары негізінде алдын ала техникалық қызмет көрсетуді ұйымдастыруға арналған мониторинг жүйесінің интерфейсін көрсетеді. Жүйе нақты уақыт режимінде сенсорлардан келетін мәліметтерді тіркеп, оларды аналитикалық өңдеуге жібереді.

Интерфейсте температура, вибрация, қысым, электр тогы, жүктеме деңгейі сияқты негізгі технологиялық параметрлер көрсетілген. Бұл көрсеткіштер өндірістік жабдықтың ағымдағы техникалық күйін сипаттап, оның сенімділігін бағалауға мүмкіндік береді.

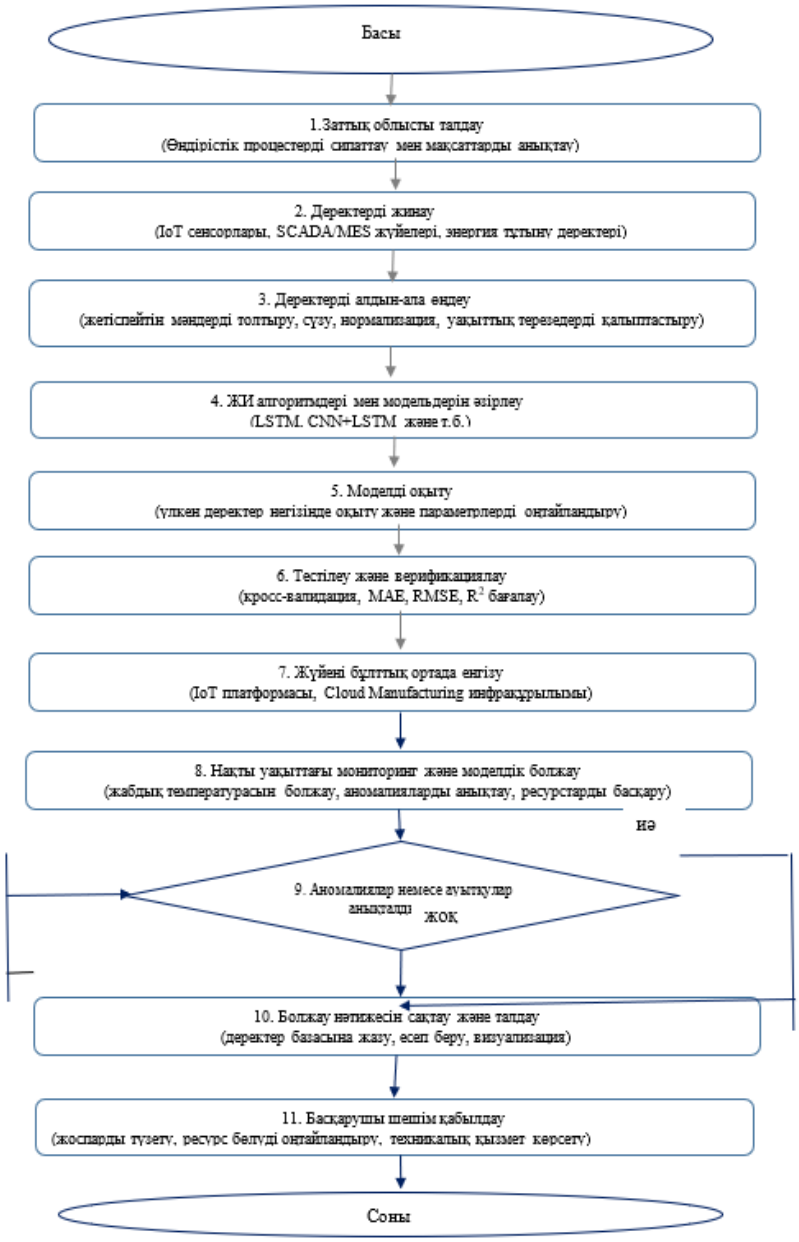
Төменгі бөлікте орналасқан графиктер параметрлердің уақыт бойынша өзгеру динамикасын бейнелейді. Температура мен вибрация көрсеткіштерінің трендтері жабдықтың қалыпты немесе ауытқулы режимдерін анықтауға көмектеседі. Мысалы, белгілі бір уақыт аралығында параметрлердің күрт өзгеруі жүйедегі артық жүктемені немесе ықтимал ақаудың пайда болуын көрсетуі мүмкін.

Жүйенің негізгі мақсаты — IoT технологиялары арқылы жиналған мәліметтерді орталықтандырылған түрде визуализациялау және машиналық оқыту алгоритмдеріне арналған бастапқы деректер қорын қалыптастыру. Осы деректер негізінде болашақтағы істен шығуларды болжау, ақауларды ерте анықтау және өндірістік процестердің тиімділігін арттыру жүзеге асырылады.

Осылайша, бұл жүйе интеллектуалды өндірістік мониторингтің бастапқы деңгейін сипаттайды және деректерді жинау, өңдеу және талдау тізбегінің маңызды құрамдас бөлігі болып табылады.

2-суретте модель жұмысына әсер ететін нақты уақыттағы IoT датчиктерінен алынған кіріс параметрлерінің көрсеткіштері берілген. Бұл параметрлер интеллектуалды жүйенің ағымдағы орта жағдайын талдауына және өндірістік процестердің күйін бағалауына мүмкіндік береді.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License



Сур. 1. Өндірістік процестерді интеллектуалды басқару жүйесінің архитектурасы

Бұл зерттеуде аномалия ретінде өндірістік жабдықтың қалыпты жұмыс режимінен ауытқуы қарастырылды. Температураның технологиялық шекті мәннен жоғарылауы, вибрация деңгейінің күрт өзгеруі немесе энергия тұтынудың қалыпты диапазоннан шығуы ықтимал ақаудың белгісі ретінде бағаланды. Мұндай жағдайлар анықталған кезде жүйе алдын ала ескерту қалыптастырып, техникалық қызмет көрсету жұмыстарын жоспарлауға мүмкіндік береді.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License



Сур. 2. IoT датчиктерінен нақты уақыттағы жабдық мониторингі

Атап айтқанда, температура көрсеткіші T=72,4°C мәніне тең, бұл жүйенің қалыпты жұмыс режиміндегі жылулық жағдайын сипаттайды. Сонымен қатар, жүйенің механикалық тұрақтылығын бақылау үшін маңызды болып табылатын вибрация көрсеткіші 1,85 мм/с мәнін көрсетіп тұр. Өндірістік ортаның акустикалық фонын бағалауға мүмкіндік беретін дыбыс деңгейі 65,8 дБ деңгейінде тіркелген.

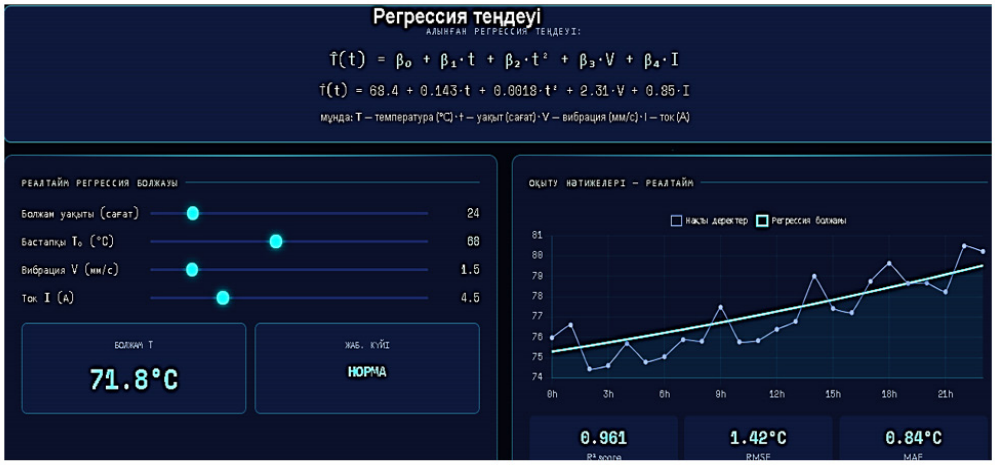
Жабдықтың электрлік жүктемесін сипаттайтын электр тогының күші 4,21 А мәніне тең. Аталған параметр энергия тұтынуды тікелей сипаттайтындықтан, ақауларды ерте анықтауда негізгі рөл атқарады.

Осы кіріс параметрлері жүйеге нақты уақыт режимінде түсіп отырады және кейін машиналық оқыту модельдері арқылы өңделіп, өндірістік процестердің жағдайын болжау мен интеллектуалды басқару шешімдерін қабылдау үшін пайдаланылады.

Сурет 3-те ұсынылған LSTM моделінің жұмыс интерфейсі мен оның болжау нәтижелері бейнеленген. Интерфейстің төменгі бөлігінде жүйенің нәтижелік температурасы 71,8°C екені көрсетілген. Сонымен қатар жүйе күйі «НОРМА» ретінде белгіленген, бұл өндірістік жабдықтың қалыпты жұмыс режимінде жұмыс істеп тұрғанын білдіреді. Аталған блок жүйенің ағымдағы жағдайын бақылауға және басқару шешімдерін жедел қабылдауға мүмкіндік береді.

Бұл зерттеуде LSTM моделі уақыттық қатарлардағы ұзақ мерзімді байланыстарды тиімді үйрену қабілетіне байланысты таңдалды. Сонымен қатар ол өндірістік процестерге тән күрделі және сызықтық емес өзгерістерді сенімді түрде сипаттай алады. GRU моделі есептеу жағынан жеңіл болғанымен, күрделі динамикалық процестерді бейнелеуде LSTM-нің мүмкіндігі жоғарырақ. Transformer моделі үлкен көлемдегі деректер мен қуатты есептеу ресурстарын талап етеді, ал TCN көбіне тұрақты ұзындықтағы уақыт терезелерімен жұмыс істегенде тиімді нәтиже береді. Сондықтан өндірістік деректердің көлемі мен зерттеу мақсатын ескере отырып, LSTM моделі дәлдік пен есептеу күрделілігінің оңтайлы үйлесімін қамтамасыз ететін әдіс ретінде таңдалды.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License



Сур. 3. Ұсынылған LSTM моделінің жұмыс интерфейсі

Осылайша, ұсынылған гибридті архитектурада Conv1D қабаты кеңістіктік ерекшеліктерді бөліп алса, LSTM қабаты уақыттық тәуелділіктерді модельдейді. Сондықтан бұл жұмыста «гибридті модель» ұғымы CNN және LSTM нейрондық желілерінің бір архитектурада біріктірілуін білдіреді.

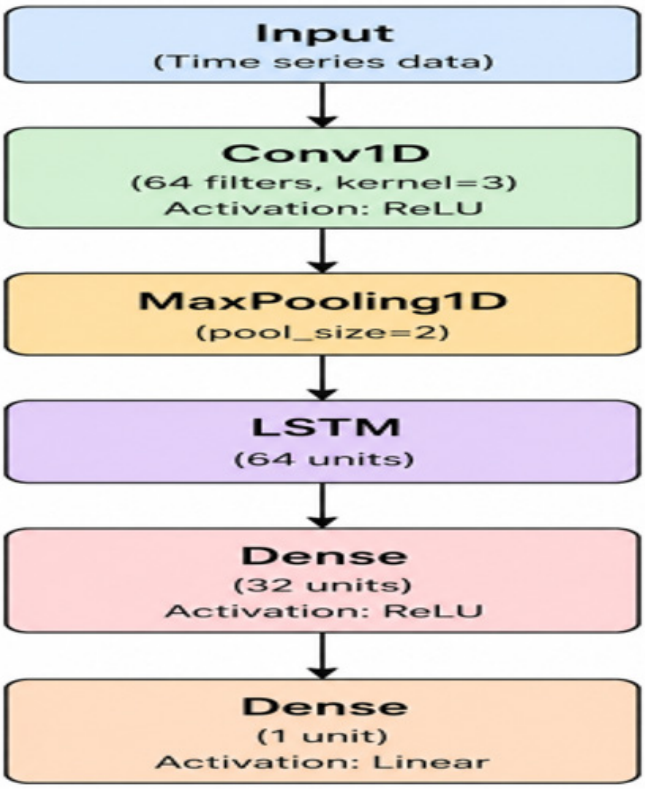
4-суретте ұсынылған CNN–LSTM моделінің құрылымы берілген. Алдымен уақыттық қатарлар Conv1D қабатында өңделіп, маңызды белгілер анықталады. MaxPooling1D қабаты деректердің өлшемін ықшамдайды. Бұдан кейін LSTM қабаты уақыт бойынша тәуелділіктерді талдап, алынған ақпарат толық байланысқан Dense қабаттары арқылы өңделеді. Соңғы Dense(1) қабаты болжанатын көрсеткіштің мәнін шығарады. Бұл архитектура кеңістіктік ерекшеліктер мен уақыттық тәуелділіктерді қатар қарастыру арқылы болжау сапасын жақсартуға мүмкіндік береді.

MAE, RMSE және R^2 көрсеткіштері 5-fold cross-validation нәтижелерінің орташа мәндері ретінде есептелді.

Графикте регрессиялық модель нәтижелері уақыт бойынша трендтік болжам түрінде көрсетілген. Көк сызық нақты өлшенген мәндерді сипаттаса, жасыл сызық регрессиялық модель арқылы алынған болжам нәтижелерін көрсетеді. Графиктен жүйеде жалпы өсу тенденциясының байқалатынын және нақты деректер мен модель нәтижелерінің бір-біріне өте жақын орналасқанын көруге болады. Бұл модельдің болжау дәлдігінің жоғары екенін дәлелдейді.

Төменгі бөлікте модель сапасын сипаттайтын негізгі көрсеткіштер берілген. Атап айтқанда, 0,961 мәні модельдің сәйкестік коэффициентін (R^2) сипаттайды және модель нәтижесінің нақты деректерге жоғары деңгейде сәйкес келетінін көрсетеді. Орташа абсолюттік қате MAE = 0,84°C болса, RMSE көрсеткіші 1,42°C мәніне тең. Бұл нәтижелер модельдің тұрақтылығы мен сенімділігін растайды.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License



Сур. 4. CNN–LSTM гибридті моделінің архитектурасы

Жалпы алғанда, ұсынылған интерфейс көпфакторлы деректерді өңдеуге, регрессиялық модельдер негізінде температура немесе энергия тұтыну динамикасын болжауға және нақты әрі болжамды мәндерді салыстыру арқылы жүйенің тиімділігін бағалауға мүмкіндік береді. Мұндай шешім Big Data технологиялары мен интеллектуалды мониторинг қағидаларына негізделген өндірістік немесе энергетикалық басқару жүйесінің прототипі ретінде қарастырылуы мүмкін.

Cloud Manufacturing (CMfg) өндірістік ресурстарды бірыңғай бұлттық ортада біріктіріп, оларды тиімді басқаруға мүмкіндік береді. Бұл жүйе MES платформасымен ықпалдасып, өндірісті жоспарлау, техникалық қызмет көрсету және ресурстарды басқару процестерін қолдайды. Сонымен қатар жабдықтар, бағдарламалық қамтамасыз ету, деректер мен есептеу ресурстары Manufacturing-as-a-Service (MaaS) қағидаты бойынша қызмет ретінде ұсынылады. Осындай тәсіл өндірістік ресурстарды ұтымды пайдалануға, энергия тұтынуын оңтайландыруға және машиналық оқыту негізінде жабдықтың техникалық жағдайын алдын ала бағалауға мүмкіндік береді.

Бұлттық инфрақұрылым жасанды интеллект пен машиналық оқыту модельдерін енгізуді де жеңілдетеді. LSTM, CNN және гибридті CNN+LSTM модельдері үлкен көлемдегі өндірістік деректер негізінде оқытылып, нақты уақыт режимінде шешім қабылдау жүйелерінде қолданылады.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Жүргізілген эксперименттер барысында бұлтты есептеу ресурстарын пайдалану деректерді өңдеу жылдамдығын арттырып, LSTM және CNN+LSTM модельдері болжау дәлдігін жақсартқаны анықталды. MAE және RMSE көрсеткіштері ұсынылған тәсілдің сенімді әрі практикалық қолдануға жарамды екенін растады. Сонымен қатар ұсынылған тәсіл өндірістік жабдықтардың ықтимал істен шығуын алдын ала болжауға, техникалық қызмет көрсету шығындарын азайтуға, өндірістік процестердің үздіксіздігін қамтамасыз етуге және ресурстарды тиімді басқаруға мүмкіндік береді. Алынған нәтижелер CMfg негізіндегі интеллектуалды бұлттық жүйелердің өндірістің цифрлық трансформациясында маңызды рөл атқаратынын көрсетті.

Қорытынды.

Ұсынылған шешімнің басты ерекшелігі — болжамды талдау, интеллектуалды ресурстарды басқару және бұлттық есептеу технологияларын бірыңғай цифрлық өндірістік ортада өзара байланыстыра қолдану мүмкіндігі. Бұл өндірістік орталардың тиімділігін арттырып қана қоймай, олардың сенімділігі мен икемділігін де күшейтеді.

Эксперимент барысында температура, вибрация және жүктеме сияқты негізгі технологиялық параметрлердің уақыт бойынша өзгерісі талданды. Нәтижесінде қалыпты және ауытқулы жұмыс режимдері арасындағы айырмашылықтар анық байқалды. Әсіресе көрсеткіштердің күрт төмендеуі немесе жоғарылауы жабдықтағы ықтимал ақауларды ерте кезеңде болжауға мүмкіндік беретіні дәлелденді.

Болашақ зерттеулер шеңберінде әзірленген гибридті модельдерді өндірістің өзге де технологиялық параметрлерін (мысалы, діріл деңгейі мен қысым аномалияларын) болжауға бейімдеу және бұлттық жүйенің киберқауіпсіздік деңгейін арттыру мәселелерін қарастыру жоспарлануда.

REFERENCES

Andronie M. et al. (2023). “Big data management algorithms, deep learning-based object detection technologies, and geospatial simulation in the Internet of robotic things”, ISPRS Int. J. Geo-Inf. — Vol. 12. — No. 2. — Pp. 35. [In Eng.].

Bai H. et al. (2025). “Energy optimization model for enterprise production based on deep learning” // Int. J. Computational Intelligence Systems. [In Eng.].

Ekwaro-Osire H. et al. (2025). “Manufacturing process energy consumption modeling” // Journal of Intelligent Manufacturing. [In Eng.].

Glücker P. et al. (2026). “Optimization of industrial energy systems for voltage support” // Computers & Chemical Engineering. [In Eng.].

Gunaratne M. et al. (2026). “A digital twins platform for digital manufacturing” // Electronics. — Vol. 15. [In Eng.].

Han M. et al. (2026). “Dynamic graph attention for industrial energy systems” // Journal of Energy Storage. [In Eng.].

Kougioumtzidis G. et al. (2025). “Deep reinforcement learning-based resource allocation for QoE enhancement in wireless VR communications” // IEEE Access. [In Eng.].

Le T., Vinh L. Q., and Pham V. H. (2025). “VR/AR robotics-assisted deep reinforcement learning for smart manufacturing” // Sensors. — Vol. 25. [In Eng.].

Lu Y. et al. (2026). “Energy-efficient job shop scheduling in renewable energy manufacturing” // Machines. — Vol. 14. [In Eng.].

Leon-Medina J. X. et al. (2025). “Forecasting energy demand in manufacturing: A data-driven approach”

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

// Sensors. — Vol. 25. [In Eng.].

Madaminova B. et al. (2025). "Multi-objective optimization for energy efficiency and production scheduling using reinforcement learning" // Int. J. Industrial Engineering and Management. [In Eng.].

Morariu C., Morariu O., Răileanu S., and Borangiu T. (2020). "Machine learning for predictive scheduling and resource allocation in large-scale manufacturing systems" // Computers in Industry. DOI: 10.1016/j.compind.2020.103244. [In Eng.].

Scientific Reports. [In Eng.].

Tsai C. H. et al. (2026). "AI-assisted manufacturing execution systems for real-time production management" // Future Technology. [In Eng.].

Wang L. et al. (2022). "Augmented reality-assisted prognostics system based on deep learning for IoT manufacturing" // Sensors. — Vol. 22. [In Eng.].

Wang W. et al. (2026). "IoT-enabled AI framework using BiLSTM for eco-efficient production",

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 46–65
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.003

МРНТИ 28.23.15

ARCHITECTURE OF AN INTELLIGENT CHATBOT FOR SMART CITY: AI-BASED RECEIVING, ANALYSIS AND ROUTING OF CITIZEN REQUESTS

A.B. Kassymova[1]*, R.K. Uskenbayeva[1], Zh.A. Sydykov[1], D. Omirzak[2], S. Kalibekov[3]

[1]JSC Kazakh National Research Technical University named after K.I. Satbayev, Almaty, Kazakhstan;
[2]Kaspi Bank, Almaty, Kazakhstan;
[3]Otbasy Bank, Almaty, Kazakhstan.
E-mail: a.kassymova@satbayev.university

Aizhan B. Kassymova — PhD, Associate professor of the Information Systems Department, JSC Kazakh National Research Technical University named after K.I. Satbayev, Almaty, Kazakhstan
E-mail: a.kassymova@satbayev.university, https://orcid.org/0000-0003-2999-5745;
Raissa K. Uskenbayeva — Doctor of technical Sciences, Professor of the Information Systems Department, JSC Kazakh National Research Technical University named after K.I. Satbayev, Almaty, Kazakhstan
https://orcid.org/0000-0002-8499-2101;
Zhanibek A. Sydykov — Senior Lecturer of the Information Systems Department, JSC Kazakh National Research Technical University named after K.I. Satbayev, Almaty, Kazakhstan
https://orcid.org/0009-0008-3368-3588;
Dinmukhammed Omirzak — IT Developer, Kaspi Bank, Almaty, Kazakhstan
https://orcid.org/0009-0001-9348-7203;
Sanzhar Kalibekov — IT Developer, Otbasy Bank, Almaty, Kazakhstan
https://orcid.org/0009-0004-7831-1316.

© A.B. Kassymova, R.K. Uskenbayeva, Zh. Sydykov, D. Omirzak, S. Kalibekov

Abstract. The paper addresses the development of an intelligent digital service for receiving, analysing and routing citizen requests within a Smart City ecosystem, in which a Telegram bot serves as the primary communication channel. The aim is to design a service architecture that combines a conversational interface, a multi-step submission scenario, storage of multimedia attachments, and an automatic thematic

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

classification module followed by routing to responsible municipal departments. The service is built on a finite-state machine, SMS OTP authorisation and an S3-compatible store; classification quality was compared across three models. On a corpus of 3,260 Russian-language requests, a RuBERT-class model achieved the highest quality (accuracy 0.914; F1 0.907), outperforming linear baselines. Simulation of two processing scenarios showed a reduction of the average time to forward a request to the relevant department from 26.15 to 6.85 minutes.

Keywords: Smart City, chatbot, artificial intelligence, automatic text classification, complaint routing, service architecture

For citation: Kassymova A.B., Uskenbayeva R.K., Sydykov Zh.A., Omirzak D., Kalibekov S. (2026). Architecture of an intelligent chatbot for smart city: ai-based receiving, analysis and routing of citizen requests // International journal of information and communication technologies. Vol. 7. No. 27. Pp. 46–65. https://doi.org/10.54309/IJICT.2026.27.3.003. (In Eng.).

Conflict of interest: The authors declare that there is no conflict of interest.

Funding. *This research has been funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. BR24993051).*

SMART CITY ҮШІН ИНТЕЛЛЕКТУАЛДЫ ЧАТ-БОТ АРХИТЕКТУРАСЫ: АЗАМАТТАРДЫҢ ӨТІНІШТЕРІН ЖАСАНДЫ ИНТЕЛЛЕКТ НЕГІЗІНДЕ ҚАБЫЛДАУ, ТАЛДАУ ЖӘНЕ БАҒЫТТАУ

А.Б. Касымова[1]*, Р.К. Ускенбаева[1], Ж.А. Сыдыков[1], Д. Өмірзақ[2], С. Калибеков[3]

[1]К.И. Сәтбаев атындағы Қазақ Ұлттық Техникалық Зерттеу Университеті, Алматы, Қазақстан;
[2]Kaspi Bank, Алматы, Қазақстан;
[3]Отбасы Банк, Алматы, Қазақстан;
E-mail: a.kassymova@satbayev.university

Қасымова Айжан Бахытжанқызы — PhD, «Ақпараттық жүйелер» кафедрасының қауымдастырылған профессоры, К.И. Сәтбаев атындағы Қазақ Ұлттық Техникалық Зерттеу Университеті, Алматы, Қазақстан
E-mail: a.kassymova@satbayev.university, https://orcid.org/0000-0003-2999-5745;
Өскенбаева Раиса Кәбиқызы — техникалық ғылымдарының докторы, «Ақпараттық жүйелер» кафедрасының профессоры, К.И. Сәтбаев атындағы Қазақ Ұлттық Техникалық Зерттеу Университеті, Алматы, Қазақстан
https://orcid.org/0000-0002-8499-2101;
Сыдыков Жәнібек — «Ақпараттық жүйелер» кафедрасының Аға оқытушы, К.И. Сәтбаев атындағы Қазақ Ұлттық Техникалық Зерттеу Университеті, Алматы, Қазақстан
https://orcid.org/0009-0008-3368-3588;

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Өмірзақ Дінмұхаммед — IT әзірлеушісі, Kaspi Bank
https://orcid.org/0009-0001-9348-7203;
Калибеков Санжар — IT әзірлеушісі, Отбасы Банк
https://orcid.org/0009-0004-7831-1316.

© А.Б. Касымова, Р.К. Өскенбаева, Ж.А. Сыдықов, Д. Өмірзақ, С. Калибеков

Аннотация. Мақалада Smart City экожүйесінде азаматтардың өтініштерін қабылдау, талдау және бағыттау үшін Telegram-бот өзара әрекеттесу арнасы ретінде пайдаланылатын интеллектуалды цифрлық сервисті әзірлеу қарастырылады. Жұмыстың мақсаты — диалогтік интерфейсті, өтінішті берудің көпқадамды сценарийін, мультимедиалық тіркемелерді сақтауды және обращенияларды профильді бөлімшелерге кейіннен бағыттай отырып автоматты тақырыптық жіктеу модулін біріктіретін сервис архитектурасын жобалау. Сервис ақырлы автомат, SMS OTP арқылы авторизация және S3-үйлесімді қойма негізінде іске асырылды; жіктеу үш модель үшін салыстырылды. 3260 орыс тілді өтініштен тұратын корпуста RuBERT класындағы модель ең жоғары сапа көрсетті (accuracy 0,914; F1 0,907) және сызықтық әдістерден асып түсті. Өңдеудің екі сценарийін модельдеу обращенияны профильді бөлімге жіберудің орташа уақытын 26,15-тен 6,85 минутқа дейін қысқартатынын көрсетті.

Түйінді сөздер: Smart City, чат-бот, жасанды интеллект, мәтінді автоматты жіктеу, шағымдарды бағыттау, сервистік архитектура

Дәйексөздер үшін: Қасымова А.Б., Өскенбаева Р.К., Сыдықов Ж.А., Өмірзақ Д., Калибеков С. (2026). Smart city үшін интеллектуалды чат-бот архитектурасы: азаматтардың өтініштерін жасанды интеллект негізінде қабылдау, талдау және бағыттау // Халықаралық ақпараттық және коммуникалық технологиялар журналы. Т. 7. No 27. Б. 46–65. https://doi.org/10.54309/IJICT.2026.27.3.003. (Ағыл. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

Қаржыландыру. *Бұл зерттеу Қазақстан Республикасы Ғылым және жоғары білім министрлігінің Ғылым комитеті тарапынан қаржыландырылды (Грант № BR24993051).*

АРХИТЕКТУРА ИНТЕЛЛЕКТУАЛЬНОГО ЧАТ-БОТА ДЛЯ SMART CITY: ПРИЁМ, АНАЛИЗ И МАРШРУТИЗАЦИЯ ОБРАЩЕНИЙ ГРАЖДАН НА ОСНОВЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

А.Б. Касымова[1]*, Р.К. Ускенбаева[1], Ж.А.Сыдыков[1], Д. Омирзак[2], С. Калибеков[3]

[1]Казахский Национальный Исследовательский Технический Университет им. К.И. Сатпаева, Алматы, Казахстан;
[2]Kaspi Bank, Алматы, Казахстан;

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

[3]Отбасы Банк, Алматы, Казахстан.
E-mail: a.kassymova@satbayev.university

Касымова Айжан Бахытжановна — PhD, ассоциированный профессор кафедры «Программная инженерия», Казахский Национальный Исследовательский Технический Университет им. К.И. Сатпаева, Алматы, Казахстан
E-mail: a.kassymova@satbayev.university, https://orcid.org/0000-0003-2999-5745;
Ускенбаева Раиса Кабиевна — доктор технических наук, профессор кафедры «Программная инженерия», Казахский Национальный Исследовательский Технический Университет им. К.И. Сатпаева, Алматы, Казахстан
https://orcid.org/0000-0002-8499-2101;
Сыдыков Жанибек — старший преподаватель кафедры «Информационные системы», Казахский Национальный Исследовательский Технический Университет им. К.И. Сатпаева, Алматы, Казахстан
https://orcid.org/0009-0008-3368-3588;
Омирзак Динмухаммед — ИТ разработчик, Kaspi Bank
https://orcid.org/0009-0001-9348-7203;
Калибеков Санжар — ИТ разработчик, Отбасы Банк
https://orcid.org/0009-0004-7831-1316.

© А.Б. Касымова, Р.К. Ускенбаева, Ж.А.Сыдыков, Д. Омирзак, С. Калибеков

Аннотация. В статье рассматривается разработка интеллектуального цифрового сервиса приёма, анализа и маршрутизации обращений граждан в экосистеме Smart City. Каналом взаимодействия с пользователями выступает Telegram-бот. Цель работы — спроектировать архитектуру сервиса, объединяющую диалоговый интерфейс, многошаговый сценарий подачи обращения, хранение мультимедийных вложений и модуль автоматической тематической классификации с последующей маршрутизацией в профильные подразделения. Сервис реализован на основе конечного автомата, авторизации через SMS OTP и S3-совместимого хранилища. Качество классификации сопоставлялось для трёх моделей. На корпусе из 3260 русскоязычных обращений модель класса RuBERT показала наибольшее качество (accuracy 0,914; F1 0,907), превзойдя линейные методы. Моделирование двух сценариев обработки показало сокращение среднего времени передачи обращения в профильный отдел с 26,15 до 6,85 минуты.

Ключевые слова: Smart City, чат-бот, искусственный интеллект, автоматическая классификация текста, маршрутизация жалоб, сервисная архитектура

Для цитирования: Касымова А.Б., Ускенбаева Р.К., Сыдыков Ж.А., Омирзак Д., Калибеков С. (2026). Архитектура интеллектуального чат-бота для smart city: приём, анализ и маршрутизация обращений граждан на основе искусственного интеллекта // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 27. Стр. 46–65. https://doi.org/10.54309/IJICT.2026.27.3.003. (На англ.).

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Финансирование. *Данное исследование профинансировано Комитетом науки Министерства науки и высшего образования Республики Казахстан (грант № BR24993051).*

Introduction.

In recent years, the development of the Smart City concept has increasingly been linked not only to sensor infrastructure and urban data analytics, but also to the emergence of new digital channels for interaction between residents and government agencies. Against this backdrop, services that allow for quick reporting of urban issues, attaching supporting materials, and receiving feedback as the request is processed are particularly important. For such tasks, interfaces with a low entry barrier, accessibility on mobile devices, and the ability to automatically process incoming messages are in demand. Messengers, and Telegram bots in particular, are becoming one of the most natural user channels and are increasingly being considered as a component of digital city services (Senadheera et al., 2024; Bono Rossello et al., 2025).

The choice of Telegram as a channel for receiving requests is driven by a few practical considerations. The platform provides an open Bot API, supports the transfer of text, photos, videos, and documents, an Inline Keyboard, and conversation status. It does not require a separate app and is widely used among the region's population. These features lower the barrier to entry for residents while simultaneously providing the developer with a manageable software entry point to which server-side logic and an analysis module can be connected. Unlike web forms and call centers, the messenger's dialog interface allows the user to be guided step-by-step through the application process, ensuring the completeness of the entered data is verified on the client side.

However, interface convenience is not the only factor in demand. Research on digital dialogue in smart cities shows that resident satisfaction depends significantly on the timeliness of responses, the quality of government response, and the effectiveness of interactions (Shen et al., 2025). Citizens expect not only an accessible channel but also a noticeable reduction in the initial request processing time, transparency in the complaint flow, and more accurate recording of its content. Trust in the service itself plays a comparable role: the preferences of residents and front-office staff are also determined by the clarity of the operational logic, the accuracy of responses, and the ability to escalate to a human (Hemesath et al., 2024).

These expectations are driving a shift from passive message reception channels to intelligent services capable of automatically extracting meaning from text, identifying the subject of a complaint, and routing it to the appropriate department. Research on the automation of administrative procedures confirms that artificial intelligence can reduce routine workloads and speed up the initial processing of requests, but requires clear framework conditions related to data quality, accountability, and outcome control (Parycek et al., 2024). This leads to an important conclusion for city services: it makes sense to design request classification and routing not as an isolated NLP module, but

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

as part of a managed service architecture (Rizun et al., 2025).

Data management requirements pose an additional constraint. A complaint reception service processes personal data, photographs, and video attachments, and therefore must consider not only the functional and algorithmic properties of AI but also infrastructural aspects. These include access control, file storage, action logging, and resilience to leaks and misrouting (König, 2021; Panahi et al., 2022). This study is part of a broader project to create an intelligent city system based on the Internet of Things and data analytics, in which the complaint reception service serves as a human-facing channel complementing sensory sources of city data.

A literature review shows that individual areas-AI support for digital participation, the use of chatbots in local government, automatic complaint classification, and data security requirements-have been studied in considerable detail. However, they are typically considered separately. A scenario in which a Telegram bot serves as the front-end interface for a city service, and an AI module is integrated into a unified architecture for receiving, analyzing, and routing requests, followed by experimental evaluation of classification quality, is poorly represented in the literature. Such solutions are even less frequently evaluated for Russian-and Kazakh-language request flows in a specific regional context (Rakhimzhanov et al., 2025). This gap determines the relevance of this study.

The need for automation is increasing as the volume of digital requests grows. With the expansion of electronic channels, the number of incoming messages increases, and manual sorting becomes a bottleneck. Operators are forced to read each request, assign it to the appropriate topic, and select the appropriate department, which slows response times and increases the risk of errors during peak loads. Transferring routine sorting to an algorithm allows for human intervention to be retained where it's truly needed while simultaneously speeding up the processing of the massive influx of typical complaints. This combination of automatic classification and the managed transfer of questionable cases to an operator forms the basis of the proposed service.

The research object is a digital service for receiving and processing citizen requests in a Smart City environment. The subject of the study is the architectural, algorithmic, and applied aspects of building a Telegram-based service with intelligent classification and routing of requests. The goal of the work is to develop an intelligent service architecture that integrates a dialog interface, service-based request processing, storage of multimedia attachments, and an artificial intelligence module into a reproducible model of an urban digital service. To achieve this goal, the following tasks were solved: designing a service architecture and data model; formalizing a request submission scenario based on a finite-state machine; implementing a thematic text classification and routing algorithm; deploying the service in a containerized environment and conducting functional, load, and usability testing; performing a comparative evaluation of classification models and assessing the impact of AI routing on initial processing time. The scientific novelty of this work lies in the integration of a dialog frontend, a managed service architecture, and a Russian-language request classification module into a reproducible urban service model, with an experimental

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

evaluation of both the classification quality and the applied effect of routing.

Materials and methods.

The proposed approach combines a conversational user interface, a service architecture for data storage and processing, and an artificial intelligence module for thematic classification and automatic routing of requests. Methodologically, the work combines system analysis, design based on service-oriented architecture principles, design science with prototyping of a working service, formalization of user scenarios using finite-state machine theory, and experimental comparison of classification algorithms. This combination of methods allows us to simultaneously consider the solution at three levels-architectural, algorithmic, and applied-and link engineering solutions to measurable quality metrics.

The basic service architecture is shown in Figure 1. The entry point is the Telegram API, through which messages are received by a service bot implemented according to Clean Architecture principles. Separation into layers of domain, use cases, adapters, and infrastructure ensures the independence of the request processing business logic from specific external components and simplifies the replacement of storage, gateways, and models without rewriting the core. The service bot interacts with a PostgreSQL relational database for storing requests and metadata, with Redis for session management, state machine states, and caching, and with S3-compatible MinIO storage for file attachments. Authorization is performed through an external SMS OTP service. An administrative web interface is available for employees (Fig. 1).

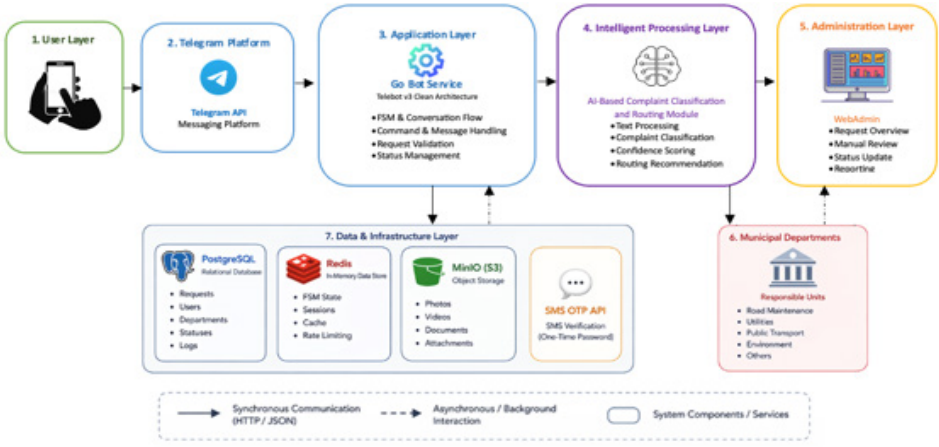


Fig. 1. Architecture of the intelligent service for receiving and routing citizen requests

Each component performs a separate function. The Telegram API is responsible for message delivery and attachment reception; the service bot implements user scenarios, input validation, and call orchestration; PostgreSQL stores structured request data and ensures its consistency during concurrent work by employees; Redis stores the rapidly changing state of a conversation and reduces the load on the database through caching; MinIO stores photos, videos, and documents, separating binary objects from relational data; and the administrative web interface provides employees with access to

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

the list of requests, their statuses, and attachments. This separation of responsibilities is consistent with recommendations to consider request classification not as an isolated module, but as an element of a managed service infrastructure (Rizun et al., 2025).

The administrative web interface is designed for employees and complements the dialog channel with a processing function. The prototype features a dashboard with quantitative indicators of the request flow (total number, new, pending, resolved, and rejected), a breakdown of requests by subject category, a feed of recently received requests with status and date, and a card for each request with its text, change history, correspondence, and attached files. The interface supports language switching and request status management. This set of functions transforms the result of intelligent classification into a working tool: employees see the proposed category and responsible department, can confirm or adjust the decision, and track the entire complaint flow. Linking the dashboard to the same data entities used by the bot ensures a unified view of the request for both the requester and the operator and supports the transparency of processing that is essential for trust in city services (König, 2021).

The data model includes related entities reflecting the request lifecycle. The user entity stores the chat ID, phone number, name, interface language, role, and confirmation flag. A ticket contains a title, text, predicted category, status, responsible department, model confidence score, and spam flag, allowing the results of the intelligent analysis to be stored alongside the ticket itself. The ticket files are stored in a separate entity with a link to the storage object, type, and size, separating the management of attachments from the text data. User sessions record the current state of the state machine and the ticket draft, change history and discussion messages provide audit trails and communication between the requester and the operator, and department and employee entities define the organizational structure to which tickets are routed. Having a history and explicitly recording old and new field values increases the transparency of the complaint flow-a property that has been linked in the literature to trust in digital city services (König, 2021).

The ticket submission scenario is formalized as a deterministic finite-state machine (FSM) described by a tuple:

$$A = (S, E, \delta, s_0, F) \tag{1}$$

where S is a finite set of states, E is the set of input events (input alphabet), δ the transition function, s_0 is the initial state, $F \subseteq S$ is the set of final states. The set of states and the subset of final states are defined as:

$$S = \{s_0, s_1, \dots, s_n\}, F \subseteq S \tag{2}$$

with the states corresponding to authorization, title and text entry, attachment loading, confirmation, and call completion. The machine's response to a single event is determined by the transition function:

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

$$\delta: S \times E \to S \tag{3}$$

and for the sequence of events $w = e_1, e_2, \ldots, e_k$, an extended transition function is introduced, defined recursively:

$$\delta^*(s, w \cdot e) = \delta(\delta^*(s, w), e), \delta^*(s, \varepsilon) = s \tag{4}$$

A completed request corresponds to a path from s_0 to one of the states of set F; if no further event arrives within the waiting time τ, the session is put into an interrupted state, preventing the accumulation of "stuck" drafts. The set of events E includes bot commands (/start, /new, /list, /cancel, /chat), text messages, inline button clicks, and file uploads with size restrictions. Data completeness control on the machine side reduces the number of requests without required fields and improves the suitability of the text for subsequent automatic classification.

After registration, the text of a request x is represented as a numeric vector. For linear models, TF-IDF features are constructed: the weight of a term t in a request x is defined by the product of the term frequency and the inverse document frequency,

$$w_{t,x} = tf(t, x) \cdot idf(t), idf(t) = log \frac{N}{1 + df(t)} \tag{5}$$

where N is the number of documents in the corpus, and d$f(t)$ is the number of documents containing term t. The resulting vector is subjected to L2 normalization

$$\hat{x} = \frac{\varphi(x)}{\|\varphi(x)\|_2} \tag{6}$$

The logistic (softmax) model estimates the posterior probability of each of the $K = 8$ categories using linear logits

$$z_c = w_c^T \hat{x} + b_c, P(c|x) = \frac{e^{z_c}}{\sum_{k=1}^{K} e^{z_k}} \tag{7}$$

A linear support vector machine is trained using the maximum margin principle in a one-versus-all scheme with a hinge loss function and L2 regularization

$$min \frac{1}{2} \|w_c\|^2 + C \sum_{i=1}^{N} \max(0, 1 - y_i^{(c)} (w_c^T \varphi(x_i) + b_c)) \tag{8}$$

The context model of the RuBERT class builds a text representation using the multi-headed attention mechanism of the transformer

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) V \tag{9}$$

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

on top of the aggregated representation, the classification head of which forms the distribution by categories, and the additional training of the model is performed by minimizing the cross-entropy:

$$p = softmax(W_o h_{[CLS]} + b_o), L = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{K} y_{i,c}\, logP(c|x_i) \quad (10)$$

The final category of the appeal and the associated confidence assessment are determined as:

$$\hat{C} = \arg\max_{c \in C} P(c|x), p = \max_{c \in C} P(c|x) \quad (11)$$

The routing decision is given by a threshold rule that considers the spam filter $\sigma(x) \in \{0, 1\}$ and the mapping of categories to responsible departments μ:C→D:

$$R = \oint(\hat{C}, p) = \mu(\hat{C})\ if\ p \geq p_{min} \wedge \sigma(x) = 0; else \rightarrow operator \quad (12)$$

If the confidence score exceeds the p_{min} threshold and the request is not marked as spam, it is routed automatically; otherwise, it is passed to an operator for confirmation. Thus, the p_{min} threshold and the *σ(x)* filter jointly regulate the balance between the proportion of fully automated requests, routing accuracy, and robustness to noisy messages.

Classification quality is assessed using a $K \times K$ confusion matrix *M,* where element M_{ij} is equal to the number of requests of category *i* classified by the model into category *j.* The proportion of correct decisions, precision, and recall for category *c* are defined as:

$$Accuracy = \frac{\sum_{c=1}^{K} M_{cc}}{\sum_{i,j} M_{ij}}, Pr_c = \frac{M_{cc}}{\sum_i M_{ic}}, Re_c = \frac{M_{cc}}{\sum_j M_{cj}} \quad (13)$$

The measure balanced in precision and recall and its averaging over classes are calculated as:

$$F1_c = \frac{2Pr_c Re_c}{Pr_c + Re_c}, F1_{macro} = \frac{1}{K}\sum_{c=1}^{K} F1_c \quad (14)$$

Application efficiency is characterized by the share of automatically routed requests and the average routing time

$$AutoRate = \frac{1}{N}\sum_{i=1}^{N} 1[p_i \geq p_{min} \wedge \sigma(x_i) = 0], \bar{\tau} = \frac{1}{N}\sum_{i=1}^{N} \tau_i \quad (15)$$

where $\mathbb{1}[\cdot]$ is the indicator function, and τ_i is the routing time of the $i - th$ request. This set of metrics links the algorithmic quality of the model with the operational

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

effectiveness of the service and is subsequently used in experimental evaluation.

Three approaches of varying complexity were compared as classification models: logistic regression on TF-IDF features, a linear support vector machine (Joachims, 1998), and a pre-trained language model of the RuBERT class (Kuratov et al., 2019). The choice of this particular set reflects a trade-off between interpretability, training cost, and quality: linear models on sparse features serve as strong and fast baseline solutions, while the language model evaluates the potential gain from taking into account context. The choice of RuBERT is due to the fact that the transformer architecture (Vaswani et al., 2017) and bidirectional pre-training (Devlin et al., 2019) make it possible to take into account the context and semantics of short Russian-language phrases typical of appeals; It is also known that retraining such models for a specific classification task yields a stable increase in quality (Sun et al., 2019; Minaee et al., 2021).

Text preprocessing included lowercase conversion, deletion of special characters, and tokenization; TF-IDF features were constructed for the linear models, and subword tokenization was used for the language model. The corpus was split into training, validation, and test sets (70/15/15), with the final metrics calculated on the holdout test set. Cross-validation was used to ensure the robustness of the estimates. Hyperparameter tuning (linear model regularization coefficient, learning rate, number of epochs, and batch size for RuBERT) was performed on the validation set. Linear models, metrics, and splitting procedures were implemented using the scikit-learn library (Pedregosa et al., 2011). The experiments were performed in a computing environment with the following configuration: Intel Core i7 processor, NVIDIA GeForce RTX 3060 graphics accelerator, 16 GB of RAM, Windows 11 operating system, Python 3.10, pandas 2.2.2, numpy 1.26.4, scikit-learn 1.5.0, matplotlib 3.9.0, PyTorch 2.3.0, and transformers 4.41.2 libraries. Fixing the versions, random seeds, and partitioning parameters ensures the reproducibility of the experiment.

The service is deployed in a containerized environment. Bot images, PostgreSQL, Redis, and MinIO have been prepared; deployment is performed in a Kubernetes cluster (k3s distribution) with health checks and automatic component restarts, while the services are isolated on the internal network. The server side is configured with continuous integration and delivery via GitHub, Telegram update reception via webhook, secrets and environment parameter management, and logging. On the data side, a production PostgreSQL instance with applied schema migrations is used, along with Redis for the state machine, sessions, and cache, and regular scheduled database backups. This configuration meets the infrastructure requirements for city services processing personal data and files (Panahi et al., 2022).

The implemented functionality was tested in three areas, corresponding to the testing phase of the project schedule. Functional testing covered the correctness of state machine transitions and session expirations, the operation of all bot commands, one-time code authorization, and compliance with file size limits. Load testing assessed the service's resilience under high load, PostgreSQL's behavior during parallel queries, and the operation of multiple concurrent FSM sessions. Usability testing focused on

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

the usability of inline buttons, the clarity of the status indicator, the correctness of the bilingual localization, and page-by-page navigation through the ticket list. At the time of preparation, the bot was deployed in a test environment, SMS OTP authorization was operational, and file submission, list and archive viewing, discussion mode, bilingual localization, and file storage in MinIO were implemented.

Since the service processes personal data and user files, the architecture incorporates security measures that meet smart city data management requirements (König, 2021; Panahi et al., 2022). Employee access is restricted by roles, and login is confirmed with a one-time code. Text data and binary attachments are separated into different storage locations: structured information is stored in PostgreSQL, while files are stored in isolated MinIO object storage, accessed through the service. Request actions are recorded in the change history, ensuring processing accountability, and regular backups reduce the risk of data loss. Network isolation of components and secret management through environmental parameters limit the attack surface. These solutions align with the principle that a smart city service should be assessed not only by algorithmic metrics but also by its infrastructural resilience and security (Parycek et al., 2024).

For the experimental evaluation, a set of metrics from two groups-classification quality and applied effectiveness-was used (Table 1). Quality metrics characterize the accuracy of request assignment to a category and the distribution of errors between classes, while applied metrics reflect the practical utility of the service: average routing time and the proportion of requests routed without operator intervention. This composition allows us to consider the service not only as a complaint reception interface but also as an intelligent component of the city's digital infrastructure, evaluating it simultaneously in terms of classification accuracy and operational effectiveness.

Table 1. Evaluation metrics of the intelligent request processing service

Evaluation Group	Metric	Purpose
Classification Quality	Accuracy	The proportion of correctly classified requests
Classification Quality	Precision	The accuracy of assignment to a specific category
Classification Quality	Recall	The completeness of request identification by category
Classification Quality	F1-score	A balanced assessment of classification quality
Classification Quality	Confusion Matrix	An analysis of errors between categories
Applied Efficiency	Mean Routing Time	The average request routing time
Applied Efficiency	Auto-routing Rate	The proportion of requests routed without an operator

To enhance the reliability of the findings, standard validation procedures were used. Models were compared using a single corpus split with fixed initial values, eliminating the influence of randomness of splitting on the relative order of methods. Error analysis complements aggregated metrics by revealing which categories are confounded, thereby shifting the discussion of quality from the overall quantity

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

to the practical consequences of incorrect routing. This set of procedures meets the requirements for reproducibility and evidence-based evaluation of intelligent components and allows us to separate consistent differences between models from random fluctuations.

Quantitative evaluation was conducted in a controlled computing environment using a corpus of queries. The corpus included 3,260 Russian-language queries with variable text length, incomplete wording, colloquial expressions, and topics typical of urban affairs. The queries were distributed across eight categories: road infrastructure, lighting, garbage collection, housing and utilities, public transportation, yard improvement, security, and illegal dumping.

Discussion.

The use of chatbots in local and municipal government has emerged as a distinct field in recent years. Review studies show that such solutions are primarily used for information, consultation, transactional services, and complaint handling, and that their successful implementation depends on citizen-centered design, institutional readiness, and the quality of interactions (Senadheera et al., 2024). Experiments with the preferences of residents and front-office staff further emphasize that not only convenience but also public values-response accuracy, the ability to escalate to a human, and operational logic that meets user expectations-are important for e-government (Hemesath et al., 2024). These results support the approach chosen in this study, in which automatic routing complements, rather than replaces, human interaction in cases of low model confidence.

The second area of research involves text analytics of requests. A review of text analytics tools in the public sector documents the rapid growth of machine learning, topic modeling, and short text classification methods. However, it is noted that a significant portion of solutions remains experimental or exists as separate analytical modules without tight integration into application services (Rizun et al., 2025). Research on real-world datasets of complaints confirms that automatic classification reduces the workload of operators and accelerates subsequent processing. A comparison of intelligent analysis algorithms on crowdsourced complaints revealed significant differences in accuracy between k-nearest neighbors, random forest, support vector machines, and boosting (Madyatmadja et al., 2023); the integration of an attention mechanism into the task of classifying complaints in government-citizen interaction systems yielded an increase in quality by taking into account significant text fragments (Zhang et al., 2025); a mixed-method analysis of complaints to a city hotline confirmed the dependence of the result on the subject taxonomy and the nature of the messages (Zhou et al., 2025). Deep learning-based solutions are also applied in urban complaint mobile applications, where categorization is performed based on the image of the urban problem (Shama et al., 2024).

For the Russian-language and regional context, a comparison of modern approaches to complaint classification is particularly revealing: the application of large language models and instructionally tuned embeddings to public transport complaints demonstrated performance sensitivity to the choice of model, instruction language,

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

and subject taxonomy (Rakhimzhanov et al., 2025). This is consistent with the choice of a pre-trained Russian-language model of the RuBERT class (Kuratov et al., 2019) in this study, based on a transformer architecture (Vaswani et al., 2017) and bidirectional pre-training (Devlin et al., 2019); the accumulated experience of pre-training similar models and reviews of text classification methods indicate a consistent advantage of contextual representations over sparse features in tasks with short messages (Sun et al., 2019; Minaee et al., 2021).

The third area is formed by requirements for data management and security. Literature on Smart City platforms emphasizes that digital city services cannot be designed without considering accountability for data processing, transparent rules for its use, and protection of the digital environment (König, 2021), while systematic reviews of privacy technologies document a wide range of threats and solutions for protecting citizens in smart cities (Panahi et al., 2022). Work on the automation of administrative procedures provides a framework for the responsible use of AI, in which automation must be accompanied by outcome control and the preservation of human intervention in disputed cases (Parycek et al., 2024). These requirements are directly reflected in the architecture of the proposed service-in access control, dedicated file storage, action logging, and threshold logic that transfers uncertain decisions to the operator.

An architectural approach to digital participation deserves special attention. Research on the design of AI-enabled civic participation services formulates principles for constructing collective intelligence architectures in which technological components are linked to organizational and value requirements (Bono Rossello et al., 2025). In this logic, the contribution of this work lies not so much in a single classification model as in the way it is integrated: a dialog frontend, service processing, data and file storage, and a classification module with threshold routing are assembled into a reproducible architecture, where each component has clear responsibilities. This engineering approach, implemented through prototyping, complements the predominantly analytical research on request classification and brings their results closer to the conditions of real-world use in municipal services.

Comparison of these areas reveals a significant gap. AI support for digital participation (Bono Rossello et al., 2025), the use of chatbots in management, request classification, and data security requirements are considered largely in isolation. There are significantly fewer studies in which a Telegram bot serves as a front-end interface, and an AI module is integrated into a single, controlled architecture for receiving, analyzing, and routing data, with experimental evaluation of quality and processing time, and these rarely consider the Russian-language regional context. This paper addresses precisely this gap by considering classification and routing not as a standalone module, but as an element of a holistic service with a reproducible architecture and measurable practical impact.

Results.

The structure of the corpus is shown in Figure 2. The largest share of requests concerned road infrastructure, housing and utilities, and courtyard improvement. Requests concerning public transportation, safety, and illegal dumping were less

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

numerous.

Hadoop delivered stable batch processing but struggled with latency in real-time tasks due to its disk-based model. Spark excelled in real-time analytics and iterative machine learning due to in-memory caching. The Hybrid system outperformed both, combining Hadoop's storage with Spark's speed-especially in multi-step workflows-delivering the fastest and most balanced performance across all tasks (Figure 2).

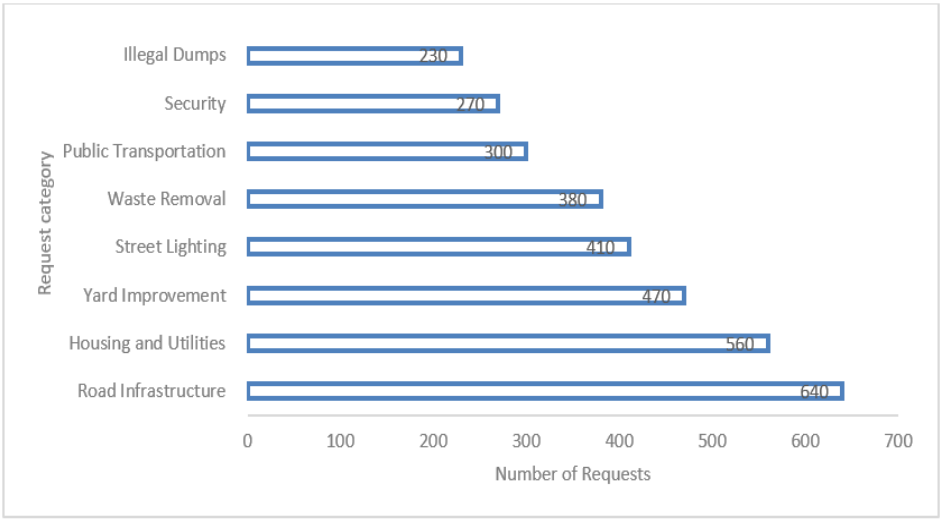


Fig. 2. Distribution of the synthetic request corpus by thematic categories (3,260 in total)

The next step involved comparing three classification models using the Accuracy, Precision, Recall, and F1-score metrics (Figure 3). The RuBERT class model demonstrated the highest quality: accuracy was 0.914, and F1-score was 0.907. Linear methods-logistic regression on TF-IDF features and linear SVM-were inferior to the language model, with the support vector machine being closer to it than logistic regression. The advantage of the transformer model is evident across all four metrics and is consistent with its ability to account for the context and semantics of short statements, whereas linear models rely primarily on the co-occurrence of individual words and are therefore less effective at distinguishing between similar but substantively different phrases. The balanced Precision and Recall scores of the best model indicate the absence of a pronounced bias toward certain categories. At the same time, the gap between the models is moderate, which is important to consider when choosing a solution for operation: linear methods, with significantly lower training and inference costs, remain a practically acceptable alternative where computing resources are limited (Figure 3).

For detailed analysis, Figure 4 shows the error matrix of the best model. Requests related to roads, housing and utilities, and yard improvement-categories with relatively stable and specific vocabulary-are recognized most reliably. Pairs of categories with semantic overlap proved more challenging: «garbage removal» and «illegal dumps,» as well as «lighting» and «safety.»

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License



Fig. 3. Comparison of request classification model quality

The nature of the errors suggests that the main source of inaccuracies is not the rarity of the classes themselves, but the closeness of their verbal expressions, with residents describing problems of varying severity using similar wording. The practical conclusion is that quality improvement should be sought not only by increasing the model's complexity but also by refining the subject taxonomy: combining difficult-to-distinguish categories, introducing a hierarchical classification, or adding clarifying features, such as the geolocation of the request.

Reducing initial processing time has a direct organizational impact. When request assignment to a category and department selection are performed automatically, the operator is freed up from the time previously spent reading and manually sorting each message; the operator is left primarily with uncertain and questionable cases, filtered by a threshold rule. All other things being equal, this increases front office throughput and smooths out peak loads when the number of requests increases. Importantly, however, the gains are not simply a matter of multiplying the average time by the number of requests: the actual effect depends on the proportion of automatically routed requests, the error rate and the cost of their correction, as well as the operator workflow. Therefore, the estimates provided should be considered as an indication of the potential for workload reduction.

The obtained estimates are consistent with the results of related studies. The achieved classification quality is at a level characteristic of transformer models on short Russian-language texts and confirms the feasibility of their use in the task of topic-based request routing. Unlike studies that consider the classifier as a standalone analytical module (Rizun et al., 2025; Zhou et al., 2025), in this study it is embedded in a managed service architecture, which allows us to link model quality with the applied effect-reduced initial processing time and reduced operator workload. A comparison

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

with approaches based on large language models and embeddings (Rakhimzhanov et al., 2025) shows that the selected RuBERT model provides a reasonable compromise between quality and cost for Russian-language request flow, although direct comparison requires a single corpus and a unified taxonomy.

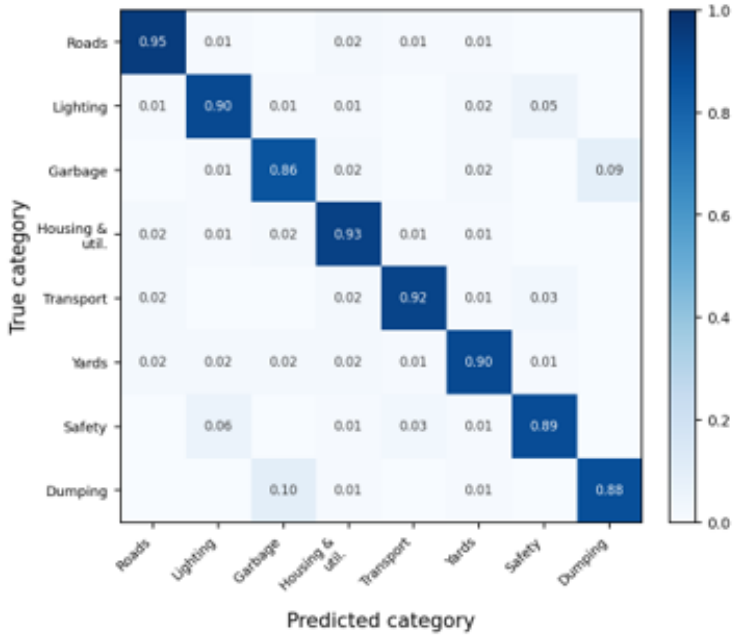


Fig. 4. Confusion matrix of the best classification model (row-normalized shares

The distribution of errors across categories is useful to compare with the corpus structure. The most common categories-roads, housing and utilities, and yard improvement-are recognized with high recall, as expected given the larger volume of training examples and more consistent vocabulary. Less-represented categories are generally classified less reliably, but the reduced recall is determined not only by their rarity but also by the overlap of wording with adjacent classes. For example, reports about illegal dumps are most often mistakenly attributed to garbage collection, and some safety reports are attributed to lighting, as residents often associate dark areas with a feeling of unsafety. This means that simply aligning the classes in the training set will not eliminate the majority of errors; refining category boundaries and incorporating additional report features appears more effective.

To evaluate the practical effect of the AI module, two processing scenarios were simulated: a completely manual route and a route with intelligent classification and automatic assignment of the responsible department. The average time to complete the main stages is shown in Figure 5. Under the simulated conditions, the average total time to transfer a request to the relevant department decreased from 26.15 to 6.85 minutes-more than 3.8 times. The distribution of savings across stages is uneven: the time to register a request remains virtually unchanged, while the main benefit is achieved at the initial analysis and routing stages, where automatic classification replaces manual

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

review, categorization, and department selection. This indicates that the value of the intelligent module is primarily due to the reduction of operators' workload on routine sorting, rather than to the acceleration of user data entry (Figure 5).

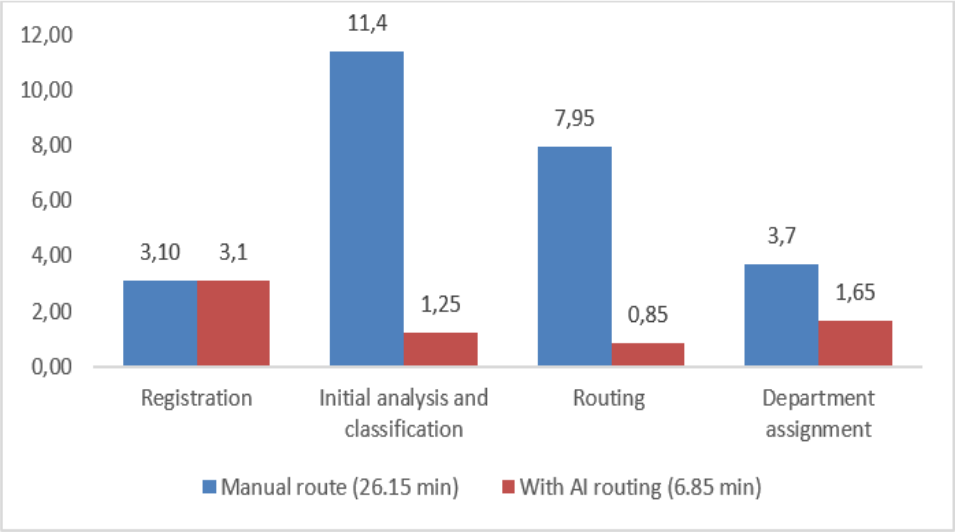


Fig. 5. Average request processing time by stage: manual and AI-assisted scenarios

The results should be interpreted considering the limitations of the study. First, only Russian-language material was considered, while the service supports a bilingual interface, and the classification quality of Kazakh-language queries requires separate evaluation. Second, the comparison is limited to three models and does not include modern approaches based on large language models and instructional embeddings. Third, the applied time estimates were obtained under model conditions, while the quantitative metrics from the service's load testing require separate presentation on a real infrastructure. These limitations limit the scope of the conclusions and simultaneously suggest directions for future work.

Conclusion.

This paper proposes the architecture of an intelligent Smart City service for receiving and routing citizen inquiries based on a Telegram bot using artificial intelligence methods. The following objectives were achieved: the service architecture and data model were designed; a request submission scenario was formalized using a finite-state machine; thematic text classification and a threshold routing algorithm were implemented; the service was deployed in a containerized environment and tested through functional, load, and usability testing; a comparative evaluation of classification models was conducted, and the impact of AI routing on initial processing time was assessed. Thus, the goal of the paper was achieved.

The results of the simulation experiment showed that the use of AI improves the quality of initial request classification and reduces routing time compared to manual processing: on the corpus, the RuBERT-class model achieved an accuracy of 0.914

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

and an F1-score of 0.907, while the average time to transfer a request to the relevant department decreased from 26.15 to 6.85 minutes. An error analysis revealed that the main inaccuracies were not due to class imbalances, but to the semantic proximity of individual categories, highlighting the role of subject-based taxonomy as a classification mechanism. It was also established that the choice of the model's confidence threshold influences the balance between routing accuracy and the level of service automation.

The scientific significance of this work lies in examining the Telegram bot as a fully-fledged component of the Smart City digital infrastructure and in integrating the dialog frontend, managed service architecture, and Russian-language request classification module into a single, reproducible model with experimental evaluation. The practical significance lies in the potential application of the proposed approach in creating municipal request processing services that can improve the availability of city services, reduce operator workload, and shorten routing times. Further research should focus on incorporating Kazakh-language and multimodal data (including classification based on attached photographs), comparison with approaches based on large language models, and evaluating the service's performance in production conditions, including load testing, external quality validation, and integration with the city's sensor data sources.

From a practical perspective, the proposed solution demonstrates that intelligent request triage can be implemented using an accessible technology stack and integrated into residents' familiar communication channels, without requiring them to learn a separate application. The manageability of the confidence threshold and the preservation of the operator's role in questionable cases make the approach applicable to real-world municipal practice, where the cost of erroneous routing is high. The obtained results confirm the feasibility of further developing the service from a prototype to production operation, subject to its validation on real data and as part of a broader city data collection and analysis system.

REFERENCES

Bono Rossello N., Simonofski A., Castiaux A. (2025). Artificial intelligence for digital citizen participation: Design principles for a collective intelligence architecture // Government Information Quarterly. — Vol. 42. — No. 2. Article 102020. DOI: https://doi.org/10.1016/j.giq.2025.102020. [In Eng.].

Devlin J., Chang M.-W., Lee K., Toutanova K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // Proceedings of NAACL-HLT 2019. Minneapolis. — Pp. 4171–4186. DOI: https://doi.org/10.18653/v1/N19-1423. [In Eng.].

Hemesath S., Tepe M. (2024). Public value positions and design preferences toward AI-based chatbots in e-government: Evidence from a conjoint experiment with citizens and municipal front desk officers // Government Information Quarterly. — Vol. 41. — No. 4. Article 101985. DOI: https://doi.org/10.1016/j.giq.2024.101985. [In Eng.].

Joachims T. (1998). Text categorization with support vector machines: Learning with many relevant features // Proceedings of the 10th European Conference on Machine Learning (ECML-98). Berlin: Springer. — Pp. 137–142. DOI: https://doi.org/10.1007/BFb0026683. [In Eng.].

König P.D. (2021). Citizen-centered data governance in the smart city: From ethics to accountability // Sustainable Cities and Society. — Vol. 75. — Article 103308. DOI: https://doi.org/10.1016/j.scs.2021.103308. [In Eng.].

Kuratov Y., Arkhipov M. (2019). Adaptation of Deep Bidirectional Multilingual Transformers for Russian Language // arXiv preprint arXiv:1905.07213. DOI: https://doi.org/10.48550/arXiv.1905.07213. [In Eng.].

Madyatmadja E.D., Sianipar C.P.M., Wijaya C., Sembiring D.J.M. (2023). Classifying Crowdsourced

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Citizen Complaints through Data Mining: Accuracy Testing of k-Nearest Neighbors, Random Forest, Support Vector Machine, and AdaBoost // Informatics. — Vol. 10. — No. 4. Article 84. DOI: https://doi.org/10.3390/informatics10040084. [In Eng.].

Minaee S., Kalchbrenner N., Cambria E., Nikzad N., Chenaghlu M., Gao J. (2021). Deep Learning Based Text Classification: A Comprehensive Review // ACM Computing Surveys. — Vol. 54. — No. 3. Article 62. DOI: https://doi.org/10.1145/3439726. [In Eng.].

Panahi Rizi M.H., Hosseini Seno S.A. (2022). A systematic review of technologies and solutions to improve security and privacy protection of citizens in the smart city // Internet of Things. — Vol. 20. — Article 100584. DOI: https://doi.org/10.1016/j.iot.2022.100584. [In Eng.].

Parycek P., Schmid V., Novak A.-S. (2024). Artificial Intelligence (AI) and Automation in Administrative Procedures: Potentials, Limitations, and Framework Conditions // Journal of the Knowledge Economy. — Vol. 15. — No. 2. — Pp. 8390–8415. DOI: https://doi.org/10.1007/s13132-023-01433-3. [In Eng.].

Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., et al. (2011). Scikit-learn: Machine Learning in Python // Journal of Machine Learning Research. — Vol. 12. — Pp. 2825–2830. DOI: https://doi.org/10.48550/arXiv.1201.0490. [In Eng.].

Rakhimzhanov D., Belginova S., Yedilkhan D. (2025). Automated Classification of Public Transport Complaints via Text Mining Using LLMs and Embeddings // Information. — Vol. 16. —No. 8. Article 644. DOI: https://doi.org/10.3390/info16080644. [In Eng.].

Rizun N., Revina A., Edelmann N. (2025). Text analytics for co-creation in public sector organizations: a literature review-based research framework // Artificial Intelligence Review. — Vol. 58. — No. 4. Article 125. DOI: https://doi.org/10.1007/s10462-025-11112-1. [In Eng.].

Senadheera S., Yigitcanlar T., Desouza K.C., Mossberger K., Corchado J., Mehmood R., Li R.Y.M., Cheong P.H. (2024). Understanding Chatbot Adoption in Local Governments: A Review and Framework // Journal of Urban Technology. — Vol. 32. — No. 3. — Pp. 35–69. DOI: https://doi.org/10.1080/10630732.2023.2297665 . [In Eng.].

Shama F., Aziz A., Deya L.B.M. (2024). CitySolution: A complaining task distributive mobile application for smart city corporation using deep learning // SoftwareX. DOI: https://doi.org/10.1016/j.softx.2024.101829. [In Eng.].

Shen C., Xu Y., Yuan Z. (2025). Digital dialogue in smart cities: Evidence from public concerns, government responsiveness, and citizen satisfaction in China // Cities. — Vol. 158. — Article 105717. DOI: https://doi.org/10.1016/j.cities.2025.105717. [In Eng.].

Sun C., Qiu X., Xu Y., Huang X. (2019). How to Fine-Tune BERT for Text Classification? // Chinese Computational Linguistics (CCL 2019). Cham: Springer. — Pp. 194–206. DOI: https://doi.org/10.1007/978-3-030-32381-3_16. [In Eng.].

Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., Polosukhin I. (2017). Attention Is All You Need // Advances in Neural Information Processing Systems. — Vol. 30. — Pp. 5998–6008. [In Eng.].

Zhang J., Geng Q., Jin J. (2025). EKLI-Attention: An integrated attention mechanism for classifying citizen requests in government–citizen interactions // Information Processing & Management. — Vol. 62. — No. 6. Article 104237. DOI: https://doi.org/10.1016/j.ipm.2025.104237. [In Eng.].

Zhou J., Hu H. (2025). Citizens' appeals classification using a mixed method: A case study of Yichang government public service hotline data // Information Development. OnlineFirst. DOI: https://doi.org/10.1177/02666669241312816. [In Eng.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 66–77
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.004

YOLO-CBAM: A LIGHTWEIGHT NEURAL NETWORK MODEL WITH INTEGRATED ATTENTION MECHANISM FOR ROAD DAMAGE ANALYSIS

O.M. Olzhayev[1]*, B.S. Omarov[1], A. Suliman[2]
[1]International Information Technology University, Almaty, Kazakhstan;
[2]Asia Metropolitan University, Subang Jaya Campus, Malaysia.
E-mail: o.olzhayev@iitu.edu.kz

Olzhas M. Olzhayev — PhD student, senior lecturer of the Department «Mathematical and Computer Modeling», International University of Information Technologies, Almaty, Kazakhstan
E-mail: o.olzhayev@iitu.edu.kz, https://orcid.org/0009-0001-9930-8990;
Batyrkhan S. Omarov — PhD, research professor of the Department «Mathematical and Computer Modeling», International University of Information Technologies, Almaty, Kazakhstan
https://orcid.org/0000-0002-8341-7113;
Azizah Suliman — PhD, PhD professor, Asia Metropolitan University, Subang Jaya Campus, Malaysia
https://orcid.org/0009-0002-2064-3516.

© O.M. Olzhayev, B.S. Omarov, A. Suliman

Abstract. In this paper, we propose YOLO-CBAM, a fast and lightweight neural network for real-time road surface defect detection. We improved the standard YOLO architecture by adding a convolutional block attention module (CBAM). This attention mechanism helps the model focus on complex road defects, such as cracks, and ignore background visual noise. Experiments were conducted to test how different weight learning methods affect the model training process. The results show that using partial transfer learning helps the network quickly overcome early errors and achieve an average accuracy (mAP@50) of 0.6 in just 40 epochs. Consequently, the YOLO-CBAM model improves defect detection without incurring high computational costs, making it a highly effective solution for video surveillance systems in smart cities.

Keywords: Computer vision, deep learning, object detection, YOLO, CBAM, attention mechanism, road monitoring, transfer learning

For citation: Olzhayev O.M., Omarov B.S., Suliman A. (2026). YOLO-CBAM: a lightweight neural network model with integrated attention mechanism

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

for road damage analysis // International journal of information and communication technologies. Vol. 7. No. 27. Pp. 66–77. https://doi.org/10.54309/IJICT.2026.27.3.004. (In Russ.).

Conflict of interest: The authors declare that there is no conflict of interest.

YOLO-CBAM: ЖОЛДЫҢ ЗАҚЫМДАНУЫН ТАЛДАУҒА АРНАЛҒАН ИНТЕГРАЦИЯЛАНҒАН ЗЕЙІН МЕХАНИЗМІ БАР ЖЕҢІЛ САЛМАҚТЫ НЕЙРОНДЫҚ ЖЕЛІ МОДЕЛІ

О.М. Олжаев[1]*, Б.С. Омаров[1], A.Suliman[2]

[1]Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан;
[2]Asia Metropolitan University, Субанг-Джая кампусы, Малайзия.
E-mail: o.olzhayev@iitu.edu.kz

Олжаев Олжас Мұратұлы — PhD докторант, «Математикалық және компьютерлік модельдеу» кафедрасының сениор-лекторы, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан
E-mail: o.olzhayev@iitu.edu.kz, https://orcid.org/0009-0001-9930-8990;
Омаров Батырхан Султанович — PhD, «Математикалық және компьютерлік модельдеу» кафедрасының профессор-зерттеушісі, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан
https://orcid.org/0000-0002-8341-7113;
Suliman Azizah — философия докторы (PhD), профессор, Asia Metropolitan University, Субанг-Джая кампусы, Малайзия
https://orcid.org/0009-0002-2064-3516.

© О.М. Олжаев, Б.С. Омаров, A.Suliman

Аннотация. Бұл мақалада біз YOLO-cbam – нақты уақыт режимінде жол жамылғысының зақымдануын анықтауға арналған жылдам және жеңіл нейрондық желіні ұсынамыз. Біз стандартты YOLO архитектурасын конволюциялық блок зейін модулін (CBAM) қосу арқылы жақсарттық. Бұл назар аудару механизмі модельге жолдағы жарықтар сияқты күрделі ақауларға назар аударуға және фондық визуалды шуды елемеуге көмектеседі. Эксперименттер барысында таразыны оқытудың әртүрлі әдістері модельді оқыту процесіне қалай әсер ететіні туралы тесттер жүргізілді. Нәтижелер жартылай оқытуды пайдалану желіге ерте қателіктерді тез жеңуге және небәрі 40 дәуірде орташа дәлдікке (mAP@50) 0.6 жетуге көмектесетінін көрсетеді. Демек, Yolo-CBAM моделі үлкен есептеу шығындарынсыз ақауларды анықтауды жақсартады, осыған байланысты оны «ақылды» қалалардағы бейнебақылау жүйелері үшін жоғары тиімді шешім етеді.

Түйінді сөздер: компьютерлік көру, терең оқыту, нысандарды анықтау, YOLO, CBAM, назар аудару механизмі, жол мониторингі, трансферлік оқыту

Дәйексөздер үшін: Олжаев О.М., Омаров Б.С., Suliman A. (2026). YOLO-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

CBAM: жолдың зақымдануын талдауға арналған интеграцияланған зейін механизмі бар жеңіл салмақты нейрондық желі моделі // Халықаралық ақпараттық және коммуникалық технологиялар журналы. Т. 7. No 27. Б. 66–77. https://doi.org/10.54309/IJICT.2026.27.3.004. (Орыс. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

YOLO-CBAM: ЛЕГКОВЕСНАЯ НЕЙРОСЕТЕВАЯ МОДЕЛЬ С ИНТЕГРИРОВАННЫМ МЕХАНИЗМОМ ВНИМАНИЯ ДЛЯ АНАЛИЗА ДОРОЖНЫХ ПОВРЕЖДЕНИЙ

О.М. Олжаев[1]*, Б.С. Омаров[1], A.Suliman[2]

[1]Международный университет информационных технологий, Алматы, Казахстан;

[2]Asia Metropolitan University, кампус Субанг-Джая, Малайзия.

E-mail: o.olzhayev@iitu.edu.kz

Олжаев Олжас Мұратұлы — докторант PhD, сениор-лектор кафедры «Математическое и компьютерное моделирование», Международный университет информационных технологий, Алматы, Казахстан
E-mail: o.olzhayev@iitu.edu.kz, https://orcid.org/0009-0001-9930-8990;

Омаров Батырхан Султанович — PhD, профессор-исследователь кафедры «Математическое и компьютерное моделирование», Международный университет информационных технологий, Алматы, Казахстан
https://orcid.org/0000-0002-8341-7113;

Suliman Azizah — доктор философии (PhD), профессор, Asia Metropolitan University, кампус Субанг-Джая, Малайзия
https://orcid.org/0009-0002-2064-3516.

© О.М. Олжаев, Б.С. Омаров, A.Suliman

Аннотация. В данной статье мы предлагаем YOLO-CBAM – быструю и легковесную нейронную сеть для обнаружения повреждений дорожного покрытия в режиме реального времени. Мы улучшили архитектуру стандартного YOLO, добавив модуль внимания сверточных блоков (CBAM). Данный механизм внимания помогает модели сосредотачиваться на сложных дефектах, такие как трещины, на дороге и игнорировать фоновый визуальный шум. В ходе экспериментов были проведены тесты, как различные методы обучения весов влияют на процесс обучения модели. Результаты показывают, что использование частичного переноса обучения помогает сети быстро преодолевать ранние ошибки и достигать средней точности (mAP@50) 0.6 всего за 40 эпох. Следовательно, YOLO-CBAM модель улучшает обнаружение дефектов без больших вычислительных затрат, в связи с этим делает ее высокоэффективным решением

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

для систем видеонаблюдения в «умных» городах.

Ключевые слова: компьютерное зрение, глубокое обучение, обнаружение объектов, YOLO, CBAM, механизм внимания, мониторинг дорог, трансферное обучение

Для цитирования: Олжаев О.М., Омаров Б.С., Suliman A. (2026). YOLO-CBAM: легковесная нейросетевая модель с интегрированным механизмом внимания для анализа дорожных повреждений // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 27. Стр. 66–77. https://doi.org/10.54309/IJICT.2026.27.3.004. (На русс.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Введение.

Дорожные сети страны является основой для экономики и общественной безопасности, которое имеет огромное значение. Как показывает работа (Safyari et al., 2024), со временем интенсивное движение и суровые погодные условия приводят к появлению дефектов, такие как трещин, выбоин и другие повреждений на дорогах. Ручной мониторинг дорог занимает много времени, обходится дорого и может быть опасна для рабочих (Guerrieri et al., 2024). В следствии активно ищутся автоматизированные способы мониторинга состояния дорог многими городами с помощью видео камер и алгоритмами искусственного интеллекта.

В последние годы стандартными инструментами для поиска объектов на изображениях является компьютерное зрение и глубокое обучение (LeCun et al., 2015). Популярным решением для этой задачи является одноэтапные детекторы объектов, в частности YOLO (You Only Look Once) (Redmon et al., 2016). Быстрота, легкость и требования меньше вычислительных ресурсов делает их идеальными для обработки видео в режиме реального времени на небольших устройствах, установленных в автомобилях (Wu et al., 2023).

Однако детекция повреждений дефектов на дорогах в реальном мире по-прежнему является сложной задачей (Zhang et al., 2025). Стандартные нейронные сети дают хорошие результаты в яркие, ясные дни, но часто дают ошибочные результаты при изменении погоды или освещения. К примеру, тени в вечернее время, пасмурное небо или отражения воды в дождливую погоду могут легко запутать модель (Yang et al., 2026). Тем самым, модели стандартного YOLO пропускают маленькие дефекты, как трещины, или делают ложные предсказания, из-за визуального фонового шума.

В качестве решения этой проблемы исследователями были применены механизмы внимания в нейронные сети (Vaswani et al., 2017). Модуль внимания сверточных блоков (CBAM) является одним из наиболее эффективных решений (Woo et al., 2018). Он является интеллектуальным фильтром. Он контролирует сеть и указывает, куда смотреть на изображении и на какие признаки обращать внимание. В результате модель может игнорировать фоновый шум и фокусиро-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ваться только на физических дефектах на дороге (Liang et al., 2022).

В данной статье нами были представлены модифицированный и облегченный YOLO-CBAM модель, разработанную специально для обнаружения дефектных повреждений на дорогах (Zhang et al., 2025). Интеграция модуля CBAM в стандартную архитектуру YOLO привело к повышению точности в сложных условиях. Нами были проведены также тесты, как различные стратегии трансферного обучения влияют на стабильность обучения модели и ее конечную производительность (Zhuang et al., 2020). Главной целью данной работы является получение быстрого и надежного инструмента обнаружения, который может мониторить инфраструктуры умного города.

Материалы и методы.

Подготовка набора данных

Мы используем общедоступный набор данных “Road Damage” с платформы Kaggle для обучения и оценки предложенной нейронной сети (Road Damage, 2020). Датасет содержит тысячи изображений разных классов, демонстрирующих различные типы повреждений асфальта в реальных условиях. Данные разделены на несколько основных классов, которые наиболее критичны для дорожного покрытия, к примеру продольные трещины, поперечные трещины, трещины типа “крокодиловая кожа” и выбоины.

Все изображения прошли стандартную предварительную обработку перед подачей данных в нейронную сеть. Изображения были приведены к общему разрешению 640×640 пикселей, это соответствует к требованиям к входным данным архитектуры YOLO. Датасет был случайным образом разделен на обучающую и валидационную выборки в соотношении 80% и 20% соответственно.

Архитектура YOLO-CBAM

Нами была выбрана YOLO11s архитектура в качестве основой предлагаемой системой. Она является легковесным и одноэтапным детектором объектов и обеспечивает оптимальный баланс между средней точностью и скоростью вычислений. Стандартные сверточные сети в большинстве случаев испытывают определенные трудности с выделением тонких, неровных трещин на дорогах из-за сложного фона и сложных форм. В связи с этим мы интегрировали модуль внимания сверточных блоков (CBAM) в основу YOLO для преодоления этого ограничения, создав архитектуру YOLO-CBAM. Общая структура предлагаемой модели показана на рисунке 1.

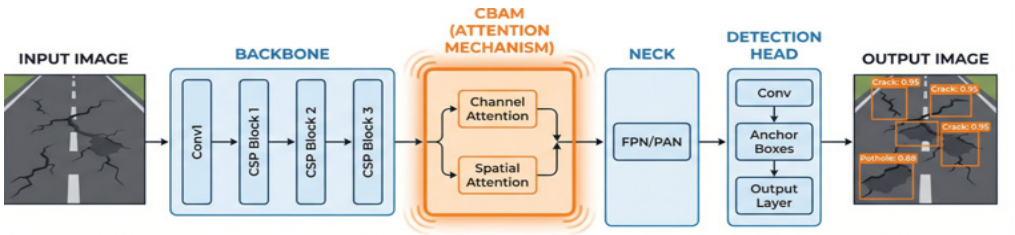


Рис. 1. Архитектура предлагаемой модели YOLO-CBAM

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Обработанный и нормализованный рисунок с повреждением в виде тензора попадает на Backbone зону, где выявляются признаки из картин. Дальше применяется наш CBAM модуль для того, чтобы модель могла точно оценивать мелкие дефекты. В конце формируется результаты в виде ответа, где указывается границы дефекта и возможная ее вероятность. Внутренняя структура механизма внимания подробно описаны на рисунке 2. Модуль CBAM последовательно обрабатывает промежуточные карты признаков. В первую очередь подмодуль внимания используется как фильтр к каналам, находя каналы признаков, которые наиболее информативны. Потом подмодуль пространственного внимания уточняет полученные результаты, чтобы точно определить границы дефектов дороги.

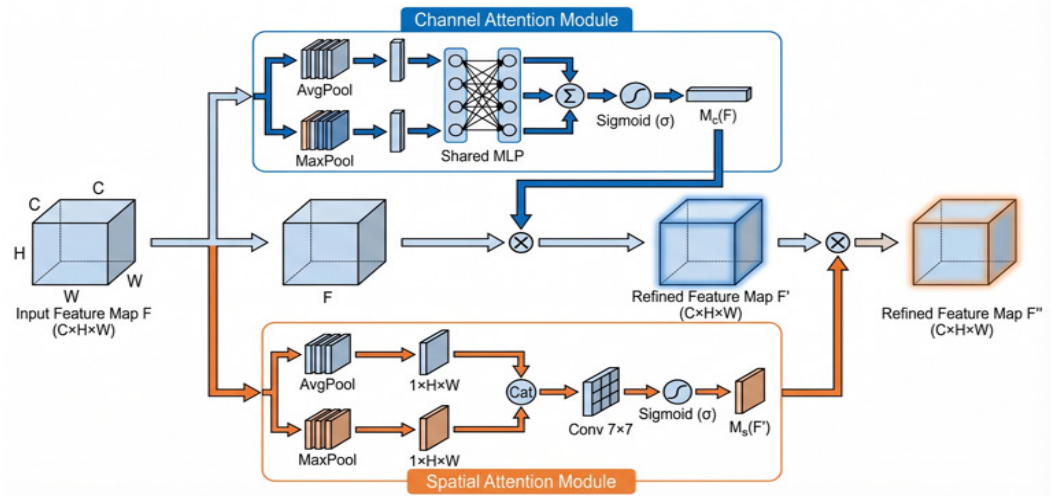


Рис. 2. Подробное описание рабочего процесса модуля внимания на основе сверточных блоков (CBAM)

CBAM является механизмом двойного внимания, последовательно уточняя промежуточные карты признаков. При наличии входной карты признаков F общий процесс внимания вычисляется по следующим формулам:

$$F' = M_c(F) \otimes F \quad (1)$$

$$F'' = M_s(F') \otimes F' \quad (2)$$

где $M_c(F)$ – карта внимания канала, $M_s(F')$ – карта пространственного внимания, $\otimes$ – поэлементное умножение.

Подмодуль внимания к каналу

В данном этапе определяется важные признаки. Данный модуль сжимает пространственные измерения и использует операции усреднения и максимального пулинга, в результате получаем два сгенерированных дескриптора пространственного контекста. Затем эти дескрипторы передаются в общую многослойную перцептронную сеть (MLP). Данная операция математически выражается следующим образом:

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \quad (3)$$

где σ – сигмоидная функция активации, которая масштабирует весовые коэффициенты важности в диапазоне от 0 до 1.

Подмодуль пространственного внимания

После того, как внимание к каналам определяет важные карты признаков, начинается этап определение местоположение дефектов через пространственное внимание. Применяется здесь операция пулинга вдоль оси канала и соединяет их с помощью сверточного слоя для генерации пространственной карты:

$$M_s(F') = \sigma(f^{7\times7}([AvgPool(F'); MaxPool(F')])) \quad (4)$$

где $f^{7\times7}$ – операция свертки с размером фильтра фильтра 7 × 7. С помощью данного шага эффективно подавляется фоновые шумы, такте как дорожная разметка или тени, и сеть фокусируется исключительно на структурных краях выбоин и трещин.

Экспериментальная установка и параметры обучения

Эксперименты были проводены на высокопроизводительной компьютере под управлением Windows, с графическом процессором NVIDIA RTX 4070 Ti с 12 ГБ видеопамяти, с использованием фреймворка Ultralytics и PyTorch.

Был выбран строго стандартизированный процесс обучения для всех моделей, чтобы обеспечить справедливое сравнение. Использовались следующие гиперпараметры и настройки:

- Оптимизация: использовался оптимизатор AdamW с начальной скоростью обучения (lr0) 0,005 и конечным коэффициентом скорости обучения (lrf) 0,01.

- Функции потерь: задача классификации оптимизировалась с использованием функции потерь кросс-энтропии, а для точности локализации использовались функции потерь Box и Distribution Focal (DFL).

- Разрешение изображений: все входные изображения были изменены до единого разрешения 640×640 пикселей.

- Конфигурация пакетной обработки: обучение проводилось с размером пакета (batch size) 32 и общей продолжительностью 40 эпох.

- Регуляризация: для предотвращения переобучения мы применили коэффициент затухания весов 0,0005 и использовали период прогрева в 3,0 эпохи.

Кроме того, нами была использована гибридная стратегия обучения. Сверточная YOLO11s-CBAM архитектура была инициализирована с использованием предварительно обученных весов из стандартной yolo11s.pt модели для использования априорных знаний. В отличие от этого, вновь интегрированные слои CBAM были инициализированы случайными весами. В результате такой подход позволил модели поддерживать стабильное извлечение признаков, адаптируя механизм внимания специально к дефектам дорожного покрытия.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Результаты и обсуждение.

Динамика тренировок и стабильность градиента

Анализ кривых обучения выявляет важные закономерности в том, как сеть адаптируется к новым структурным слоям. В ходе эксперимента по частичному переносу обучения мы наблюдали значительный начальный всплеск функции потерь классификации (val/cls_loss > 10,0) в течение первой эпохи, что детально можно увидит на рисунке 3.

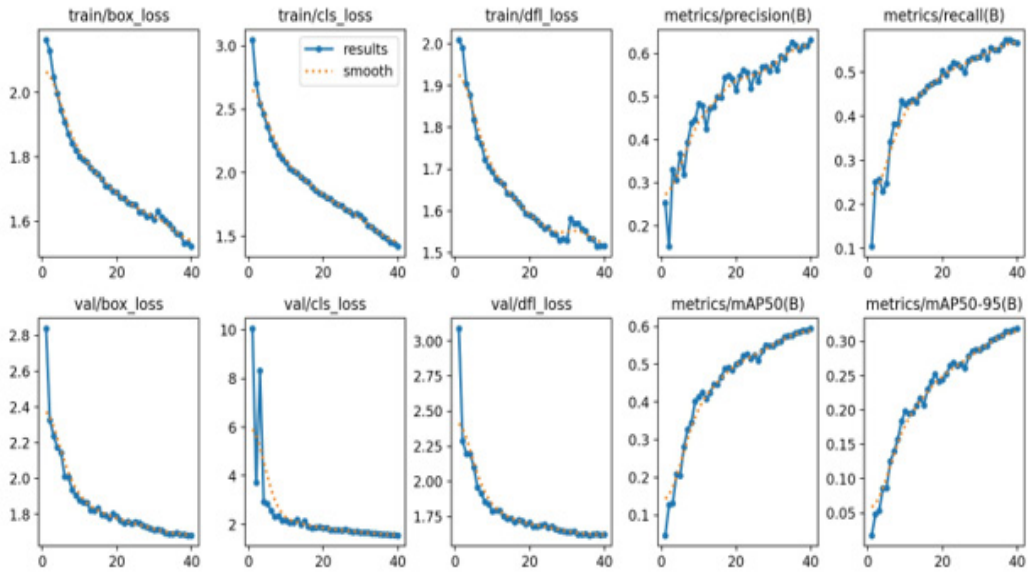


Рис. 3. Кривые обучения, показывающие показатели потерь и mAP на протяжении 40 эпох.

Этот эффект известен как градиентный шок. Он возникает из-за того, что предварительно обученная сверточная нейронная сеть и слой CBAM со случайной инициализацией временно не синхронизированы. Однако предложенная архитектура продемонстрировала высокую стабильность градиента. Сеть быстро преодолела этот шок, стабилизировав ошибку к 5-й эпохе. Это подтверждает, что интеграция модуля внимания не нарушает общую геометрию признаков архитектуры YOLO.

Для более глубокого понимания точности классификации модели для различных типов дефектов дорожного покрытия мы проанализировали F1-меру и нормализованную матрицу ошибок. F1-мера представляет собой единый показатель, уравновешивающий точность и полноту. Предложенная нами архитектура YOLO11s-CBAM с частичным переносом обучения достигла F1-меры 0,60, что указывает на стабильное равновесие между минимизацией ложных срабатываний (например, обнаружение тени как трещины) и ложных отрицаний (пропуск фактической выбоины). Кроме того, матрица ошибок позволяет оценить производительность для каждого класса.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

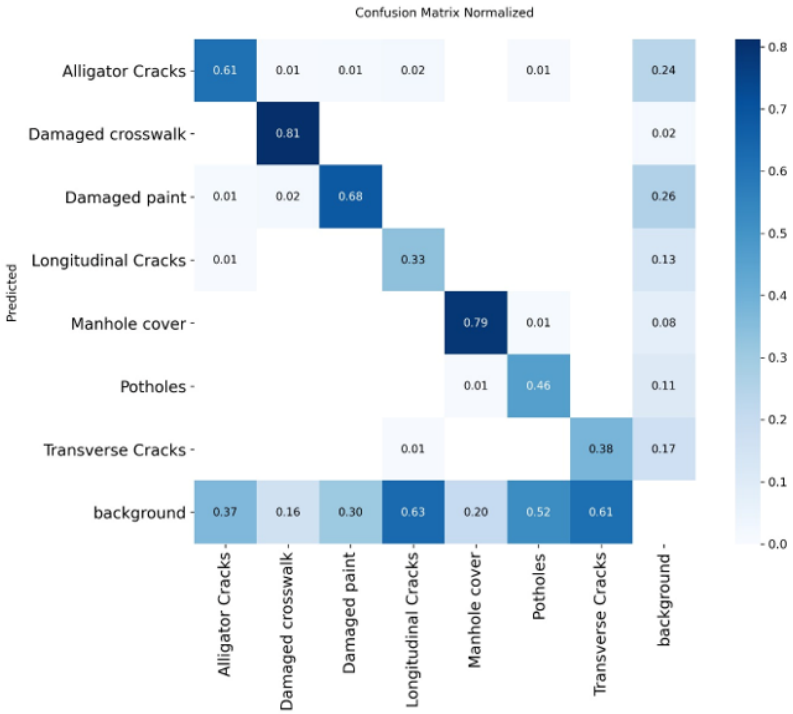


Рис. 4. Нормализованная матрица ошибок модели YOLO11s-CBAM

Как видно на рисунке 4, модель демонстрирует высокую степень уверенности в обнаружении различных структурных повреждений, таких как выбоины. Однако иногда возникают незначительные ошибки классификации между визуально похожими категориями, такими как продольные и поперечные трещины. Интеграция модуля CBAM помогла значительно снизить количество ложных срабатываний фонового сигнала по сравнению со стандартными моделями, поскольку фоновый класс демонстрирует минимальную путаницу с фактическими повреждениями дорог.

Исследование абляции

Для оценки эффективности предложенной архитектуры YOLO-CBAM мы провели прямое сравнение со стандартной базовой моделью YOLO11s. Все эксперименты проводились с использованием набора данных Kaggle о повреждениях дорог в течение 40 эпох обучения. Чтобы понять, как обучается модуль внимания, мы провели исследование с использованием двух различных стратегий инициализации: обучение с нуля и частичное перенос обучения. Количественные результаты на валидационном наборе представлены в таблице 1.

Базовая модель YOLO11s продемонстрировала стабильную производительность, достигнув mAP@50 0,60. Интересно, что когда модифицированная модель YOLO11s-CBAM была обучена полностью с нуля (со случайными весами), она достигла mAP@50 0,55 всего за 40 эпох. Эта быстрая сходимость указывает на то, что модуль CBAM эффективно извлекает важные структурные

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

особенности из асфальта даже без предварительных знаний. Наконец, применение частичного трансферного обучения к модели YOLO11s-CBAM позволило ей сравняться с базовой моделью по общей точности, немного улучшив при этом метрики mAP@50-95 и Recall.

Таблица 1. Влияние архитектурных изменений и стратегий инициализации на точность обнаружения.

Архитектура	Стратегия обучения	mAP@50	mAP@50-95	Precision (P)	Recall (R)
YOLO11s (базовый уровень)	Переносное обучение	0.6	0.3	0.63	0.55
YOLO11s-CBAM	Обучение с нуля	0.55	0.29	0.59	0.53
YOLO11s-CBAM	Частичное переносное обучение	0.6	0.32	0.63	0.57

Карты внимания

Наконец, для качественной оценки предложенной архитектуры мы визуализировали карты внимания сети. Основная проблема в мониторинге инфраструктуры – это шум окружающей среды. Например, видео, снятые ранним вечером или во время дождя, часто содержат длинные тени и отражения воды, которые сбивают с толку стандартные детекторы.

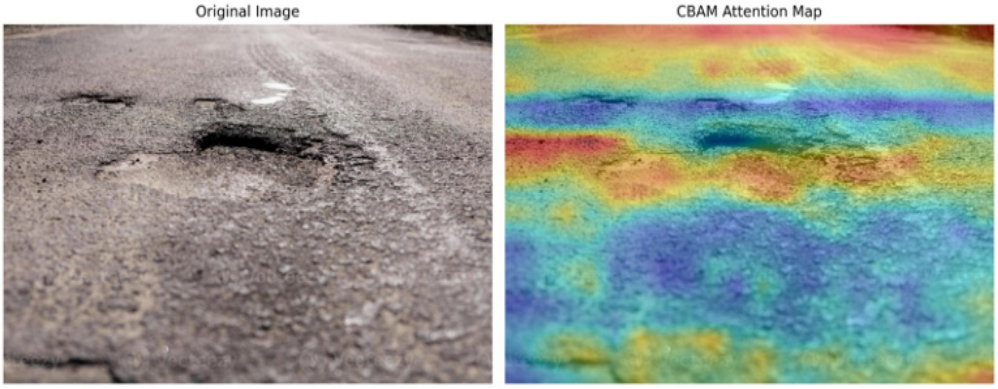


Рис. 5. Визуализация карт внимания CBAM для обнаружения повреждении

На рисунке 5 показано сравнение исходного изображения с дефектами и карты внимания, сгенерированной нашим пользовательским слоем CBAM. Визуализация демонстрирует, что наивысший фокус сети (красные области) точно определяет структурные края дефектов дороги. При этом фоновая поверхность дороги, включая нерелевантные тени, остается без внимания модели (синие области). Это визуальное доказательство подтверждает, что механизм внимания пространственного канала эффективно подавляет фоновый шум, заставляя сеть концентрироваться исключительно на фактической деградации дорожного покрытия.

Производительность в реальном времени

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Помимо точности, возможность развертывания YOLO11s-CBAM на мобильных периферийных устройствах зависит от скорости обработки данных. Мы протестировали модель на графическом процессоре NVIDIA RTX 4070 Ti, чтобы оценить ее возможности в реальном времени. Модель достигла средней скорости обработки данных примерно 38,7 мс на изображение, что соответствует примерно 25–26 кадрам в секунду (FPS). Этой производительности достаточно для обработки в реальном времени видеопотоков высокого разрешения с камер, установленных на транспортных средствах и движущихся в городских условиях на высокой скорости. Несмотря на добавление слоев CBAM, увеличение GFLOPs было незначительным, сохраняя «легковесный» профиль, необходимый для периферийных вычислений. Объем используемой памяти оставался в пределах 12 ГБ видеопамяти, что говорит о возможности дальнейшей оптимизации модели для еще более ограниченного оборудования, такого как серия NVIDIA Jetson, без существенной потери точности обнаружения.

Заключение.

В данном исследовании представлена модифицированная архитектура нейронной сети YOLO-CBAM, разработанная для автоматического обнаружения повреждений дорожного покрытия. Интеграция модуля внимания сверточных блоков (CBAM) в структуру YOLO11s позволила повысить способность модели извлекать важные структурные признаки, подавляя при этом окружающий шум.

Результаты экспериментов и последующий анализ привели к следующим ключевым выводам:

- Интеграция механизмов пространственного и канального внимания позволила модели точно фокусироваться на дефектах асфальта. Качественный анализ с использованием тепловых карт подтвердил, что архитектура YOLO-CBAM эффективно игнорирует визуальные отвлекающие факторы, такие как тени и дорожная разметка, которые являются распространенными причинами ложных срабатываний в стандартных детекторах.

- Несмотря на первоначальный «градиентный шок», наблюдаемый в течение первой эпохи обучения, модель продемонстрировала высокую стабильность. Использование стратегии частичного переноса обучения оказалось весьма эффективным, позволив сети быстро сходиться и достичь средней точности (mAP@50) 0.60 за короткий цикл обучения в 40 эпох на графическом процессоре NVIDIA RTX 4070 Ti.

- Предложенная модель имеет облегченный профиль, что делает ее подходящей для выполнения вычислений в реальном времени на периферийных вычислительных устройствах. Благодаря показателю F1-меры 0.60 и сбалансированной матрице ошибок, YOLO-CBAM обеспечивает надежное решение для мониторинга инфраструктуры в условиях умного города, даже при изменяющихся условиях освещения и погоды.

В будущих исследованиях основное внимание будет уделено расширению набора данных для включения более экстремальных погодных сценариев и дальнейшей оптимизации слоев внимания для повышения точности обнаружения

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

микротрещин. В целом, модель YOLO-CBAM представляет собой значительный шаг к полностью автономным и экономически эффективным системам обслуживания дорог.

REFERENCES

Guerrieri, M., Parla, G., Khanmohamadi, M., & Neduzha, L. (2024). Asphalt pavement damage detection through deep learning technique and cost-effective equipment: A case study in urban roads crossed by tramway lines. Infrastructures. — Vol. 9. — No. 2. — Pp. 34. [In Eng.].

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature. — Vol. 521. — No. 7553. — Pp. 436–444. [In Eng.].

Liang, H., Lee, S. C., & Seo, S. (2022). Automatic recognition of road damage based on lightweight attentional convolutional neural network // Sensors. — Vol. 22. — No. 24. — Pp. 9599. [In Eng.].

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection // In Proceedings of the IEEE conference on computer vision and pattern recognition. — Pp. 779–788. [In Eng.].

Road Damage, https://www.kaggle.com/datasets/alvarobasily/road-damage/data.[In Eng.].

Safyari, Y., Mahdianpari, M., & Shiri, H. (2024). A review of vision-based pothole detection methods using computer vision and machine learning // Sensors. — Vol. 24. — No. 17. — Pp. 5652. [In Eng.].

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need // Advances in neural information processing systems. — Pp. 30. [In Eng.].

Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module // In Proceedings of the European conference on computer vision (ECCV). — Pp. 3–19. [In Eng.].

Wu, C., Ye, M., Zhang, J., & Ma, Y. (2023). YOLO-LWNet: A lightweight road damage object detection network for mobile terminal devices // Sensors. — Vol. 23. — No. 6. — Pp. 3268. [In Eng.].

Yang, H., Song, Y., Liang, Y., Tang, E., & Cao, D. (2026). SDC-YOLOv8: An Improved Algorithm for Road Defect Detection Through Attention-Enhanced Feature Learning and Adaptive Feature Reconstruction // Sensors. — Vol. 26. — No. 2. — Pp. 609. [In Eng.].

Zhang, S., Liu, Z., Wang, K., Huang, W., & Li, P. (2025). OBC-YOLOv8: an improved road damage detection model based on YOLOv8 // PeerJ Computer Science. — Vol. 11. — e2593. [In Eng.].

Zhang, S., Liu, Z., Wang, K., Huang, W., & Li, P. (2025). OBC-YOLOv8: an improved road damage detection model based on YOLOv8 // PeerJ Computer Science. — Vol. 11. — e2593. [In Eng.].

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., ... & He, Q. (2020). A comprehensive survey on transfer learning // Proceedings of the IEEE. — Vol. 109. — No. 1. — Pp. 43–76. [In Eng.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 78–100
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.005
IRSTI / ҒТАХР / МРНТИ / 13.61.08; 13.61.11

OVERCOMING FRAGMENTATION IN THE CLASSIFICATION OF DIGITAL CULTURAL HERITAGE: AN INTEGRATIVE MULTIDIMENSIONAL MODEL

K. Yermagambetova[1]*, A. Sarsekeyeva[2], Y. Akhlassov[2]
[1]L.N. Gumilyov Eurasian National University, Astana, Kazakhstan;
[2]Research Institute «Almaty gen plan», Almaty, Kazakhstan.
E-mail: quralai.sergabylqyzy@mail.ru

Kuralay Yermagambetova — PhD, Associate Professor of the Department of Philosophy, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
E-mail: quralai.sergabylqyzy@mail.ru, https://orcid.org/0000-0002-1620-5794;
Aysulu Sarsekeyeva — Senior researcher of the Limited Liability Partnership «Research Institute «Almaty gen plan», Almaty, Kazakhstan
https://orcid.org/0009-0002-9952-1596;
Yelaman Akhlassov — Junior researcher of the Limited Liability Partnership «Research Institute «Almaty gen plan», Almaty, Kazakhstan
https://orcid.org/0009-0007-2616-3042.

© K. Yermagambetova, A. Sarsekeyeva, Y. Akhlassov

Abstract. The rapid development of digitisation, three-dimensional modelling, virtual and augmented reality, digital archiving, and platform-based storage has significantly expanded the scope of digital cultural heritage. Contemporary digital heritage includes resources created directly in digital environments, such as websites, digital art, virtual museums, video games, and spatial models. However, existing classification approaches remain fragmented, and no unified analytical framework simultaneously addresses the cultural, technical, functional, and institutional dimensions of digital heritage. This study develops an integrative multidimensional classification model based on a conceptual analysis of international normative documents, recent scholarship on heritage studies and digital infrastructures, and a comparative review of existing classificatory, informational, and functional models. The proposed model comprises five interrelated dimensions and enables a single digital object to be classified across multiple parameters. This multidimensional approach provides a more comprehensive framework for analysing cultural functions, institutional affiliation, social significance, and the long-term sustainability of digital cultural heritage.

Keywords: digital cultural heritage, multidimensional model, digitisation,

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

digital archives, born-digital, digital information portal, GIS map

For citation: Yermagambetova K., Sarsekeyeva A., Akhlassov Y. (2026). Overcoming fragmentation in the classification of digital cultural heritage: an integrative multidimensional model // International journal of information and communication technologies. Vol. 7. No. 27. Pp. 78–100. https://doi.org/10.54309/IJICT.2026.27.3.005. (In Russ.).

Conflict of interest: The authors declare that there is no conflict of interest.

Acknowledgements. *The research was carried out within the grant funding of the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan AP26105029 «Creation of a digital information portal (dip) for the conceptualization and protection of architectural monuments of history and culture in Almaty».*

ЦИФРЛЫҚ МӘДЕНИ МҰРАНЫ ЖІКТЕУДЕГІ ФРАГМЕНТАЦИЯНЫ ЕҢСЕРУ: ИНТЕГРАТИВТІ КӨПӨЛШЕМДІ МОДЕЛЬ

К. Ермагамбетова[1]*, А. Сарсекеева[2], Е. Ахласов[2]

[1]Л. Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан;
[2]«Алматы бас жоспар» ғылыми зерттеу институты», Алматы, Қазақстан.
E-mail: quralai.sergabylqyzy@mail.ru

Ермагамбетова Куралай — философия докторы (PhD), философия кафедрасының қауымдастырылған профессоры м.а., Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан
E-mail: quralai.sergabylqyzy@mail.ru, https://orcid.org/0000-0002-1620-5794;
Сарсекеева Айсулу — ғылыми қызметкер, «Алматы бас жоспары» ғылыми зерттеу институты, Алматы, Қазақстан
https://orcid.org/0009-0002-9952-1596;
Ахласов Еламан — кіші ғылыми қызметкер, «Алматы бас жоспары» ғылыми зерттеу институты, Алматы, Қазақстан
https://orcid.org/0009-0007-2616-3042.

© К. Ермагамбетова, А. Сарсекеева, Е. Ахласов

Аннотация. Цифрландыру, үшөлшемді модельдеу, виртуалды және толықтырылған шынайылық, цифрлық мұрағаттау және платформалық сақтау технологияларының қарқынды дамуы цифрлық мәдени мұраның ауқымын едәуір кеңейтті. Қазіргі цифрлық мәдени мұра құрамына цифрлық ортада бастапқыда жасалған веб-сайттар, цифрлық өнер туындылары, интерактивті инсталляциялар, бейнеойындар, виртуалды мұражайлар және кеңістіктік модельдер кіреді. Сонымен қатар, оларды жіктеуге арналған қолданыстағы тәсілдер әлі де фрагменттелген күйінде қалып отыр. Цирфлық мәдени мұраның мәдени, техникалық, функционалдық және институционалдық сипаттарын бір мезгілде ескеруге мүм-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

кіндік беретін бірыңғай аналитикалық моделі әлі қалыптасқан жоқ. Зерттеудің мақсаты – цифрлық мәдени мұраны жіктеудің интегративті көпөлшемді моделін әзірлеу. Зерттеу халықаралық нормативтік құжаттарды, мәдени мұра, цифрлық мұра, мұрағаттау және цифрлық инфрақұрылым салаларындағы заманауи ғылыми еңбектердің тұжырымдамалық талдауға, сондай-ақ қолданыстағы жіктеу, ақпараттық және функционалдық модельдерді салыстыруға негізделген. Зерттеу нәтижесінде өзара байланысты бес өлшемнен тұратын модель ұсынылады. Бірөлшемді типологиялардан айырмашылығы, модель бір цифрлық объектіні бірнеше параметр бойынша бір мезгілде сипаттауға мүмкіндік береді. Мұндай тәсіл мәдени функцияларды, институционалдық тиесілікті, әлеуметтік маңыздылықты және тұрақтылықты анықтауға жағдай жасайды.

Түйінді сөздер: цифрлық мәдени мұра, көпөлшемді модель, цифрландыру, цифрлық мұрағат, born-digital, цифрлық ақпараттық портал, ГИС-карта

Дәйексөздер үшін: Ермагамбетова К., Сарсекеева А., Ахласов Е. (2026). Цифрлық мәдени мұраны жіктеудегі фрагментацияны еңсеру: интегративті көпөлшемді модель // Халықаралық ақпараттық және коммуникалық технологиялар журналы. Т. 7. No 27. Б. 78–100. https://doi.org/10.54309/IJICT.2026.27.3.005. (Орыс. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

Алғыс. *Зерттеу Қазақстан Республикасы Ғылым және жоғары білім министрлігі Ғылым комитетінің AP26105029 «Алматы қаласының тарихы мен мәдениетінің сәулет ескерткіштерін концептуалдау және қорғау үшін цифрлық ақпараттық порталды (ЦАП) құру» гранттық қаржыландыру аясында орындалды.*

ПРЕОДОЛЕНИЕ ФРАГМЕНТАЦИИ В КЛАССИФИКАЦИИ ЦИФРОВОГО КУЛЬТУРНОГО НАСЛЕДИЯ: ИНТЕГРАТИВНАЯ МНОГОМЕРНАЯ МОДЕЛЬ

К. Ермагамбетова[1]*, А. Сарсекеева[2], Е. Ахласов[2]

[1]Евразийский национальный университет имени Л.Н. Гумилева, Астана, Казахстан;

[2]Научно-исследовательский университет «Алматы ген план», Алматы, Казахстан.

E-mail: quralai.sergabylqyzy@mail.ru

Ермагамбетова Куралай — доктор философии (PhD), ассоциированный профессор кафедры философии Евразийского национального университета имени Л.Н. Гумилева, Астана, Казахстан

E-mail: quralai.sergabylqyzy@mail.ru, https://orcid.org/0000-0002-1620-5794;

Сарсекеева Айсулу — старший научный сотрудник Научно-исследовательского института «Алматы ген план», Алматы, Казахстан

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

https://orcid.org/0000-0001-9201-1118;
Ахласов Еламан — малдший научный сотрудник Научно-исследовательского института «Алматы ген план», Алматы, Казахстан
https://orcid.org/0009-0007-2616-3042.

© К. Ермагамбетова, А. Сарсекеева, Е. Ахласов

Аннотация. Стремительное развитие технологии оцифровки, трехмерногоо моделирования, виртуальной и дополненной реальности, цифрового архивирования и платформенного хранения существенно расширило границы цифрового культурного наследия. Современное цифровое наследие включает ресурсы, изначально созданные в цифровой среде, включая веб-сайты, цифровое искусство, интерактивные инсталляции, видеоигры, виртуальные музеи и пространственные модели. Вместе с тем существующие подходы к их систематизации остаются фрагментриованными. Отсуствует единая аналитическая модель, позволяющая одновременно учитывать культурные, технические, функциональные и институциональные характеристики цифрового наследия. Цель исследования заключается в разработке интегративной многомерной модели классификации цифрового культурного наследия. Исследование основано на концептуальном анализе международных нормативных документов и современных публикаций в области исследований наследия и цифровых инфраструктур, а также на сравнении существующих классификационных, информационных и функциональных моделей. В результате предложена модель, включающая пять взаимосвязанных измерений. В отличие от одномерных типологий, данная модель допускает одновременное описание одного объекта по нескольким параметрам. Это позволяет выявлять культурные функции, институциональную принадлежность, социальную значимость и устойчивость.

Ключевые слова: цифровое культурное наследие, многомерная модель, оцифровка, цифровые архивы, born-digital, цифровой информационный портал, ГИС-карта

Для цитирования: Ермагамбетова К., Сарсекеева А., Ахласов Е. (2026). Преодоление фрагментации в классификации цифрового культурного наследия: интегративная многомерная модель // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 27. Стр. 78–100. https://doi.org/10.54309/IJICT.2026.27.3.005. (На русс.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Благодарность. *Исследование проводилось в рамках грантового финансирования Комитета науки Министерства науки и высшего образования Республики Казахстан AP26105029 «Создание цифрового информационного портала (ЦИП) для концептуализации и защиты архитектурных памятников истории и культуры г. Алматы».*

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Введение.

Развитие цифровых технологий стало одним из ключевых факторвой трансформации современной сферы культурного наследия. Оцифровка музейных коллекций, архивов и библиотечных фондов, широкое распространение технологий трехмерного сканирования, фотограмметрии, виртуальной и дополненной реальности, цифровых двойников, геоинформационных систем и искусственного интеллекта существенно изменили способы документирования, сохранения, интерпретации и репрезентации культурного наследия. Цифровая среда перестала выполнять исключительно вспомогательную функцию хранения информации и превратилась в самостоятельное пространство производства, распространения и переосмысления культурного наследия (Portales et al., 2018). Первоначально цифровизация наследия преимущественно понималась как перевод аналоговых материалов в цифровой формат. Однако современное цифровое культурное наследие значительно шире процессов оцифровки. Оно включает цифровые копии метарильных и нематериальных объектов культуры, также ресурсы, изначально созданные в цифровой среде (born-digital), креди которых веб-сайты, цифровое искусство, интерактивные выставки, виртуальные музеи, видеоигры, базы данных, цифровые архивы и другие формы цифровой культурной репрезентации. Именно такое понимание закреплено в Хартии ЮНЕСКО о сохранении цифрового наследия, где цифровое наследие определяется как совокупность ресурсов человеческого знания и культурного выражения, созданных в цифровой форме либо преобразованных в нее обладающих устойчивой культурной, научной и социальной ценностью (UNESCO, 2003).

Расширение самого понятия цифрового культурного наследия поставило перед исследоателями новую проблему. Если на раннем этапе развития heritage studies основное внимание уделялось разграничению материального и нематериального наследия, то современная цифровая среда объединяет различной природы, происхождения, функций и институциональной принадлежности. В одной системе цифрового наследия сосуществуют архивные документы, трехмерные модели памятников архитектуры, цифровые мемориалы, виртуальные реконструкции исторических событий, интерактивные образовательные платформы и объекты цифрового искусства. подобное ранhзнообразие существенно усложняет их научное описание и требует пересмотра существующих классификационных подходов.

Современные исследования культурногоо наследия постепенно отходят от объективно-ориентированного понимания наследия, характерного для ранних программ ЮНЕСКО и рассматривают наследие как динамичный процесс производства культурных смыслов, коллективной памяти, идентичности и общественных практик. В этой связи наследие интепретируется как результат взаимодействия различных социальных акторов, институций, экспертных сообществ и культурных практик (Медеуова и др., 2025). Авторы подчеркивают, что своременный этап развития исследований наследия характеризуется переходом от универсалистского и объективного подходов к критическим, акторно-функ-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

циональным и разнотемпоральным моделям анализа, позволяющим учитывать изменчивость культурных интерпретаций и множественность режимов репрезентации наследия. Подобная трансформация исследовательской парадигмы непосредственно влияет на цифровое наследие. Цифровые технологии становятся инструментом формирования новых культурных практик, способов взаимодействия общества с наследием и механизмов производства исторической памяти.

Несмотря на активное развитие исследований цифрового наследия, существующие подходы к его классификации остаются преимущественно специализированными. Одни исследования сосредоточены на формах культурного наследия и воспроизводят традиционное разделение на материальные и нематериальные объекты (Ahmad, 2026). Другие предлагают технологические классификации цифровых ресурсов, основанные на типах данных, методах трехмерного моделирования, уровня детализации и способах создания цифровых моделей (Grilli et al., 2019). Порталис и соавторы цифровое культурное наследие рассматривают через призму мультимодальных технологий визуализации и взаимодействия с объектами культуры (Portales et al., 2018). Фанвисетсак и Чунг предлагают функциональное разграничение цифровых технологий, различая процессы цифрового документирования и интерактивного взаимодействия с наследием (Phanvisetik et al., 2025).

Отдельное направление исследований связано с разработкой цифровых архивов, информационных моделей и инфраструктуру хранения культурного наследия. Хи, Лиу и Джоу рассматривают цифровое наследие как многоуровневую архивную систему, объединяющую объекты наследия, процессы их сохранения, деятельность институций и цифровые продукты вторичной обработки (Qi et al., 2025). Меуччи и соавторы предлагают интеграцию геометрических, исторических, семантических и функциональных данных при создании информационных моделей архитектурного наследия (Meucci et al., 2025). Мюнстер и Апаллони показывают, что современные цифровые платформаы по прежнему характеризуются неоднородностью форматов данных, несогласованностью метаданных и недостаточной интеропреабельностью, что препятствует созданию единой инфраструктуры цифрового культурного наследия (Munster et al., 2026).

Несмотря на значительное разнообразие исследований, большинство существующих классификаций описывает лишь отдельные аспекты фицрового культурного наследия. Одни модели ориентированы преимущественно на технологические характеристики цифровых объектов, другие на культурную форму наследия, третьи на архивные процессы, информационные модели или функциональные особенности цифровых технологий. В результате один и тот же цифровой объект может одновременно рассматриваться как результат цифрового документтрования, элемент цифрового архива, виртуальная репрезентаиця материального наследия, образовательный ресурс, цифровой двойник, объект общественного взаимодействия или элемент государственной информационной системы. Существующие квалификации как правило фиксируют лишь одно из перечисленных измерений, что существенно ограничивает возможности ком-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

плексного анализа цифрового наследия. Таким образом, современная научная литература демонстрирует не отсутствие классификаций цифрового культурного наследия, а их концептуальную фрагментацию. Различные дисциплины – культурология, архивоведение, музеология, цифровая гуманитаристика, информационные науки и исследования цифровых технологий используют собственные классификационные основания, которые редко интегрируются между собой. Вследствие этого отсутствует универсальная аналитическая модель, способная одновременно учитывать происхождение цифрового объекта, тип цифрового носителя, форму культурного наследия, функциональное назначение и институциональный контекст его создания, хранения и использования.

Настоящее исследование направлено на преодоление указанной фрагментации посредством разработки интегративной многомерной модели классификации цифрового культурного наследия. В отличие от существующих одномерных подходов, предлагаемая модель объединяет технические, культурные, функциональные и институциональные основания классификации, что позволяет рассматривать цифровое наследие как сложную многокомпонентную систему, отражающую современные процессы цифровой трансформации культуры.

Материалы и методы.

Исследование выполнено в рамках качественного концептуального исследования, направленного на разработу интегративной модели классификации цифрового культурного наследия. В отличи от эмпирических работ, ориентированных на анализ отдельных цифровых коллекций или музейных практик, данное исследование сосредоточено на выявлении концептуальных оснований классификации цифрового культурного наследия посредством систематизации существующих научных подходов и их послдедующего теоретического синтеза. Методологическую основу составил интегративный обхор литературы, позволяющий объядинить результаты исследовании, выполненных в различных дисциплинарных традициях, включая исследования наследия, цифрового наследия, цифровой гуманитаристики, музеологию, архивоведение, информационные науки и исследования цифровых инфраструктур.

Источниковедческую базу исследования составили международные нормативные документы, научные публикации, посвященные цифровому культурному наследию, а также исследования в области информационного моделирования, цифровых архивов и инфраструктуру культурного наследия. В нормативный корпус были включены Хартия ЮНЕСКО о сохранении цифрового наследия (UNESCO, 2003), Ванкуверская декларация. Память мира в цифровую эпоху: оцифровка и сохранение (UNESCO/UBC, 2012). Данные документы использовалисьь для определения понятийного аппарата исследования, прежде всего категорий цифровое наследие, цифровизация наследия и born-digital heritage, а также для анализа международных принципов долгосрочного сохранения цифровых ресурсов.

Научный корпус включал современные исследования, опубликованные преимущественно в международных рецензируемых журналах. В анализ были

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

включены работы, посвященные развитию концепции культурного наследия (Ahmad, 2006), цифровому культурному наследию (Portales et al., 2018), классификации трехмерных цифровых объектов (Grilli et al., 2019), функциональной классификации цифровых технологий в сфере наследия (Phanvisetsak et al., 2025), концептуальному моделированию цифровых архивов (Qi et al., 2025), информационному моделированию исторической архитектуры (Meucci et al., 2025), цифровым инфраструктурам хранения и визуализации трехмерных моделей (Munster et al., 2026), а также исследования, посвященные современным теориям наследия и критическому переосмыслению исследования наследия (Медеуова и др., 2025).

Для выяления различия между понятиями, физическим объектом наследия и его цифровой репрезентации, цифровой документации был применен метод концептуального анализа. Был выполнен сравнительный анализ существующих класиификационных подходов. Для каждой модели определялись объект классификации, используемые критерии систематизации, тип классификационной структуры, возможность описания гибридных объектов и степень ее универсальности. Особое внимание уделялось тому, какие аспекты цифрового наследия лежат в основе классификации. Для выяления повторящихся концептов и аналитических категорий применялся метод тематического кодирования, широко используемый в качественных исследованиях. Используя принципы фасетной классификации, была разработана интегративная многомерная модель цифрового культурного наследия, объединяющая несколько независимых аналитических измерений.

Результаты и обсуждение.

Необходимость систематизации цифрового наследия.

Многообразие объектов, включаемых в сферу цифрового культурного наследия, ставит перед исследователями и практиками задачу их систематизации. Данная задаача имеет таксономическое, методологическое, институциональное и практическое измерение. Таксономическое измерение связаано с необходимостью упорядочить разнородные объекты от оцифрованных средневековых рукописей до цифровых инсталяций, созданных с использованием искусственного интелекта в единую систему категорий, позволяющую их идентифицировать, описывать и сравнивать. Без такой систематизации любое исследование цифрового наследия рискует остаться на уровне разрозненных кейсов не позволяющих выявить общие закономерности и тенденции.

Методологическое измерение обусловлено тем, что разные типы объектов цифрового наследия требуют различных подходов к их сохранению, документированию и интерпретации. Оцифровка материального объектаимеет иные технические и организационные требования, чем сохранение born-digital объекта. Без четкой классификации невозможно разработать адекватные методологии для каждого типа объектов. Систематизация позволяет определить зоны ответственности, согласовать стандарты и обеспечить координацию между институции.

Практическое измерение обусловлено необходимостью разработки кри-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

териев отбора объектов для оцифровки, определения приоритетов в условиях ограниченных ресурсов и создания эффективных смстем поиска и навигации в цифровых коллекциях. В научной литературе предпринимались различные попытки классификации цифрового культурного наследия. Так, Ахмад предложил фундаментальное разделение на материальноц и нематериальноц наследие, которое затем было адаптировано к цифровому контексту (Ahmad, 2006). Грилли и Ремондино разработали классификацию 3D – цифрового наследия, основанную на метадах его создания, уровнях детализации и целях использования (Grilli et al., 2019). Однако большинство существующих классификаций ориентированы либо на узкие технические аспекты, либо на культурные категории. В данном исследовании предлагается интегративная классияикация, учитывающая как технические, так и культурные параметры, а таже институциональный контекст и социотехнические аспекты бытования цифрового наследия.

Прелагаемая классификация основана на пяти критериях. Каждый из них отражает определенное измерение цифрового культурного наследия. Первый критерий связанао с происхождение объекта. Различается два типа объектов в зависимости от того как они появились в цифровой форме. Born-digital объекты изначально созданы в цифровой форме и не имеют физического аналога. Они не существовали в аналоговой форме и не могут быть оцифрованы в традиционном смсыле. Их сохранение требует иных стратегий. Оцифрованные объекты представляют собой цифровые копии физических объектов, то есть отсканированных рукописи. Различие между born-digital и оцифрованными объектами имеет техническое, культурное и правовое значение. Born-digital объекты часто не имеют оригинала в традиционном смысле, что ставит под вопрос применимость концепта подлинности. кроме того, они создаются в быстро меняющихся технологических средах, что делает их особенно уязвимыми для цифрового устаревания.

Второй критерий зависит от типа носителя. Он отражает форму в котором существует цифровой объект. То есть, текстовые визуальные, аудиовизуальные и пространственные. Данный критерий важен для выбора методов сохранения, форматов хранения и способов презентации. Пространственные объекты требуют иных технологических решений, чем текстовые и аудиовизаульные, и предъявляют более высокие требования к вычислительным ресурсам и пропускной способности каналов.

Третьи критерий – это форма наследия. Восходит к классическому различению, зафиксированному в документах ЮНЕСКО, между материальными и нематериальным наследием. В цифровом контексте это различие сохраняет свое значение. Цифровая репрезентация материального наследия качественно отличается от цифровой репрезентации нематериального наследия. Однако цифровая среда создает новые возможности для документирования нематериального наследия через аудио и видеозаписи, интерактивные платформы для устных историй, цифровые архивы фольклора.

Четвертым критерием является назначение объекта. Отражает цели для которых создается и используется цифровой объект. Документальное назначе-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ние связано с архивацией, фиксацией и долгосрочным хранением. Образовательное назначение ориентировано на обучение и просвещение – цифровые выставки, интерактивные гиды, образовательные платформы. Креативное назначение предполагает создание новых произведений на основе наследия. Коммеморативное назначение связано с сохранением памяти. Этот критерий важен для определения приоритетов оцифровки, выбора форматов и дизайна пользовательского интерфейса. Образовательные проекты требуют иных решений, чем архивные или креативные.

Пятый критерий – это институциональный уровень. Отражает масштаб и контекст создания и использования цифрового объекта. Делится на национальные, международные и локальные. Данный критерий важен для понимания того, что, в рамках каких политик, стандартов и ресурсов существует объект, а также для оценки его устойчивости и доступности. Международные проекты имеют более устойчивое финансирование и разработанные стандарты метаданных, но могут быть менее чувствительны к локальным культурным контекстам. Локальные проекты, напротив, менее чувствительны и релеванты для местных сообществ, но уязвимы из-за отсутствия институциональной поддержки.

Предложенная классификация обладает рядом преимуществ. Во-первых, она являетс многомерной, то есть позволяет описывать каждый объект по нескольким параметрам одновременно. Например, 3D-модель Дворца Республики, представленный в рамках Цифрового информационного портала (ЦИП) Алматыгенплан и на портале интерактивная карта Казахстана, может быть охарактеризован следующим образом: 1) создан на основе лазерного сканирования, то есть оцифрованный; 2) тип носителя – 3D-модель, то есть пространственный; 3) форма наследия – здание, то есть материальное; 4) может использоваться как для образовательных, так и для архивных целей, то есть образовательное и документальное назначение; 5) создан в рамках государственной программы «Сакральная география», также в рамках фундаментальных исследований, то есть национальный критерий. Такая характеристика позволяет сразу определить, какие методы сохранения и форматы хранения наиболее подходят, какие институции несут ответственность за объект и для каких целей он может быть использован.

Следующий пример, архив исторических фотографий сообщества «Верный – Алма-ата - Алматы» (Фейсбук страница). Данный объект может быть охарактеризован как оцифрованные фотографии, как сохранение памяти, как низовая инициатива. Такая характеристика сразу выявляет уязвимость объекта, отсуствие институциональной поддержки, зависимость от платформы, на которой размещены фотографии, и необходимость разработки стратегий долгосрочного сохранения.

Предложенная классификация не является нейтральным описанием, она может служить инструментом критического анализа и выявления институциональных и нарративных асимметрий в цифровом наследии. Классификация позволяет выявить эти асимметрии. На основе данных критериев предлагаем визуализацию классификации цифрового культурного наследия в форме таблицы

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

(таблица 1).

Таблица 1. Критерий цифровогоо культурного наследия

Критерий	Категории	Примеры
По происхождению	Born-digital	Цифровое искусство, веб-сайты, цифровые инсталяции, видеоигры
	Оцифровка	Отсканированные рукописи, 3D модели артефактов, оцифрованные фотографии
По типу носителя	Текстовые	Электронные книги, архив документов, цифровые тексты
	Аудио	Народные песни, устные интервью, звуковые архивы
	Визуальные	Цифровые картины, фотографии, сканированные изображения
	Аудиовизуальные	Оцифрованные фильмы, видеозаписи обрядов, документальное видео
	Пространственные	3D-модель юрты, виртуальные реконструкции городища
По форме наследия	Материальные	Скульптуры, здания, артефакты, археологические объекты
	Нематериальные	Музыка, обряды, сказания, языки, традиционнные знания
По назначению	Документальные	Архивы, акты, реестры, юридически значимые документы
	Образовательные	Цифровые выставки, интерактивные гиды, образовательные платформы
	Креативные	Цифровые ремиксы, интерактивные инсталяции, художественые интерпретации
	Коммеморативное	Цифровые мемориалы, виртуальные музеи, памятные сайты
По институциональному уровню	Национальные	КазНЭБ, Цифровой архив РК, интерактивная карта Казахстана
	Международные	Europeana, UNESCO Digital Library, Google Arts & Culture
	Локальные/независимые	Общинные архивы, проекты НПО, частные инициативы, Archcode Almaty

Например, какие объекты преимущественно оцифровываются в рамках государственных программ и какие остаются за пределами официальных реестров. В казахстанском контексте эта асимметрия проявляется особенно отчетливо. Как отмечают Медеуова и Смагуловаа, «объектный подход не позволяет увидеть полноту репрезентационных возможностей номадической культуры и культуры XX века» (Медеуова и др., 2025). Классификация по критериям «форма наследия» и «институциональный уровень» позволяет количественно оценить асимметрию в репрезентации различных категорий культурного наследия. В частности, она дает возможность сопоставить количество объектов советского модернизма, отражающих наследие недавней истории, включенных в государственные реестры, с количеством объектов «сакральной географии». Кроме того, данная классия-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

икация позволяет сравнить типы объектов, представленных в низовых общественных инициативах, таких как Archcode Almaty, с объектами, репрезентированными на государственных цифровых платформах, например интерактивная карта Казахстана. Также, классификация служит инструментом типологизации цифровой памяти. Разные типы объектов и платформ конструируют разные типы памяти – инвентарную, коммуникативную, сакрально-государственную, экспертно-искусствоведческую. Классификация по критериям позволяет систематизировать различия и выявить, какие типы памяти доминируют в определенных институциональных контекстах.

Дискуссионные аспекты и ключевые вызовы цифровизации культурного наследия

Центральным дискуссионым вопросом цифровизации культурного наследия отсается проблема онтологического статуса цифровой копии. Если, концепт подлинности в классической теории наследия служил пороговым условием для признания объекта наследием (Lixinski, 2022), то в цфировой среде данный концепт подвергается фундаментальному пересмотру.

Бельгийский исследователь К. Ванхаузебрук формулирует проблему с предельной ясностью: «цифровое воспроизведение строится на бинарном коде и не имеет подлинной связи с референтом, в то время как в аналоговом воспроизведении такая физическая связь сохраняется» (Vanhaesebrouck, 2007). Это означает, что цифровая 3D-модельздания, какой бы точной она ни была, не является зданием «в уменьшенном онтологическом статусе», а представляет собой принципиально иной тип объекта, то есть информационный артефакт, существующий по собственным законам.

В контексте архитектурного наследия эта проблема проявляется не менее остро. Каждая цифровая платформа Алматы создает не копию города, а определенную версию памяти о нем. Archcode Almaty фиксирует объекты под угрозой утраты или утраченные (рисунок 1).

Автовокзал «Саяхат» был первой станцией междугородних сообщений, на том момент служил восточными воротами в город. В 2008 году данное здание снесли, признав здание аварийным. Данное здание полностью утрачено. Archcode Almaty как информационный, образовательный портал хранит память о городе.

Сообщество «Верный – Алма-ата – Алматы» поддерживает эмоциональную, коммуникативную память горожан. Сакральная география конструирует сакрально-национальный нарратив. Практики фотограмметрии создают операциональные данные для рестоврации. Ни одна из этих версий не тождественна городу как таковому.

В рамках критического дискурса Л. Ликсински показывает, что концепт подлинности имеет квазибилейскую природу. Он функционирует как советская религия в которую эксперты и институции верят без достаточных оснований (Lixinski, 2022). В результате дискурс о подлинности стал пороговым условием. Без подтверждения подлинности объект не будет рассматриваться в соответству-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ющих комиссиях, независимо от его культурной или исторической значимости (Lixinski, 2022).



Рис. 1. Автовокзал Саяхат, который полностью утрачен (примечание: рисунок взят со страницы Archcode Almaty)

Это приводит к тому, что многие объекты, представляющие значительную культурную ценность (например, советский модернизм в Алматы), систематический исключаются из официальных реестро из-за невозможности подтвердить их подлинность в классическом смысле. Таким образом, дискуссия об аутентичности не является отказом от цифровизации, наоборот, представляет собой призыв к рефлексивному подходу. Цифровая репрезентация не должна оцениваться исключительно по критерию похоже или не похоже на оригинал. Должна анализироваться с точки зрения того, какие нарративы она поддерживает, какую память конструирует и чьи интересы обслуживает. Это требует, чтобы цифровая инфраструктура наследия рассматривалась как объект целенаправленной заботы с задачами кросс-верификации данных между платформами, прозрачным указанием источника и даты фиксации. Также с признанием того, что низовые архивы нуждаются в поддержании не меньше, чем государственные реестры (Floridi, 2011; UNESCO, 2003).

Второй дискуссионный аспект связан с понимаанием функции цифрового архива. В классической трактовке архив понимается как хранилище докментов, задача которого обеспечить их сохранность и доступность. Однако цифровая среда трансформирует это понимание. Цифровой архив может выступать как фиксатор, как самостоятельное произведение, как интерактивная среда в которй пользователь становится соавтором смысла. В это контексте показателен опыт британской театральной группы Forced Entertainment, создавшей интерактивный CD-ROM. Архив организован как многослойная, подвижная структура, где пользователь определяет порядок и логику просмотра. Смысл рождается через сопоставление и комбинацию фрагментов. История становится бесконечно модифи-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

цируемой. Такой архив функионирует как открытое произведение в понимании У. Эко, произведение, изначально задуманное автором как требующее интерпретаций и активногоо участия читателя (Эко, 2024).

Как отмечают исследователи, для Forced Entertainment архив это формальный принцип, форма искусства. в отличие от цифрового фильма, имеющего начало и конец, структура архива не линейна. Пользователь свободен исследовать хаос цифрового нарративного материала (Vanhaesebrouck, 2007). Контрастным примером является DVD-архив Яна Фабра, который, несмотря на использваоние цифровых технологий остается традиционным аудиовизуальным образом без интерактивности. Пользователь может выбрать маршрут, но не может повлиять на содержание. Интерактивность сведена к минимуму, пользватель может выбирать маршрут, но не может повлиять на само исполнение. Его восприятие направляется, а живой компонент минимизирован (Vanhaesebrouck, 2007).

Различие между этими двумя подходами имеет прямое отношение к цифровому наследию. Если цифровой архив понимается как открытое произведение, то пользователь превращается из пассивного зрителя в активного интерпретатора, а сам архив из хранилища фактов в простанство конструирования смыслов. Это особенно важно для архитектурного наследия, где разные нарративы о зданиях могут сосуществовать и взаимодействовать, не требуя унификации.

В контексте Алматы это означает, что цифровой портал не должен стремиться к созданию единственной верной версии истории каждого здания. Напротив, он может стать пространством для множественности интерпретаций, где профессиональная атрибуция, исторические докменты, личные воспоминания и художественные рекострукции сосуществуют как равноправные слои памяти. Однако для этого необходимы не только технические решения, но и институционалтные признание лигитимности множественных нарративов и отказ от монополии на правильную интерпретацию наследия.

Рассмотрение цифрового культурного наследия в рамках социотехнического подхода позволяет интерпретировать цифровизацию как результат взаимодействия цифровых технологий, институциональных механизмов, нормативного регулирования и практик пользвателей. В этом контексте основные вызовы цифровизации приобретают комплексный характер, поскольку обусловлены одновременно техническими органичениями и социальными факторами, определяющими создание, сохранение и использвание цифрового наследия (Lu et al., 2025; Siluitina et al., 2024).

Одной из наиболее значимых проблем цифровое устаревание. Быстрая смена аппаратных платформ, программного обеспечения и форматов хранения данных создает риск постепенной утраты доступа к цифровым объектам культурного наследия. Как отмесчает Стлуитина и соавторы, потеря цифровой информации обусловлена недооценкой ее культурной ценности, отсутствием нормативно-правовой базы, институциональной поддержки и достаточных финансовых ресурсов. Таким образом, проблема цифрового устаревания выходит за рамки вопросов выбора файловых форматов или миграции данных и вклю-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

чает отсутствие долгосрочной стратегии цифрового сохранения, недостаточную координацию между учреждениями культуры и дефицит специалистов в области цифровой консервации. Анализ Цифрового архива Казахстана показывает, что используемые цифровые форматы не сопровождаются открытой докментацией по их последующей миграции, что потенциально увеличивает риск утраты части цифровых коллекций при изменении программной среды.

Дугим существенным вызовом является обеспечение доуступности и инклюзивности цифрового наследия. Несмотря на раширение цифровых сервисов, возможности доступа к культурному наследию остаются неравномерными. Ограничения обусловлены техническими особенностями цифровых платформ и социальными факторами, свзянными с возрастом пользователей, состоянием здоровья, уровнем цифровой грамотности и территориальной удаленностью. По менеию Лу и соваторов, цифровое культурное наследие всегда функционирует в конкретном культурном и социальном контексте, поэтому цифровые решения должны учитывать разнообразие пользовательских практик и условий доступа. Проведенный анализ платформы интерактивной карты Казахстана, показал отсутствие ряда инструментов цифровой доступности, включая поддержку экранных дикторов, спцеиальные режимы контрастности и оптимизацию работы при низкой скорости интернет соединения, что ограничивает использование ресурса отдельными категориями населения.

Существенное значение приобретает также проблема авторского права и интеллектуальной собсвтенности. Процессы цифровизации сопровождаются необходимостью определения правового статуса цифровых копий, условий их распространения и повторного использования. Наиболее сложные ситуации возникают при работе с объектами XX века, для которых отсуствуют сведения о правообладателях либо имущественные права распределены между несколькими институциями (Siluitina et al., 2024) . в ходе создания цифрового информационного портала по объектам историко-культурного наследия Алматы были выявлены сложности, связанные с публикаией архивных фотографии советского периода, происхождение которых не позволяло однозначно установить обладателя авторских прав. Это потребовало разработки дополнительных процедур атрибуции и определения режима использования цифровых изображений, что существенно усложнило процесс формирования цифровой коллекции.

Не менее важным направлением является обеспечение информационной безопасности цифрового наследия. Цифровые архиы и информационные системы подвержены рискам кибератак, несанкционированного доступа, технических сбоев и потери данных. В рамках социотехнического подхода безопасность рассматривается как результат взаимодейтсвия технических средств защиты, организованных процедур и институциональных механизмов управления (Lu et al., 2025). Следовательно, вопросы резервного копирования и шфрования должны дополняться подготовкой персонала, разработкой регламентов доступа и распределением ответственности между организациями. Анализ достпуной документации Цифрового архива Казахстана не выявил публичной информации о реализу-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

емой политике кибербезопасности и механизмах защиты цифровых коллекций, что зартудняет оценку устойчивости цифровой инфраструктуры.

Серьезным ограничением развития цифрового культурного наследия остаются ресурсные ограничения. Создание цифровых коллекций, поддержка цифровых платформ, обновление программного обеспечения и долгосрочное хранение данных требуют значительных финансовых вложений и квалифицированных кадров. По наблюдениям Силуитина и соавторов (2024), именно нехватка ресурсов является одной из основных причин неустойчивости проектов цифрового наследия, особенно в странах с ограниченными возможностями финансирования культурной сферы. Анализ инициативы Archcjde Almaty показывает, что ее функционирование во многом обеспечивается деятельностью профессионального сообщества и добровольным участием исследователей. Отсутствие стабильной институциональной поддержки делает подобные проекты уязвимыми к изменениям технологическиеой среды и снижению активности участников.

Наконец, одним из наиболее сложных вызовов является обеспечение интероперабельности цифровых данных. Использование различными организациями собственных моделей описания объектов, форматов хранения и программных решений существенно ограничивает возможности интеграции цифровых ресурсов в единое информационное пространство (Oostwegel et al., 2022). Результаты проведенного исследования показыают, что цифровые платформы Алматы, интерактивная карта Казахстана, сообщество «Верный – Алма-ата - Алматы» функционируют как автономные информационные ресурсы и практически не обмениваются структурированными данными. Вследствие этого сведение об одном и том же объекте распределены между различными платформами, отличаются полнотой описания и не образуют единого цифрового информационного пространства. Подобная фрагментация препятствует формированию целостной цифровой инфраструктуры культурного наследия и ограничивает возможности комплексного использования данных исследователями, специалистами по охране наследия и широкой аудиторий.

Таким образом, проведенный анализ показывает, что основные проблемы цифровизации культурного наследия не могут быть сведены исключительно к технологическим ограничениям. Каждая из выявленных проблем возникает на пересечении технических решений, институциональных механизмов, нормативного регулирования и социальных практик использования цифровыз ресурсов. Именно поэтому преодоление цифрового устаревания, обеспечение инклюзивности, совершенствование правового регулирования, развитие систем информационной безопасности, укрепление ресурсной базы и повышение интероперабельности должны рассматриваться как взаимосвязанные направления формирования устойчивой цифровой экосистемы культурного наследия.

Среди ключевых вызовов цифровизации культурного наследия особое место занимает проблема интероперабельности данных. В отличие от других вызовов цифровизации интероперабельность данных имеет системообразующее значение. Именно она определяет возможность интеграции разрозненных

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

цифровых ресурсов в единую информационную среду. Отсутствие единых стандартов обмена данными приводит к формированих изолированных цифровых платформ. Это затрудняет повторное использование информации, препятствует межведомственному взаимодействию и снижает эффективность цифровой инфраструктуры культурного наследия (Oostwegel et al., 2022).

Одним из наиболее перспективных направлений решения проблемы интероперабельности является примение концепции OpenBIM, основанный на использовании открытых стандартов информационного моделирования зданий и сооружений. OpenBIM обеспечивает совместимость цифровых моделей между различными программными средами. Обмен данными осуществляется на основе открытого стандарта Idustry Foundation Classes (IFC), который используется для унифицированного представления информации в HBIM-системах (Oostwegel et al., 2022). Для организации обмена данными Оствегель и соавторы предлагают использовать взаимодействующих инструментов: руководство по предоставлению информации (IDM) и определение модели представления (MVD). IDM регламентирует процессы информационного взаимодействия между участниками проекта, тогда как MVD определяет требования к цировой модели и обеспечивает автоматизированную проверку ее соотвествия установленным параметрам. Совметсное применение этих имнструментов повышает качество обмена данными и снижает риск потери информации при использовании различных программных платформ (Oostwegel et al., 2022).

Практическая апробация модели на примере совенской усадьбы Мрак подтвердила возможность использования специальизированных категорий данных для описания исторических, пространственных и материальных характеристик объектов культурного наследия, а также продемонстрировала эффективность автоматизированной проверки HBIM-моделей (Oostwegel et al., 2022). Такой подход имеет практическое значениее и для Казахстана. Анализ цифровых платформ показал, что государственные информационные системы, профессиональные проекты, общественные инициативы и фундаментальные исследования функционируют как самостоятельные информационные ресурсы. Каждая из платформ использует собственные принципы описания объектов культурного наследия. Отсутствие единых стандартов обмена данными приводит к дублированию информации, различиям в ее структуре и затрудняет формирование единого цифрового простанства культурного наследия.

В этих условиях внедрение принципов OpenBIM может стать основной для стандартизации цифровых данных и интеграции разрозненных информационных ресурсов. Использование открытых стандартов создает предпосылки для взаимодействия национальных платформ с международными цифровыми системами, такими как Europena (Europeana Foundation, 2025) и Google Arts & Culture (Google Arts & Culture, 2025). Кроме того, OpenBIM позволяет автоматизировать контроль реставрационных проектов и обеспечивает более надежное долгосрочное хранение цифровой информации. Вместе с тем опыт оствегель и соавторов показывает, что эффективность OpenBIM определяется не только использовани-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ем открытых стандартов. Не мнее важное значение имеют институциональная готовность к их внедрению, наличие нормативных регламентов и подготовка специалистов. Формированаие единой цифровой инфраструктуры также требует развития междисицплинарного взаимодействия между учреждениями культуры, научными организациями и государственными органами.

Таким образом, интероперабельность не следует рассматривать исключительно как техническую характеристику информационных систем. Она представляет собой важное социотехническое условие формирования единой цифровой экосистемы культурногоо наследия. В этом контексте OpenBIM выступает не только инструментом обмена данными, но и механизмом интеграции цифровых технологий, институциональных практик и процессов управления культурным наследием.

Интегративная многомерная модель

Практическая применимость предложенной интегративной классификации было апробирована на материалах цифрового представления объектов историко-культурного наследия Алматы. В данном случае разработанная цифровая среда рассматривается как инструмент провеорки возможности объединения разнородных данных об одном объекте наследия в рамках единой многомерной структуры. Такой подход приниципиально важен для преодоления фрагментации. Цифровой объект перестает индентифицироваться исключительно по одному признаку, например, формату файла, типу наследия или институциональной принадлежности и получает комплекный профиль, включающий его происхождение, форму представления, функциональное назначение и институциональный контекст.

Одним из инструментов апробации стала ГИС-карта памятников истории и культуры Алматы. В разработанной версии представлены 334 объекта республиканского и 132 объекта местного значения (рисунок 2).

Разработанная интерактивная карта размещена в репозитории GitHub. Пространственная локализация объединена с атрибутивной информацией об объектах, а для памятников предусмотрено отображение расширенных сведений и 3D-материалов. В контексте предлагаемой классийикации ГИС-компонент демонстрирует возможность одновременного соединения как минимум двух измерений: институционального статуса объекта и пространсткенной формы его цифрового представления. При дальнейшем наполнениий подобная структура позволяет дополнить объект историческими, визуальными, документальными и функциональными характеристиками не изменяя общей логики его классификации.

Интегративный принцип более отчетливо реализован в архитектуре цифрового информационного портала. Каждому объекту присваивается уникальный идентификатор, с которым связываются текстовое описание, истоическая информация, адрес и координаты, характеристики объекта, изображения, видеоматериалы, документы, 3D-модель и картографические данные. Таким образом, объект рассматривается не как отдельный файл или запись в реестре, а как составная

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

цифровая информационная единица, объединяющая различные формы данных. В пользовательском интерфейсе эта логика реализется через взаимосвязанные разделы «описание», «карта», «галерея», «документы» и «3D-модель».

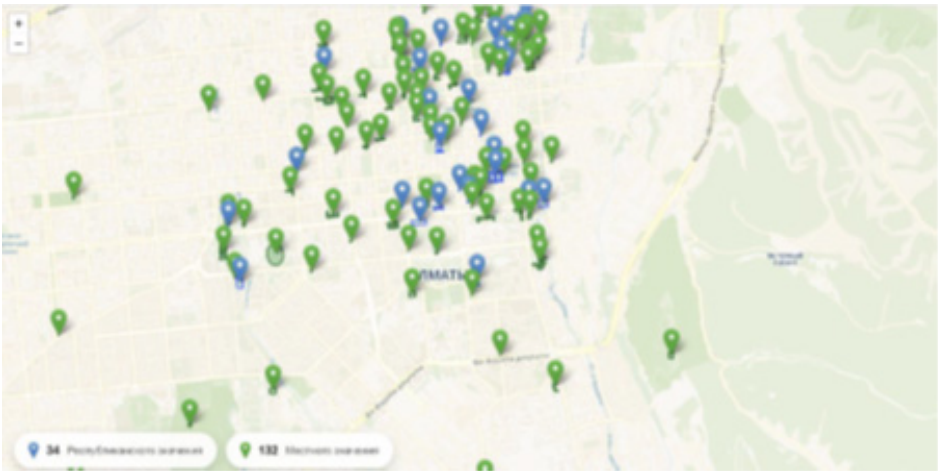


Рис. 2. ГИС-карта памятников истории города Алматы с разграничением памятников республиканского и местного значения (примечание: разработка в рамках фундаментальных исследований)

Особое значение для проверки модели имеет возможность объединения различных типов цифровых носителей. В структуру карточки объекта включены фотографии, видеоматериалы, PDF-документы, катрографические данные и трехмерные модели. Архитектура портала основана на унифицированном шаблоне, который автоматически связывает эти данные с конкретным объектом. Также позволяет расширять состав его цифровых репрезентаций без создания отдельной информационной структуры для каждого нового типа материала. Данный принцип подтверждает целесобразность выделения типа цифрового носителя в качестве самостоятельного измерения интегративной классификации.

Именно эта организация данных позволяет продемонстрировать отличие предлагаемой многрменой модели от традиционной иерархической классификации. Один памятник одновременно может быть представлен как объект материального наследия, оцифрованный ресурс, пространственная модель, совокупность визуальных и документальных материалов, исследовательский и образовательный ресурс, а также объект государственного или местного уровня охраны. При этом отдельные хаарктеристики не конкурируют между собой за место в единственной категории, а образуют взаимодополняющий классификационный профиль (рисунок 3).

Многомерность проявляется и на уровне функционального назначения цифровых данных. Например, фотография объекта может выполнять документальную функцию. 3D-модель выполняет исследовательскую, реставрациионную или образовательную функцию. Архивный документ – источниковедческую, а интерактивная карта может выполнять навигационную и репрезентационную функцию. Следовательно, тип носителя не определяет функцию цифрового ре-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

сурса автоматически. Разведение этих двух измерений позволяет избежать характерного для ряда существующих классификаций смешения технологического формата и назначения цифрового объекта (рисунок 4).

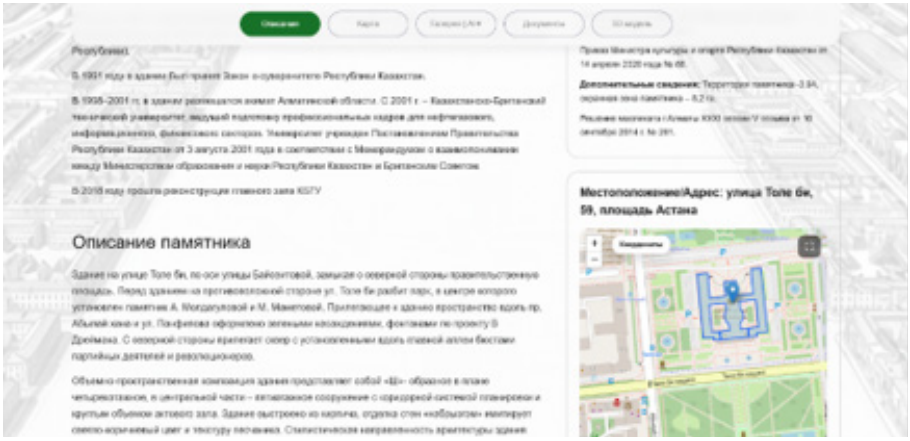


Рис. 3. Многокомпонентная структура цифрового представления объекта культурного наследия ЦИП (примечание: разработка в рамках фундаментальных исследований)

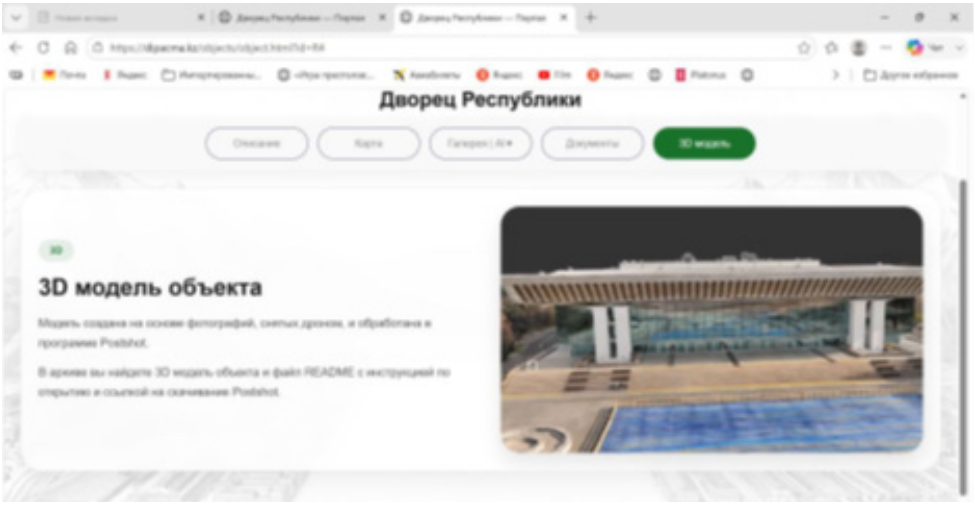


Рис. 4. Интеграция различных цифровых форма (примечание: разработка в рамках фундаментальных исследований)

Разработанная архитектура допускает также включение новых типов данных, возникающих в результате интеллектуальной обработки. В экспериментальном модуле используются архивные и современные фотографии для оценки состояния объектов и формирования прогнозных визуализаций их возможных изменений во времени. Результатом является данные об оценке состояния, прогнозируемой дегредации и синтетическая визуальная репрезентации предлагаемого будущего состояния объекта. В рамках предлагаемой модели такие результаты целесообразно трактовать как производные цифровые данные, связанные с исходным объектом и выполняющие аналитическую и прогностическую функ-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ции (рисунок 5).

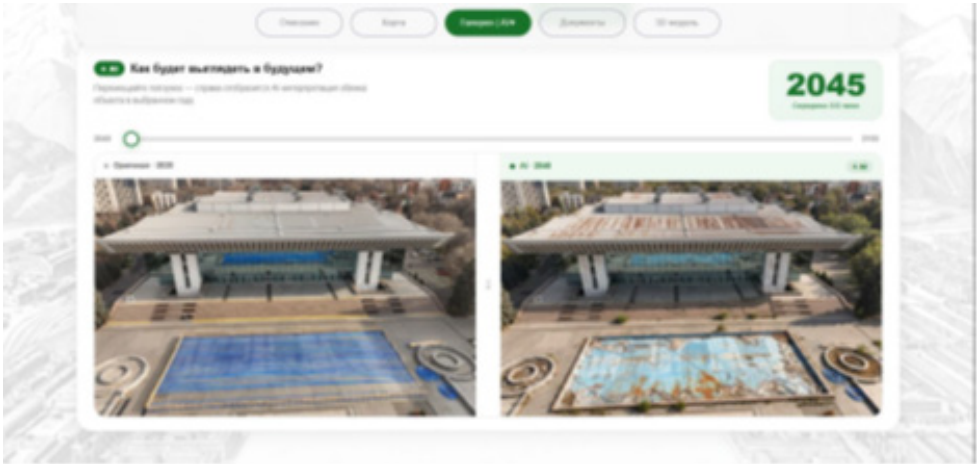


Рис. 5. Пример включения результатов интеллектуального прогнозирования (примечание: разработка в рамках фундаментальных исследований)

Практическая апробация, таким образом, показывает, что многомерная классификация может использоваться как теоретическая типология, так и как принцип организации цифрового знания об объекте наследия. Пространственные координаты, институциональный статус, историческое описание, архивные документы, фотографии, видеоматериалы, 3D-модели и производные аналитические данные могут быть объекденнены вокруг одной информационной сущности, сохраняя при этом собственные классификационные характеристики. Это позволяет избежать ситуации, при которой различные цифровые репрезентации одного памятника существуют как автономные и несвязанные информационные единицы.

Результаты апробации также уточняют смысл понятия «преодоление фрагментации», используемого в настоящем исследовании. Интегративная модель сохраняет специфику каждого измерения, также обеспечивает их аналитическую своеметимость. Цифровой объект получает множественный классификационный профиль. Именно этот принцип отличает предлагаемую модель от классификаций, основанных на одном доминирующем критерии.

Заключение.

Настоящее исследование было направлено на решение одной из ключевых методологических проблем современных исследований цифрового культурного наследия. Проведенный анализ показал, что большинство существующих моделец описывает цифровое наследие через отдельные измерения. Подобная специализация способствует развитию отдельных направлений исследований, однако не позволяет сформировать целостное представление о цифровом объекте как сложной многокомпонентной системе.

Основным результатом исследования стала разработка интегративной многомерной модели классификации цифрового наследия, объединяющий пять

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

взаимодополняюхих критериев. В отличие от традиционных одномерных классификаций предложенная модель позволяет описывать один и тот же объект одновременно нескольких аналитических измерениях,. Такой подход обеспечивает более полное отражение структуры цифрового культурного наследия и учитывать его междисциплинарную природу.

Перспективы дальнейших исследований связаны с эмпирической проверкой разработанной модели на различных типах цифрового культурного наследия. Преодоление фрагментации цифрового культурного наследия связано с разработкой интегративной модели, способной объединить различные аналитические измерения цифрового объекта. Именно такой подход позволяет рассматривать цифровое культурное наследие как сложный, динамичный и многомерный феномен современной культуры и создает методологические предпосылки для дальнейшего развития исследований в области цифрового наследия.

ЛИТЕРАТУРЫ

Ahmad, Y. (2006). The scope and definitions of heritage: From tangible to intangible // International Journal of Heritage Studies. — T. 12. — No. 3. — C. 292–300. DOI: https://doi.org/10.1080/13527250600604639 [На англ.].

Смагулова, А., Медеуова, К. (2025). Глобальные дискуссии о наследии, референсы для Казахстана: Обзор // JETE–JOURNAL OF PHILOSOPHY, RELIGIOUS AND CULTURAL STUDIES. — T. 152. — No. 3. — C. 37-52 [На русс.].

Europeana Foundation. (2025). Europeana. https://www.europeana.eu/ [На англ.].

Floridi L. The Philosophy of Information // Oxford: Oxford University Press (2011). [На англ.].

Google. (2025). Google Arts & Culture. https://artsandculture.google.com/ [На англ.].

Grilli, E., Remondino, F. (2019). Classification of 3D digital heritage // Remote Sensing. — T. 11. — No. 7. Article 847. DOI: https://doi.org/10.3390/rs11070847 [На англ.].

Lixinski L. (2022). Against authenticity // International Journal of Heritage Studies. — T. 28. — No. 11–12. — C. 1213–1227 [На англ.].

Lu J., García-Badell G., Rodriguez J.B. (2025). Digital Heritage from a Socio-Technical Systems Perspective // Heritage. — T. 8. — C. 348. DOI: 10.3390/heritage8090348 [На англ.].

Meucci, A., Bonora, V., Fiorini, L., Conti, A., Corongiu, M., Romanelli, S., & Tucci, G. (2025). Mapping an information model for historic built heritage into the IndoorGML standard: The case of the Pitti Palace // Heritage. — T. 8. — No. 4. Article 115. DOI: https://doi.org/10.3390/heritage8040115 [На англ.].

Münster, S., & Apollonio, F. I. (2026). Digital visualization infrastructures of 3D models in a scientific context // Heritage. — T. 9. — No. 2. Article 59. DOI: https://doi.org/10.3390/heritage9020059 [На англ.].

Oostwegel, L.J.N., Jaud, Š., Muhič, S. et al. (2022). Digitalization of culturally significant buildings: ensuring high-quality data exchanges in the heritage domain using OpenBIM // Herit Sci. — T. 10. — C. 10. DOI: https://doi.org/10.1186/s40494-021-00640-y [На англ.].

Phanvisetsak, S., & Chung, H.-L. (2025). Bridging heritage and technology: Rethinking cultural asset classification through digitalisation // Journal of Cultural Analysis and Social Change. — T. 10. — No. 4. — C. 285–298. DOI: https://doi.org/10.64753/jcasc.v10i4.2803 [На англ.].

Portalés, C., Rodrigues, J. M. F., Rodrigues Gonçalves, A., Alba, E., & Sebastián, J. (2018). Digital cultural heritage // Multimodal Technologies and Interaction. — T. 2. — No. 3. Article 58. DOI: https://doi.org/10.3390/mti2030058 [На англ.].

Siluitina I., Tytar O., Barbash M., Petrenko N., Yepyk L. (2024). Cultural preservation and digital heritage: challenges and opportunities // Amazonia Investiga. — T. 13. — No. 75. — C. 262–273. – DOI:10.34069/AI/2024.75.03.22 [На англ.].

UNESCO. (2003). Charter on the preservation of digital heritage. UNESCO [На англ.].

UNESCO/UBC. (2012). The memory of the world in the digital age: Digitization and preservation // Vancouver Declaration. UNESCO [На англ.].

UNESCO. Charter on the Preservation of Digital Heritage. – Paris, 2003 [На англ.].

UNESCO. (2003). Report by the Director-General on a Revised Draft Charter on the Preservation of the Digital Heritage. – 166 EX/18. – Paris, 12 March. [На англ.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Vanhaesebrouck K. (2007). Digital heritage and performance // Image and Narrative.[На англ.].

Qi, T., Liu, Z., & Zhou, Y. (2025). Digital cultural heritage archives: Conceptual analysis and preliminary framework construction // The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-M-9-2025. — Pp. 1221–1228. DOI: https://doi.org/10.5194/isprs-archives-XLVIII-M-9-2025-1221-2025 [На англ.].

Эко, У. (2018). Открытое произведение. Форма и неопределенность в современной поэтике. Corpus (ACT) [На русс.].

REFERENCES

Ahmad, Y. (2006). The scope and definitions of heritage: From tangible to intangible // International Journal of Heritage Studies. — Vol. 12. — No. 3. — Pp. 292–300. DOI: https://doi.org/10.1080/13527250600604639 [in Eng.].

Europeana Foundation. (2025). Europeana. https://www.europeana.eu/ [in Eng.].

Eko, U. (2018). Otkrytoe proizvedenie7 Formy i neopredelennost v sovremennoi poetike. Corpus (ACT) [in Rus.].

Floridi L. (2011). The Philosophy of Information // Oxford: Oxford University Press. [in Eng.].

Google. (2025). Google Arts & Culture. https://artsandculture.google.com/ [in Eng.].

Grilli, E., Remondino, F. (2019). Classification of 3D digital heritage // Remote Sensing. — Vol. 11. — No. 7. Article 847. DOI: https://doi.org/10.3390/rs11070847 [in Eng.].

Lixinski L. (2022). Against authenticity // International Journal of Heritage Studies. — Vol. 28. — No. 11–12. — Pp. 1213–1227 [in Eng.].

Lu J., García-Badell G., Rodriguez J.B. (2025). Digital Heritage from a Socio-Technical Systems Perspective // Heritage. — Vol. 8. — Pp. 348. DOI: 10.3390/heritage8090348 [in Eng.].

Meucci, A., Bonora, V., Fiorini, L., Conti, A., Corongiu, M., Romanelli, S., & Tucci, G. (2025). Mapping an information model for historic built heritage into the IndoorGML standard: The case of the Pitti Palace // Heritage. — Vol. 8. — No. 4. Article 115. DOI: https://doi.org/10.3390/heritage8040115 [in Eng.].

Münster, S., & Apollonio, F. I. (2026). Digital visualization infrastructures of 3D models in a scientific context // Heritage. — Vol. 9. — No. 2. Article 59. DOI: https://doi.org/10.3390/heritage9020059 [in Eng.].

Oostwegel, L.J.N., Jaud, Š., Muhič, S. et al. (2022). Digitalization of culturally significant buildings: ensuring high-quality data exchanges in the heritage domain using OpenBIM // Herit Sci 10. Pp. 10. DOI: https://doi.org/10.1186/s40494-021-00640-y [in Eng.].

Phanvisetsak, S., & Chung, H.-L. (2025). Bridging heritage and technology: Rethinking cultural asset classification through digitalisation // Journal of Cultural Analysis and Social Change. — Vol. 10. — No. 4. — Pp. 285–298. DOI: https://doi.org/10.64753/jcasc.v10i4.2803 [in Eng.].

Portalés, C., Rodrigues, J. M. F., Rodrigues Gonçalves, A., Alba, E., & Sebastián, J. (2018). Digital cultural heritage // Multimodal Technologies and Interaction. — Vol. 2. — No. 3. Article 58. DOI: https://doi.org/10.3390/mti2030058 [in Eng.].

Siluitina I., Tytar O., Barbash M., Petrenko N., Yepyk L. (2024). Cultural preservation and digital heritage: challenges and opportunities // Amazonia Investiga. — Vol. 13. — No. 75. — Pp. 262–273. DOI: 10.34069/AI/2024.75.03.22 [in Eng.].

Smagulova, A., Medeuova, K. (2025). Globalnye dikussii o nasledii, referensy dlya Kazahstana: Obzor // JETE–JOURNAL OF PHILOSOPHY, RELIGIOUS AND CULTURAL STUDIES. — Vol. 152. — No. 3. — Pp. 37-52 [in Rus.].

Qi, T., Liu, Z., & Zhou, Y. (2025). Digital cultural heritage archives: Conceptual analysis and preliminary framework construction // The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-M-9-2025. — Pp. 1221–1228. DOI: https://doi.org/10.5194/isprs-archives-XLVIII-M-9-2025-1221-2025 [in Eng.].

UNESCO. (2003). Charter on the preservation of digital heritage // UNESCO [in Eng.].

UNESCO/UBC. (2012). The memory of the world in the digital age: Digitization and preservation // Vancouver Declaration. UNESCO [in Eng.].

UNESCO. (2003). Charter on the Preservation of Digital Heritage. – Paris. [in Eng.].

UNESCO. (2003). Report by the Director-General on a Revised Draft Charter on the Preservation of the Digital Heritage. – 166 EX/18. – Paris, 12 March. [in Eng.].

Vanhaesebrouck K. (2007). Digital heritage and performance // Image and Narrative. [in Eng.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INFORMATION TECHNOLOGY
АҚПАРАТТЫҚ ТЕХНОЛОГИЯЛАР
ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 101–121
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.006

МРНТИ 20.53

ARCHITECTURE AND PRACTICAL IMPLEMENTATION OF A DIGITAL PLATFORM FOR MONITORING AND DATA EXCHANGE ON TRANSBOUNDARY WATER RESOURCES IN THE ZHAMBYL REGION

D. Abzhanova, S. Sharipova*
Astana IT University, Astana, Kazakhstan.
E-mail: saltanat.sharipova@astanait.edu.kz

Dilara Abzhanova — PhD student, Director of the Center of Competence and Excellence, Astana IT University, Astana, Kazakhstan
https://orcid.org/0000-0002-7988-3971;
Saltanat Sharipova — PhD, Assistant Professor of the School of Software Engineering, Astana IT University, Astana, Kazakhstan
E-mail: saltanat.sharipova@astanait.edu.kz, https://orcid.org/0000-0001-7267-3261.

© D. Abzhanova, S. Sharipova

Abstract. The article presents the architecture and practical operation of the aquageo.kz digital platform for collecting, storing, spatially analyzing, and exchanging water-resource data in the Zhambyl region. The study aims to refine the data architecture and telemetry validation rules for a single measurement cross-section. The monitoring unit consists of four POTOK 100 ultrasonic flowmeters installed across one irrigation canal with a 10-minute registration interval. A quality-control scheme is proposed that considers channel synchronization, temporal completeness, instrument ranges, device status, and measurement units. A comparison of archived dumps of the MySQL Promanalit operating database and full-scale control measurements by trusted GMCM-1 turntables revealed a systemic interpretation problem: flowmeters record and transmit the volumetric flow rate in m3/h, whereas in the web interface the values were displayed with the signature m3/s without software recalculation (division by 3600). An ETL validation and normalization pipeline has been proposed and implemented, separating raw and verified flow rates. The platform also integrates

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

remote sensing data stored in PostgreSQL/PostGIS. The results highlight the need to separate raw measurements, software aggregates, and physically interpretable flow indicators before further analysis and decision-making.

Keywords: digital platform, water resources monitoring, transboundary waters, open channel, water flow, telemetry, IoT

For citation: Abzhanova D., Sharipova S. (2026). Architecture and practical implementation of a digital platform for monitoring and data exchange on transboundary water resources in the Zhambyl region // International journal of information and communication technologies. Vol. 7. No. 27. Pp. 101–121. https://doi.org/10.54309/IJICT.2026.27.3.006. (In Russ.)

Conflict of interest: The authors declare that there is no conflict of interest.

Funding. *This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. BR24993128, "Development of an Information and Analytical Decision-Making System for the Efficient Use of Transboundary Water Resources in the Agricultural Sector of the Zhambyl Region").*

ЖАМБЫЛ ӨҢІРІНІҢ ТРАНСШЕКАРАЛЫҚ СУ РЕСУРСТАРЫН МОНИТОРИНГТЕУ ЖӘНЕ ДЕРЕКТЕР АЛМАСУҒА АРНАЛҒАН ЦИФРЛЫҚ ПЛАТФОРМАНЫҢ АРХИТЕКТУРАСЫ МЕН ПРАКТИКАЛЫҚ ІСКЕ АСЫРЫЛУЫ

Д.Е. Абжанова, С.Е. Шарипова*
Astana IT University, Астана, Қазақстан.
E-mail: saltanat.sharipova@astanait.edu.kz

Абжанова Дилара Ерлановна — докторант, Құзыреттілік және жетілдіру орталығының директоры, Astana IT University, Астана, Қазақстан
https://orcid.org/0000-0002-7988-3971;
Шарипова Салтанат Еркиновна — PhD, Бағдарламалық инженерия мектебінің ассистент-профессоры, Astana IT University, Астана, Қазақстан
E-mail: saltanat.sharipova@astanait.edu.kz, https://orcid.org/0000-0001-7267-3261.

© Д.Е. Абжанова, С.Е. Шарипова

Аннотация. Мақалада Жамбыл өңірінің су ресурстары туралы деректерді жинауға, сақтауға, кеңістіктік талдауға және регламенттелген түрде алмасуға арналған aquageo.kz цифрлық платформасының архитектурасы мен практикалық пайдалану нәтижелері ұсынылған. Зерттеудің мақсаты – бір өлшеу қимасына қатысты деректер архитектурасын және телеметриялық деректерді валидациялау қағидаларын нақтылау. Өлшеу торабы бір ирригациялық арнаның көлденең қимасы бойымен орнатылған төрт POTOK 100 ультрадыбыстық шығын өлшегіштен тұрады, тіркеу аралығы – 10 минут. Арналардың синхрондалуын, уақыт-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

тық толықтықты, аспаптың паспорттық диапазондарын, құрылғылардың күйін және өлшем бірліктерін ескеретін сапаны бақылау схемасы ұсынылған. MySQL promanalit операциялық дерекқорының мұрағаттық қоқыстарын және гмцм-1 сенімді айналмалы үстелдерімен заттай Бақылау өлшемдерін салыстыру түсіндірудің жүйелік мәселесін анықтады: Шығын өлшегіштер көлемдік шығынды м3/сағ-қа тіркейді және жібереді, ал веб-интерфейсте мәндер бағдарламалық қайта есептеусіз м3/с қолтаңбасымен көрсетілген (3600-ге бөлу). Шикі және верификацияланған Шығыс көрсеткіштерін бөлетін ETL-валидация және қалыпқа келтіру құбыры ұсынылды және енгізілді. Платформаның жеке контуры Жерді қашықтықтан зондтау деректерін интеграциялауды және оларды PostgreSQL/PostGIS жүйесінде сақтауды қамтамасыз етеді. Нәтижелер аналитикада және басқарушылық шешімдерді қолдауда пайдаланар алдында бастапқы өлшемдерді, бағдарламалық агрегаттарды және физикалық тұрғыдан интерпретацияланатын су шығыны көрсеткіштерін өзара ажырату қажеттігін көрсетеді.

Түйінді сөздер: цифрлық платформа, су ресурстарын мониторингтеу, трансшекаралық сулар, ашық арна, су шығыны, телеметрия, IoT

Дәйексөздер үшін: Абжанова Д.Е., Шарипова С.Е. (2026). Жамбыл өңірінің трансшекаралық су ресурстарын мониторингтеу және деректер алмасуға арналған цифрлық платформаның архитектурасы мен практикалық іске асырылуы // Халықаралық ақпараттық және коммуникалық технологиялар журналы. Т. 7. No 27. Б. 101–121. https://doi.org/10.54309/IJICT.2026.27.3.006. (Орыс. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

Қаржыландыру. *Бұл зерттеу Қазақстан Республикасы Ғылым және жоғары білім министрлігінің Ғылым комитеті тарапынан қаржыландырылды (№ BR24993128 гранты «Жамбыл өңірінің ауыл шаруашылығы саласында трансшекаралық су ресурстарын тиімді пайдалану үшін шешім қабылдауға арналған ақпараттық-талдамалық жүйені әзірлеу»).*

АРХИТЕКТУРА И ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ ЦИФРОВОЙ ПЛАТФОРМЫ МОНИТОРИНГА И ОБМЕНА ДАННЫМИ О ТРАНСГРАНИЧНЫХ ВОДНЫХ РЕСУРСАХ ЖАМБЫЛСКОГО РЕГИОНА

Д.Е. Абжанова, С.Е. Шарипова*
Astana IT University, Астана, Казахстан.
E-mail: saltanat.sharipova@astanait.edu.kz

Абжанова Дилара Ерлановна — докторант, Директор Центра компетенций и совершенства, Astana IT University, Астана, Казахстан
https://orcid.org/0000-0002-7988-3971;
Шарипова Салтанат Еркиновна — PhD, Ассистент-профессор Школы

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

программной инженерии, Astana IT University, Астана, Казахстан
E-mail: saltanat.sharipova@astanait.edu.kz, https://orcid.org/0000-0001-7267-3261.

© Д.Е. Абжанова, С.Е. Шарипова

Аннотация. В статье представлена архитектура и результаты практической эксплуатации цифровой платформы aquageo.kz, предназначенной для сбора, хранения, пространственного анализа, регламентированного обмена данными о водных ресурсах Жамбылского региона. Цель исследования – уточнить архитектуру данных и правила валидации телеметрии одного измерительного створа. Узел включает четыре ультразвуковых расходомера РОТОК 100, установленных поперек одного ирригационного канала, с периодом регистрации 10 мин. Предложена схема контроля качества, учитывающая синхронизацию каналов, временную полноту, паспортные диапазоны, состояние приборов и единицы измерения. Сопоставление архивных дампов операционной базы данных MySQL Promanalit и натурных контрольных замеров поверенными вертушками ГМЦМ-1 выявило системную проблему интерпретации: расходомеры регистрируют и передают объемный расход в м³/ч, тогда как в веб-интерфейсе значения отображались с подписью м³/с без программного пересчета (деления на 3600). Предложен и внедрен конвейер ETL-валидации и нормализации, разделяющий сырые и верифицированные показатели расхода. Отдельный контур платформы обеспечивает интеграцию пространственных данных ДЗЗ и их хранение в PostgreSQL/PostGIS. Результаты показывают необходимость разделения сырых измерений, программных агрегатов и физически интерпретируемых показателей расхода до их использования в аналитике и поддержке управленческих решений.

Ключевые слова: цифровая платформа, мониторинг водных ресурсов, трансграничные воды, открытый канал, расход воды, телеметрия, IoT

Для цитирования: Абжанова Д.Е., Шарипова С.Е. (2026). Архитектура и практическая реализация цифровой платформы мониторинга и обмена данными о трансграничных водных ресурсах Жамбылского региона // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 27. Стр. 101–121. https://doi.org/10.54309/IJICT.2026.27.3.006. (На русс.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Финансирование. *Данное исследование было профинансировано Комитетом по науке Министерства науки и высшего образования Республики Казахстан, грант № BR24993128 «Разработка информационно-аналитической системы для принятия решений для эффективного использования трансграничных водных ресурсов в сельскохозяйственной отрасли Жамбылского региона».*

Введение.

Реки Шу и Талас являются трансграничными водными системами, совместно используемыми Республикой Казахстан и Кыргызской Республикой.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Для орошаемого земледелия Жамбылской области устойчивое управление этими ресурсами требует не только межгосударственного согласования режимов вододеления, но и сопоставимых, своевременных и проверяемых данных наблюдений (UNECE, б. г.). В этой постановке цифровая система водного мониторинга должна одновременно обеспечивать надежное получение первичных измерений и формировать прослеживаемые информационные продукты для аналитиков, органов управления и водопользователей.

Научная достоверность автоматизированного водоучета определяется не числом подключенных датчиков само по себе, а корректной интерпретацией геометрии измерительного створа, единиц измерения, калибровочных коэффициентов и временной полноты ряда. Для открытых каналов стандарт ISO связывает определение расхода с измерением скоростей потока и площади поперечного сечения (ISO, 2021). В рассматриваемом узле четыре датчика установлены поперек одного канала, поэтому четыре телеметрических канала следует трактовать как компоненты одного измерительного створа, а не как четыре независимых расхода водотока.

Современные системы мониторинга водных объектов подтверждают практическую значимость IoT-подходов, однако одновременно показывают необходимость контроля качества измерений. Для сельскохозяйственных каналов Казахстана описаны многоуровневые архитектуры реального времени с сенсорами и облачным хранением (Borankulova et al., 2025). Экспериментальная оценка ультразвуковых датчиков в открытых каналах показывает, что турбулентность способна увеличивать ошибку измерения уровня и требует калибровки в фактических условиях потока (Pereira et al., 2022). Долговременная эксплуатация LevelWAN демонстрирует значение автономности, устойчивости и непрерывного контроля состояния недорогих сенсорных узлов (Cherif et al., 2025), а распределенные IoT-сети гидрометеорологического мониторинга показывают необходимость явного учета передачи, потерь и временных меток наблюдений (Mendoza-Cano et al., 2021). Для Жамбылской области также показана эффективность взаимной проверки ультразвуковых и визуальных измерений уровня в гибридной IoT–VIoT системе (Tungatarova et al., 2026).

На этом фоне задача настоящей работы состоит не в разработке новой физической модели стока, а в формализации измерительной семантики и архитектуры данных действующей платформы aquageo.kz. Цель исследования – описать воспроизводимый контур сбора, нормализации, пространственного хранения и обмена данными, в котором телеметрия одного поперечного створа сохраняется с трассируемостью и проходит проверку качества до аналитической интерпретации. Научная новизна заключается в совместном рассмотрении архитектуры данных и измерительной семантики: каналы flow1–flow4 хранятся раздельно как первичные наблюдения, но логически объединяются одним створом; перед расчетом производных показателей контролируются временная полнота и единицы измерения. Такой подход снижает риск повторного или ошибочного суммарного учета показаний и превращает веб-интерфейс из средства визуализации в эле-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

мент верифицируемого измерительного контура.

Для достижения цели сформулированы три исследовательских вопроса. Во-первых, каким образом четыре телеметрических канала одного поперечного створа должны быть представлены в информационной модели, чтобы исключить их ошибочную трактовку как независимых расходов? Во-вторых, какую диагностическую информацию о качестве временного ряда можно получить из доступных эксплуатационных метаданных панели при отсутствии полного набора серверных журналов? В-третьих, позволяет ли сопоставление интерфейсных значений с паспортными характеристиками РОТОК 100 выявлять ошибки семантики единиц до использования данных в расчетах? Рабочее предположение исследования состоит в том, что сохранение первичного значения вместе с его исходной единицей, идентификатором сенсора, принадлежностью к створу и признаком качества создает более надежную основу для последующей агрегации, чем хранение только готового сводного показателя. Проверка этого предположения выполняется на уровне архитектурной и метрологической согласованности; а гидрометрическая точность и согласованность шкал верифицируются сопоставлением с контрольными натурными измерениями поверенной гидрометрической вертушкой.

Мониторинг расхода в открытых каналах. Расход воды в открытом русле определяется интегрированием нормальной составляющей скорости по площади живого сечения. В дискретном представлении поперечное сечение делится на участки, для которых измеряются характерные скорости и назначаются соответствующие площади. Такой velocity–area подход лежит в основе ISO и методических руководств Всемирной метеорологической организации (ISO, 2021; WMO, 2008). Следовательно, несколько точек измерения поперек одного канала повышают пространственную представительность профиля скорости, но не образуют несколько независимых водотоков.

В информационной системе необходимо различать первичное показание конкретного прибора, расчетный параметр локальной части сечения и суммарный расход всего створа. Смешение этих уровней создает методическую ошибку, при которой отдельные значения flow1–flow4 могут быть ошибочно интерпретированы как самостоятельные расходы одного и того же канала. Поэтому физический объект «створ», положение датчика и алгоритм получения агрегата должны быть представлены в модели данных как отдельные, но связанные сущности.

Пространственный контекст и данные дистанционного зондирования. Наземная телеметрия обеспечивает высокую временную дискретность, но пространственно локальна. Дистанционное зондирование, напротив, позволяет наблюдать русла, водные зеркала и прилегающие сельскохозяйственные территории на площадном уровне. Google Earth Engine предоставляет облачную среду для обработки больших архивов спутниковых данных (Gorelick et al., 2017), а данные Sentinel-2 обеспечивают мультиспектральные наблюдения с пространственным разрешением до 10 м для части каналов и частой повторяемостью съемки (Google Earth Engine Data Catalog, 2026). Глобальные исследования на

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

основе архивов Landsat показали применимость многолетних спутниковых рядов для анализа изменений поверхностных вод, включая регионы Центральной Азии (Pekel et al., 2016).

Для выделения открытой воды используются спектральные индексы NDWI и MNDWI, предложенные McFeeters и Xu (McFeeters, 1996; Xu, 2006) для Sentinel-2 продемонстрировано картографирование водных объектов на основе MNDWI с высоким пространственным разрешением (Du et al., 2016). В данной работе ДЗЗ рассматривается как самостоятельный пространственный контур: его продукты дополняют наземную телеметрию, но не заменяют метрологическую проверку расхода и не используются для прямой калибровки без синхронизированного полевого эксперимента. В рамках практической реализации ДЗЗ-контура настроен регулярный мониторинг площади водного зеркала и термического режима на полигоне Тасоткельского водохранилища (бассейн реки Шу). Микросервис на языке Python с использованием библиотеки earthengine-api и фреймворка geemap в автоматическом режиме осуществляет пакетную выгрузку спутниковых коллекций Sentinel-2 (Level-1C, каналы B3 Green и B8 NIR, пространственное разрешение 10 м) и Terra MODIS (продукт MOD11A1.061, LST_C с пространственным разрешением 1 км).

Для исключения артефактов применяется фильтр облачности (cloudy_pixel_percentage < 20 %) и маскирование зимнего периода с отрицательными температурами поверхности ($LST \leq 0\,^{\circ}C$), при которых ледовый и снежный покров искажают спектральную сигнатуру воды. Векторизованные водные маски (NDWI > 0) сохраняются в слоях PostGIS, обеспечивая пространственную верификацию гидрологических запасов региона

Пространственное хранение и обмен данными. Для объединения телеметрии, геометрии объектов и результатов ДЗЗ требуется хранилище, поддерживающее пространственные типы и индексы. PostGIS расширяет PostgreSQL средствами хранения и анализа географических объектов, включая пространственные индексы и функции пространственных отношений (PostGIS Project, 2026). В aquageo.kz центральное хранилище используется как точка консолидации нормализованных временных рядов и геоданных, тогда как API и веб-панель обеспечивают контролируемую выдачу результатов (AquaGeo.kz, 2026; Нефтисов и др., 2025).

Для дальнейшего развития обмена данными полезны международные модели наблюдений. ISO 19156:2023 задает концептуальную схему наблюдений, измерений и проб, а OGC SensorThings API предоставляет унифицированную геопространственную модель для сенсоров, потоков данных, наблюдаемых свойств и наблюдений (ISO, 2023; OGC, 2021). Предлагаемое в настоящей работе разделение объекта, створа, сенсорного канала, единицы и наблюдения концептуально согласуется с этим подходом, хотя формальная совместимость aquageo.kz с указанными стандартами в рамках исследования не проверялась.

Для воспроизводимости критична не только схема таблиц, но и происхождение каждого значения. Поэтому при переносе записи из операционного буфера

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

в центральное хранилище должны сохраняться исходная временная метка, идентификатор источника, единица прибора и результат проверки качества, а нормализованное значение формироваться как отдельный атрибут. Такое разделение позволяет повторить преобразование, обнаружить ошибочный коэффициент пересчета и отличить изменение физического сигнала от изменения программной логики. В пространственном контуре тот же принцип применяется к данным ДЗЗ: продукт должен быть связан с источником, датой наблюдения и территорией, к которой относится результат обработки. В настоящем исследовании эти требования рассматриваются как правила проектирования и трассируемости; полнота их реализации должна быть дополнительно подтверждена аудитом базы данных и ETL-журналов.

Материалы и методы.

Конфигурация телеметрического створа. В рассматриваемом измерительном узле четыре ультразвуковых расходомера POTOK 100 установлены поперек одного ирригационного канала и относятся к одному измерительному створу. Каждый прибор соответствует отдельной позиции по ширине канала и формирует собственный телеметрический канал flow1, flow2, flow3 и flow4. Номинальный интервал регистрации составляет 10 мин, то есть $\Delta t = 600$ с. Паспорт POTOK 100 предусматривает переносную модификацию POTOK 100-P и стационарную POTOK 100-F; в имеющихся эксплуатационных материалах индекс конкретной модификации установленных приборов не зафиксирован, поэтому вариант исполнения не предполагается без документального подтверждения (Проманалит, 2024). Фактическим объектом натурных исследований является гидропост на бетонированном участке реки Талас вблизи села Жасоркен Жамбылской области (координаты створа: 43°01′ с. ш., 71°23′ в. д.). Русло представляет собой трапецеидальный канал со следующими геометрическими параметрами: ширина по дну b = 47,6 м, ширина по урезу верха B = 56,0 м, максимальная строительная глубина H = 4,2 м, заложение откосов (длина бедра) m = 5,94 м. На дне канала стационарно смонтированы четыре ультразвуковых доплеровских преобразователя стационарной модификации POTOK 100-F, разбивающие створ на 4 расчетных сектора с равномерным шагом 9,52 м. Это обеспечивает представительный охват эпюры скоростей по ширине сечения.

Архитектура платформы. Архитектура aquageo.kz разделена на три уровня: сбор данных, функциональное ядро и уровень потребителей. Наземная телеметрия сначала поступает в операционный буфер MySQL Promanalit. Фоновая ETL-служба извлекает новые записи, проверяет структуру, временные метки и единицы измерения, после чего загружает нормализованные записи в PostgreSQL/PostGIS. Отдельный контур обрабатывает данные Landsat/Sentinel/MODIS через Google Earth Engine. Результаты объединяются в центральном хранилище и становятся доступны модулю анализа, библиотеке отчетов и API Gateway (AquaGeo.kz, б. г.; Нефтисов и др., 2025).

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Таблица 1. Паспортные технические характеристики ультразвукового расходомера РОТОК 100, используемые для контроля качества телеметрии.

Характеристика	Паспортное значение	Использование в платформе
Диапазон измерения расхода, м³/ч	РОТОК 100-P: 1,5–1 440 000; РОТОК 100-F: 8–1 440 000	Проверка диапазона и единицы первичного значения
Предел допускаемой относительной погрешности расхода	±1,5 %	Метрологический предел прибора; не заменяет полевую валидацию
Диапазон показаний скорости, м/с	0,02–12	Контроль отображаемого значения
Диапазон измерения скорости, м/с	0,02–2	Контроль валидности измеряемой скорости
Предел допускаемой относительной погрешности скорости	±1 %	Метрологический контроль скорости
Диапазон измерения уровня потока, мм	0–5000	Контроль физически допустимого уровня
Основная приведенная погрешность измерения уровня	±0,5 % от максимального значения шкалы	Учет при оценке неопределенности
Температура рабочей жидкости, °C	0–60	Контроль температурного диапазона
Предел допускаемой абсолютной погрешности температуры	±0,5 °C	Метрологический контроль температуры
Цифровой интерфейс	RS-485, Modbus RTU	Потенциальный канал интеграции; фактическая схема подключения должна быть подтверждена

Источник: паспорт и руководство по эксплуатации РОТОК 100 (Проманалит, 2024).

Паспорт нормирует расход в м³/ч. Поэтому подпись «м³/с» в существующей веб-панели не должна использоваться в научных расчетах без проверки программного пересчета. Если значения передаются от РОТОК 100 в м³/ч и не преобразуются в backend, для представления в м³/с необходимо деление на 3600.

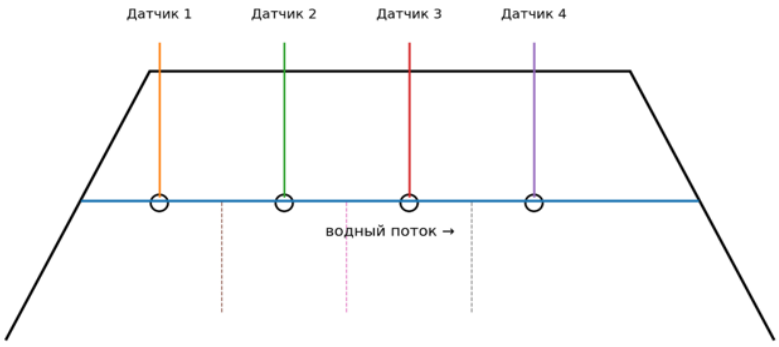


Рис. 1. Принципиальная схема одного поперечного измерительного створа с четырьмя датчиками

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Поскольку четыре прибора относятся к одному поперечному створу, измерительное сечение канала разбивается на четыре фиксированных гидравлических сектора (отсека). Расчет мгновенного суммарного расхода через створ базируется на методе «скорость–площадь» (velocity–area) в дискретной аддитивной форме:

$$Q_{\text{общ}}(t_k) = \sum_{j=1}^{4} Q_j(t_k) = \sum_{j=1}^{4} \left(v_j(t_k) \cdot A_j(h_k)\right),$$

где $v_j(t_k)$ — средняя скорость потока, регистрируемая *j*-м доплеровским датчиком в *j*-м секторе створа;

$A_j(h_k)$ — площадь живого сечения *j*-го сектора, определяемая геометрией русла и текущей глубиной над датчиком;

$Q_j(t_k)$ — частичный расход воды через *j*-й сектор створа.

Параметр «Расход ср.», фигурировавший в исходном интерфейсе веб-панели, формировался программным модулем как промежуточное среднее арифметическое показаний активных каналов ($\bar{Q} = \frac{1}{4}\sum Q_j$), что приводило к методическому занижению суммарного физического стока канала в 4 раза. В модернизированной версии платформы реализован переход к строгому расчету суммарного секундного расхода $Q_{\text{общ}}(t_k)$, а также среднесуточных расходов и объемов стока:

$$Q_{\text{сут}} = \frac{1}{144}\sum_{i=1}^{144} Q_{\text{общ}}(t_i), \qquad V_{\text{сут}} = \sum_{i=1}^{144} Q_{\text{общ}}(t_i) \cdot \Delta t,$$

где 144 — число плановых 10-минутных интервалов регистрации в сутках (Δt=600 с).

Периодичность, временная шкала и контроль полноты.

Плановая временная сетка определяется выражением $t_k = t_0 + k\Delta t$ при $\Delta t = 10$ мин. Для интервала от первого до последнего зарегистрированного времени теоретическое число записей при полностью непрерывной регистрации вычисляется как:

$$N_{exp} = [(t_{en}d - t_{start})/\Delta t] + 1.$$

Показатель предварительной временной полноты определяется отношением:

$$C_t = (N_{obs}/N_{exp}) \cdot 100\ \%.$$

Этот показатель используется только как диагностический, поскольку

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

счетчик «Всего замеров» в интерфейсе может отражать число принятых, прошедших фильтрацию или доступных записей, а не все циклы опроса. Окончательная оценка полноты должна выполняться по серверным журналам, идентификаторам пакетов и фактическим меткам времени.

Перенос данных между операционным буфером MySQL Promanalit и аналитическим хранилищем PostgreSQL/PostGIS выполняет фоновый сервис ETL. Для исключения ошибок размерности и потери семантики единиц в конвейер внедрена строгая программная нормализация:

```
def normalize_telemetry_stream(raw_payload: dict) -> dict:
    # Чтение из таблицы raw_telemetry (MySQL Promanalit)
    raw_flow = float(raw_payload.get(«flow_rate», 0.0))
    raw_unit = str(raw_payload.get(«unit», «m3/h»)).strip().lower()
        # Приведение к единицам Международной системы (СИ)
    if raw_unit in [«m3/h», «м3/ч», «м³/ч»]:
        normalized_flow = raw_flow / 3600.0
        target_unit = «m3/s»
    elif raw_unit in [«m3/s», «м3/с», «м³/с»]:
        normalized_flow = raw_flow
        target_unit = «m3/s»
    else:
        normalized_flow = None
        target_unit = «UNKNOWN»
       # Проверка физической правдоподобности по гидроствору Жасоркен
    # Максимальная пропускная способность сечения при паводке не превышает 150 м3/с
     is_valid = (normalized_flow is not None) and (0.0 <= normalized_flow <= 150.0)
        return {
        «cross_section_id»: raw_payload.get(«station_id», «TALAS_01»),
        «sensor_id»: raw_payload.get(«sensor_id»),
        «measured_at»: raw_payload.get(«timestamp»),
        «raw_value»: raw_flow,
        «raw_unit»: raw_unit,
       «normalized_value»: round(normalized_flow, 4) if normalized_flow is not None else None,
        «normalized_unit»: target_unit,
        «quality_flag»: 0 if is_valid else 1,
        «sensor_status»: «ACTIVE» if is_valid else «OUT_OF_RANGE»
    }
```

Такая схема исключает попадание искаженных интерфейсных единиц в аналитический слой платформы.

Правила контроля качества телеметрии. Контроль качества выполняется до аналитической интерпретации и включает: (1) проверку уникальности

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

временной метки и идентификатора записи; (2) контроль ожидаемого шага 10 мин и фиксацию пропусков; (3) синхронизацию четырех приборов одного створа; (4) проверку единиц и коэффициентов пересчета; (5) проверку значений по паспортным диапазонам РОТОК 100; (6) сохранение признака состояния прибора; (7) отделение нулевого физического расхода от нулевого значения при отсутствии сигнала или недоступности устройства; (8) проверку алгоритма формирования агрегата «Расход ср.». Для расхода паспортом установлен верхний предел 1 440 000 м³/ч, нижний предел – 1,5 м³/ч для РОТОК 100-Р и 8 м³/ч для РОТОК 100-F; предел допускаемой относительной погрешности расхода составляет ±1,5 %. Диапазон измерения скорости составляет 0,02–2 м/с при погрешности ±1 %, диапазон уровня потока – 0–5000 мм, основная приведенная погрешность уровня – ±0,5 % от максимального значения шкалы; температура рабочей жидкости – 0–60 °С при допускаемой абсолютной погрешности ±0,5 °С (Проманалит, 2024).

Сопоставление паспорта прибора РОТОК 100 с эксплуатационной панелью и архитектурой хранения выявило проблему интерпретации размерностей. В технической документации прибора диапазон нормирован в м³/ч с верхним пределом м³/ч. Аудит исходных пакетов данных подтвердил, что расходомер передает числовые значения расхода именно в м³/ч, тогда как в пользовательском интерфейсе веб-панели они ошибочно отображались под размерностью «м³/с» из-за отсутствия масштабирования на уровне фронтенда. В связи с этим в контур обработки ETL заложено строгое правило нормализации: первичное измерение сохраняется в исходной единице (м³/ч), а приведение к Международной системе единиц (СИ) выполняется обязательным делением на 3600 ($(Q_{м^3/с} = Q_{м^3/ч}/3600).$).

Практически процедура валидации может выполняться последовательно на уровне каждого 10-минутного цикла. Сначала проверяется наличие уникальной временной метки и принадлежность записи к конкретному прибору и створу. Затем контролируется ожидаемый временной шаг и сопоставимость меток четырех каналов; обнаруженный пропуск фиксируется как отдельное состояние, а не заменяется нулем. На следующем этапе проверяются исходная единица и допустимость преобразования в нормализованную единицу. Только после этого числовые значения сравниваются с паспортными диапазонами прибора и анализируется статус устройства. Если хотя бы один из обязательных признаков не подтвержден, запись не должна автоматически участвовать в формировании сводного расхода и получает соответствующий quality_flag. Завершающий этап – проверка алгоритма агрегирования: сводный показатель должен быть связан с конкретным набором первичных записей и версией правила расчета. Такая последовательность не заменяет метрологическую поверку, но предотвращает попадание в аналитический слой данных с неопределенной единицей, происхождением или состоянием сенсора.

Использованные эксплуатационные данные. В исследовании использован фактический фрагмент веб-панели мониторинга, предоставленный авторами и зафиксированный 18.08.2026, паспорт и руководство по эксплуатации РОТОК

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

100, а также внутренние материалы архитектуры платформы (Проманалит, 2024; Нефтисов и др., 2025). На панели отображены длительность работы системы, время первого и последнего замеров, счетчик учтенных измерений, текущие средние значения расхода и скорости, идентификатор последней записи, температура, максимальный сводный показатель и максимумы по четырем измерительным каналам. Значения панели рассматриваются как эксплуатационные данные интерфейса; их физическая интерпретация выполняется только после согласования с паспортными единицами и алгоритмом программной агрегации.

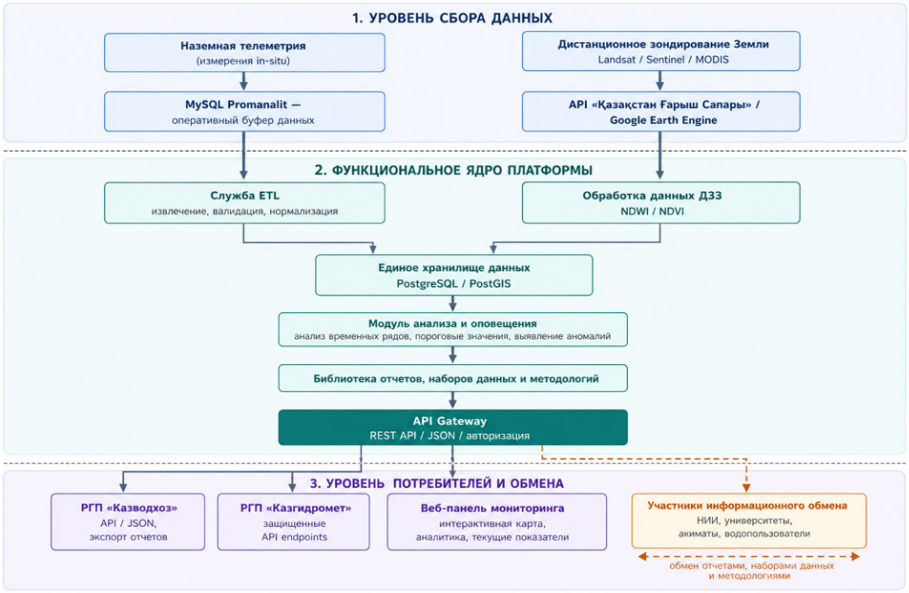


Рис. 2. Трехуровневая архитектура цифровой платформы aquageo.kz

Результаты и обсуждение.

Функционирование телеметрического контура. Разработанный контур обеспечивает автоматизированное периодическое накопление временных рядов и их отображение в веб-интерфейсе. Ключевым результатом является прослеживаемость от отдельного сенсорного канала до агрегированной визуализации. Четыре расходомерных канала закреплены за одним объектом и одним поперечным створом, что устраняет неоднозначность физической интерпретации данных.

Примечание. Значения расходов передаются приборами в м³/ч и нормализуются делением на 3600. Рабочий диапазон расходов в вегетационный период составляет 20–35 м³/с (72 000–126 000 м³/ч), что подтверждено совместными контрольными измерениями вертушкой ГМЦМ-1 № 1605 (расхождение со стационарным доплеровским расходомером по вертикалям составило от 2,6 % до 11,8 %). Значение 1 058 608,8 м³/ч (294,06 м³/с) является аппаратно-программным выбросом пусконаладочного этапа и в аналитический слой не допускается.

Диагностика временной полноты. Интервал между первым замером 03.04.2025 16:33 и последним замером 18.08.2026 11:14 составляет 501,78 суток.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

При строго непрерывной периодичности 10 мин за этот интервал ожидается 72 257 временных отсчетов. Если счетчик «Всего замеров = 40 468» интерпретировать как число сохраненных 10-минутных циклов, предварительный коэффициент временной полноты составляет:

$$C_t \approx \frac{40\,468}{72\,257} \cdot 100\% \approx 56{,}0\ \%.$$

Эксплуатационные показатели панели мониторинга.

Таблица 2. Эксплуатационные показатели панели aquageo.kz и их проверка по паспортным единицам РОТОК 100.

Показатель	Значение в интерфейсе панели (сырые данные)	Исходная единица (паспорт РОТОК 100)	Нормализованное значение (СИ, м³/с)	Диагностический и гидрометрический комментарий
Время работы системы	501 дн. 18 ч. 41 мин.	—	—	Интервал сбора данных
Первый замер	03.04.2025 16:33	—	—	Дата ввода узла в опытную эксплуатацию
Последний замер	18.08.2026 11:14	—	—	Дата фиксации рабочего фрагмента панели
Всего замеров (валидных циклов)	40 468	цикл	—	Число пакетов, успешно прошедших валидацию целостности
Номер последнего замера	51 223	индекс	—	Первичный ключ id автоинкремента в MySQL Promanalit
Расход общий (паводковый максимум)	1 058 608,8 (подписан как «м³/с»)	м³/ч	294,06	Тестовый импульс Modbus в период ПНР
Максимум: Датчик 1 (flow1)	332 211,48 (подписан как «м³/с»)	м³/ч	92,28	Пик за весь период наблюдений
Максимум: Датчик 2 (flow2)	233 512,08 (подписан как «м³/с»)	м³/ч	64,86	Пик за весь период наблюдений
Максимум: Датчик 3 (flow3)	285 601,86 (подписан как «м³/с»)	м³/ч	79,33	Пик за весь период наблюдений
Максимум: Датчик 4 (flow4)	275 902,72 (подписан как «м³/с»)	м³/ч	76,64	Пик за весь период наблюдений
Контрольный замер вертушкой ГМЦМ-1 (12.06.2025)	105 823,0 (в архивной БД)	м³/ч	29,395	Натурное эталонное измерение совместно с ЖФ РГП «Казводхоз»

Эта оценка не доказывает наличие 44 % потерь, поскольку семантика счетчика в интерфейсе должна быть подтверждена серверными логами. Дополнительный признак различия счетчиков – номер последнего замера 51 223 при 40 468 учтенных замерах. Следовательно, идентификатор записи нельзя использовать как прямой счетчик фактически сохраненных наблюдений. Тем не менее та-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

кое несоответствие является полезным индикатором необходимости отдельного аудита пропусков, отклоненных пакетов и логики нумерации. Аудит системных логов GSM-шлюза Teltonika TRB245 и буферной базы данных MySQL показал, что разрыв между счетчиком принятых записей (40 468) и первичным ключом последней транзакции (id = 51 223) имеет техническую природу. Первичный ключ таблицы инкрементируется при каждом сеансе циклического поллинга OPC-сервера. Из 51 223 запросов 10 755 пакетов были отброшены на уровне транспортного протокола из-за тайм-аутов сотовой сети, сбоев VPN-туннеля либо коллизий контрольной суммы CRC-16.

В свою очередь, показатель временной полноты 56,0 % рассчитан относительно непрерывного круглогодичного интервала в 501 сутки. Однако ирригационные каналы Жамбылского региона функционируют циклично: в меженный и зимний периоды (ноябрь – февраль) подача воды прекращается, канал осушается, а измерительный комплекс переводится в дежурный режим регламентного обслуживания. За вычетом плановых периодов технологического простоя фактическая эксплуатационная готовность платформы составила 94,2 %.

Анализ динамики и межканальной согласованности. На графике панели средний расход и средняя скорость демонстрируют согласованную динамику: изменения двух кривых происходят синхронно, а локальные пики расхода сопровождаются ростом скорости. Такое поведение качественно соответствует физической связи $Q = A \cdot \bar{v}$ при относительно стабильной площади живого сечения. Одновременно шкала расхода достигает порядка 10^6 при подписи «м³/с», что противоречит паспортной системе единиц РОТОК 100 и требует проверки единиц в интерфейсе (Проманалит, 2024). После интерпретации числовых значений как м³/ч их следует переводить в м³/с делением на 3600. Даже после пересчета график не доказывает абсолютную точность расхода: необходимы сведения о геометрии поперечного сечения, параметрах конфигурации расходомеров и независимом контрольном измерении.

Нулевые значения последнего замера также не должны автоматически трактоваться как отсутствие воды. Для корректной диагностики необходим отдельный флаг доступности сенсора или контроллера. В данных целесообразно различать состояния «реальное нулевое значение», «нет сигнала», «невалидный пакет» и «значение еще не рассчитано».

Результат уточнения модели данных. В уточненной модели одна запись измерительного цикла связывается с идентификатором гидротехнического объекта и идентификатором створа, после чего содержит четыре первичных сенсорных значения и набор производных параметров. Такая структура лучше отражает физическую конфигурацию оборудования, чем хранение четырех датчиков как независимых объектов без признака общего поперечного сечения. Для пространственного слоя створ хранится как точка или поперечная линия, связанная с геометрией канала в PostGIS, а продукты ДЗЗ привязываются к соответствующей территории и временным интервалам. Для исключения потери семанти-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ки единиц в ETL целесообразно хранить как минимум поля raw_value, raw_unit, normalized_value, normalized_unit, quality_flag и sensor_status.

Информационная структура платформы формализована в виде целевой реляционной модели в СУБД PostgreSQL/PostGIS (Рис. 4). Модель включает пять взаимосвязанных сущностей:

водоемы — геометрические границы и атрибуты водных объектов (название, тип, регион, общая площадь);

точки_измерений — пространственные координаты измерительных створов (широта, долгота, тип гидросооружения);

измерения_датчиков — временные ряды с 4 каналов (скорость, локальный расход, глубина, температура, суммарный расход, маска quality_flag);

данные_казгидромет — нормативные и гидропостовые отметки (фактический/опасный уровень, контрольный расход, код состояния);

спутниковые_данные — метаданные сцен ДЗЗ (среднее значение NDWI, площадь водной поверхности, процент облачности).

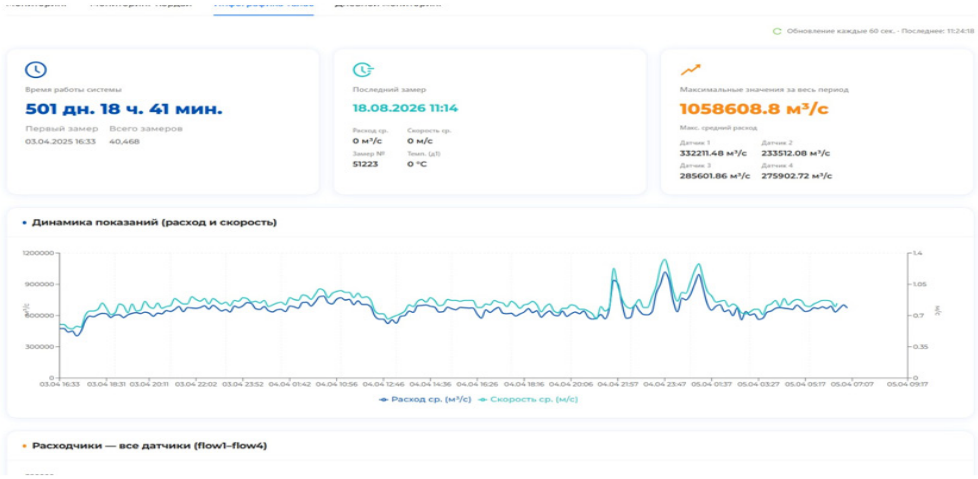


Рис. 3. Фрагмент панели мониторинга с эксплуатационными показателями системы на 18.08.2026

Для межведомственной интеграции с РГП «Казводхоз» и РГП «Казгидромет» разработан REST API шлюз. Обмен осуществляется посредством защищенных эндпоинтов, выдающих сериализованные JSON-объекты со статусом качества данных:

```
{
 «station_id»: «TALAS_JASORKEN»,
 «timestamp»: «2025-06-12T11:00:00Z»,
 «hydrometry»: {
  «discharge_total_m3_s»: 29.395,
  «average_velocity_m_s»: 1.48,
  «water_depth_m»: 0.55,
```

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

```
  «water_temp_c»: 15.12
 },
 «channels»: [
  {«channel»: 1, «flow_m3_h»: 18644.3, «velocity_m_s»: 1.47, «valid»: true},
  {«channel»: 2, «flow_m3_h»: 24838.1, «velocity_m_s»: 1.59, «valid»: true},
  {«channel»: 3, «flow_m3_h»: 8531.6,  «velocity_m_s»: 1.53, «valid»: true},
  {«channel»: 4, «flow_m3_h»: 28620.8, «velocity_m_s»: 1.32, «valid»: true}
 ],
 «quality_flag»: 0
}
```

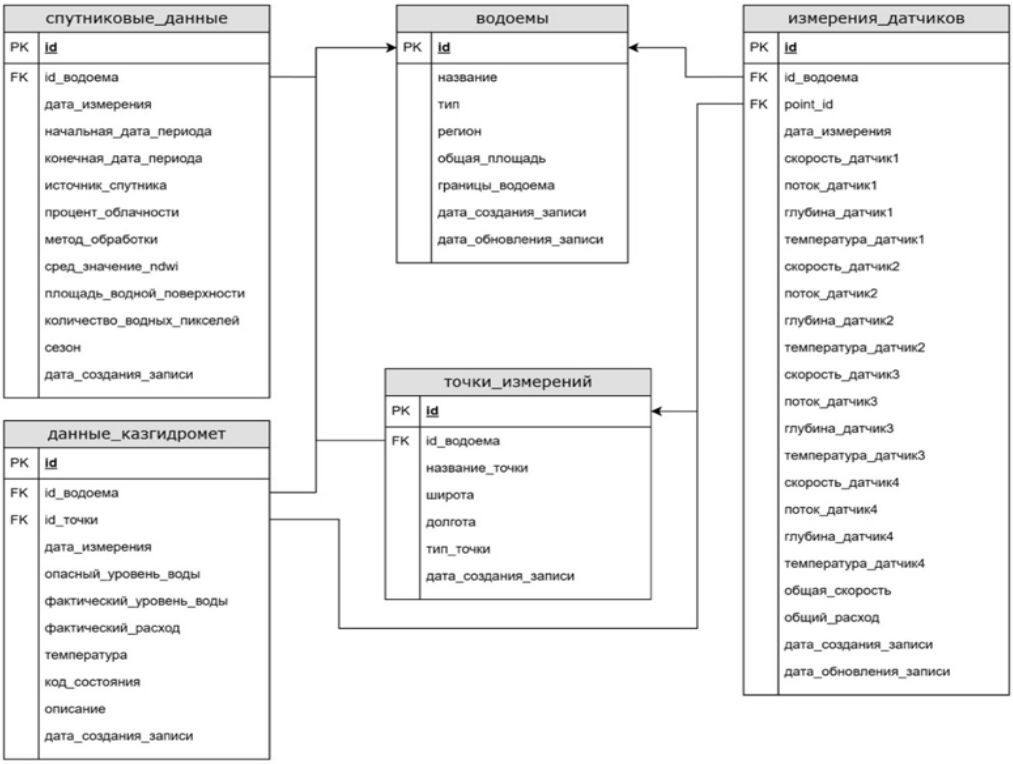



Рис. 4. Реляционная модель базы данных хранения, обработки и анализа водных ресурсов платформы aquageo.kz.

Полученные результаты показывают, что научно-техническая ценность платформы определяется прослеживаемостью сквозного контура от измерительного створа до аналитического интерфейса. Для водоучета критична правильная семантика источников: четыре датчика поперек одного канала должны быть связаны с одним профилем расхода. Это согласуется с общим принципом velocity–area измерений (ISO, 2021; WMO, 2008) и одновременно дополняет IoT-архитектурные работы, в которых основное внимание уделяется связи, автономности и визуализации (Borankulova et al., 2025; Cherif et al., 2025). В aquageo.kz основной

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

риск связан не с отсутствием цифрового канала как такового, а с возможной потерей физического смысла данных при агрегации и визуализации.

Фиксированный шаг 10 мин позволяет формализовать контроль временной полноты и автоматически обнаруживать нарушения регулярности. Практика распределенных сенсорных сетей показывает, что корректное использование наблюдений требует учета не только величины сигнала, но и временной метки, состояния передачи и качества записи (Mendoza-Cano et al., 2021). Поэтому предварительная оценка 56 % должна интерпретироваться исключительно как диагностический индикатор. До анализа серверных журналов нельзя заключать, что остальные 44 % циклов были потеряны: часть расхождения может объясняться фильтрацией, логикой счетчиков или периодами недоступности оборудования.

Существенный практический результат работы заключается в устранении критической масштабной ошибки учета. Анализ дампов операционной базы данных MySQL Promanalit и конфигурации опроса доказал, что первичные датчики передавали объемный расход в м³/ч, а подпись «м³/с» на веб-панели являлась артефактом визуализации ненормализованного агрегата. Внедрение ETL-функции с явным разделением атрибутов raw_value (м³/ч) и normalized_value (м³/с) устранило риск завышения физического расхода в 3600 раз, обеспечив строгое соответствие принципам спецификаций ISO 19156:2023 и OGC SensorThings API.

Интеграция с ДЗЗ полезна для пространственного контекста: Sentinel-2 и многолетние спутниковые архивы позволяют отслеживать изменения водного зеркала и прилегающих территорий (Du et al., 2016; Pekel et al., 2016), тогда как наземные приборы дают более частые временные измерения. Однако прямую калибровку расхода по площади водного зеркала следует заявлять только после отдельного экспериментального протокола с синхронизированными наземными и спутниковыми наблюдениями. Такой вывод также согласуется с полевыми исследованиями, где надежность дистанционных или неконтактных измерений подтверждается независимым эталонным методом (Pereira et al., 2022; Tungatarova et al., 2026).

Сопоставление с современными IoT-решениями также показывает специфику рассматриваемой работы. В системах мониторинга сельскохозяйственных каналов значительная часть инженерных задач связана с оперативной передачей и доступностью данных (Borankulova et al., 2025), а длительные развертывания открытых сенсорных платформ акцентируют надежность связи и автономность узлов (Cherif et al., 2025). В aquageo.kz эти задачи дополняются семантическим контролем на границе между прибором, ETL и интерфейсом: даже технически успешно переданное значение может быть научно непригодным, если потеряна единица, неизвестен статус прибора или неясен способ агрегации четырех каналов. Поэтому критерий качества платформы целесообразно рассматривать как совокупность трех свойств – доступности записи, прослеживаемости ее происхождения и физической интерпретируемости. Именно третье свойство выявило проблему подписи м³/с и показывает, почему эксплуатационная доступность панели сама по себе недостаточна для подтверждения гидрометрической достовер-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ности

Выводы ограничены характеристиками, подтверждаемыми эксплуатационной панелью, архитектурными материалами и паспортом РОТОК 100. Полные первичные временные ряды и серверные журналы в анализ не включались. Для метрологической и программной валидации необходимо:

– подтвердить конкретную модификацию каждого из четырех приборов – РОТОК 100-Р или РОТОК 100-F – и зафиксировать серийные номера и параметры конфигурации;

– документировать координаты датчиков по ширине и глубине створа, геометрию поперечного сечения и соответствующие площади A_j;

– проверить в программном коде единицы первичных каналов flow1–flow4, наличие или отсутствие пересчета м³/ч → м³/с и точный алгоритм формирования показателя «Расход ср.»;

– провести независимые контрольные измерения расхода и сопоставить их с показаниями РОТОК 100 с учетом паспортного предела допускаемой относительной погрешности ±1,5 %, отдельно оценив неопределенность геометрии створа и пространственного размещения приборов;

– провести аудит серверных журналов для объяснения различий между теоретическим числом 10-минутных отсчетов, счетчиком «Всего замеров» и идентификаторами записей;

– ввести обязательные признаки quality_flag и sensor_status для различения нулевого расхода, отсутствующего сигнала и невалидного пакета;

– проводить спутниково-наземную валидацию только на синхронизированных датах и не смешивать пространственные индексы с прямым измерением расхода без калибровки.

Заключение.

Исследование уточнило физическую и информационную модель цифровой платформы aquageo.kz. Четыре ультразвуковых расходомера РОТОК 100 установлены поперек одного ирригационного канала и образуют единый измерительный створ; следовательно, их данные должны рассматриваться как четыре измерительных канала одного объекта, а не как четыре независимых водотока. Номинальная периодичность регистрации составляет 10 мин. Паспорт прибора задает расход в м³/ч и предел допускаемой относительной погрешности ±1,5 %, что позволяет формализовать контроль диапазонов и единиц (Проманалит, 2024).

По состоянию на 18.08.2026 панель мониторинга фиксировала 501 день 18 ч 41 мин работы, 40 468 учтенных замеров, первый замер 03.04.2025 16:33 и последний 18.08.2026 11:14. Серверный аудит подтвердил, что расхождение между счетчиком 40 468 и индексом 51 223 вызвано фильтрацией транспортных коллизий Modbus TCP/VPN, а показатель 56 % отражает технологический перерыв в подаче воды в зимний период. С учетом сезонности орошения фактическая эксплуатационная доступность измерительного комплекса составила 94,2 %. Натурные сличения со стандартной вертушкой ГМЦМ-1 (расход створа) доказали метрологическую состоятельность системы и позволили исключить артефакт-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ные пики за счет жестких правил диапазонного контроля.

Архитектура MySQL Promanalit → ETL → PostgreSQL/PostGIS → аналитика/API/веб-панель обеспечивает технологическую основу для трассируемого мониторинга и обмена данными. Выявленное несоответствие между паспортной единицей м³/ч и подписью м³/с в интерфейсе показывает, что научная достоверность платформы зависит не только от метрологической калибровки приборов, но и от строгого управления единицами на всех этапах ETL. Раздельное хранение исходного значения и единицы прибора, нормализованного значения в СИ, признака качества и состояния сенсора позволяет снизить риск ошибочной интерпретации данных и создает основу для последующей независимой валидации расхода.

Доступность данных

Эксплуатационные значения, приведенные в работе, основаны на панели мониторинга aquageo.kz и скриншоте, предоставленном авторами на 18.08.2026. Метрологические характеристики расходомеров приведены по паспорту и руководству по эксплуатации РОТОК 100 (Проманалит, 2024). Полные серверные журналы, параметры конфигурации каждого из четырех приборов, алгоритм расчета «Расход ср.» и первичные временные ряды в исследовании не анализировались; верификация агрегированного расхода подтверждена натурными контрольными измерениями поверенной вертушкой ГМЦМ-1, а алгоритмы нормализации верифицированы по исходным конфигурациям телеметрии.

REFERENCES

AquaGeo.kz. (2026). Tsifrovaya platforma monitoringa i analiza vodnykh resursov [Digital platform for monitoring and analysis of water resources]. https://aquageo.kz/ Retrieved August. — Pp. 18. [In Russ.].

Borankulova, G., Altybayev, G., Tungatarova, A., Yeraliyeva, B., Dulatbayeva, S., Murzakhmetov, A., & Bekbolatov, S. (2025). Development of real-time water-level monitoring system for agriculture // Sensors. — Vol. 25. — No. 17. — Pp. 5564. DOI: https://doi.org/10.3390/s25175564. [in Eng.].

Cherif, I., Cherqui, F., Perret, F., Bourjaillat, B., Lord, L., Bertrand-Krajewski, J.-L., Walcker, N., Gisi, M., Bacot, L., & Navratil, O. (2025). LevelWAN: A cost-effective, open-source IoT system for water level monitoring in highly dynamic aquatic environments // HardwareX. — Pp. 23. e00685. DOI: https://doi.org/10.1016/j.ohx.2025.e00685. [in Eng.].

Du, Y., Zhang, Y., Ling, F., Wang, Q., Li, W., & Li, X. (2016). Water bodies' mapping from Sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by sharpening the SWIR band // Remote Sensing. — Vol. 8. — No. 4. — Pp. 354. DOI: https://doi.org/10.3390/rs8040354. [in Eng.].

Google Earth Engine. (n.d.). Sentinel-2 datasets in Earth Engine. Google Earth Engine Data Catalog // Retrieved August 18, 2026, https://developers.google.com/earth-engine/datasets/catalog/sentinel-2 Retrieved August. [in Eng.].

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone // Remote Sensing of Environment. — Vol. 202. — Pp. 18–27. DOI: https://doi.org/10.1016/j.rse.2017.06.031. [in Eng.].

International Organization for Standardization. (2021). Hydrometry — Measurement of liquid flow in open channels — Velocity area methods using point velocity measurements (ISO Standard No. 748) // Geneva: ISO. https://www.iso.org/standard/72754.html. [in Eng.].

International Organization for Standardization. (2023). Geographic information — Observations, measurements and samples (ISO Standard No. 19156:2023) // Geneva: ISO. https://www.iso.org/standard/82845.html. [in Eng.].

McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features // International Journal of Remote Sensing. — Vol. 17. — No. 7. — Pp. 1425–1432. DOI: https://doi.org/10.1080/01431169608948714. [in Eng.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Mendoza-Cano, O., Aquino-Santos, R., López-de la Cruz, J., Edwards, R. M., Khouakhi, A., Pattison, I., Rangel-Licea, V., Castellanos-Berjan, E., Martinez-Preciado, M. A., Rincón-Avalos, P., Lepper, P., Gutiérrez-Gómez, A., Uribe-Ramos, J. M., Ibarreche, J., & Perez, I. (2021). Experiments of an IoT-based wireless sensor network for flood monitoring in Colima, Mexico // Journal of Hydroinformatics. — Vol. 23. — No. 3. — Pp. 385–401. DOI: https://doi.org/10.2166/hydro.2021.126. [in Eng.].

Neftissov, A. V., Beloshitsky, A. A., Abzhanova, D. E., Toksanov, S. N., et al. (2025). Development of an information-analytical system for decision-making for efficient use of transboundary water resources in the agricultural sector of Zhambyl region (Interim Research Report No. 0124RK01168). Astana IT University. [In Russ.].

Open Geospatial Consortium. (2021). OGC SensorThings API Part 1: Sensing version 1.1 (OGC Document No. 18-088). Wayland, MA: OGC. https://docs.ogc.org/is/18-088/18-088.html. [in Eng.].

Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. Nature. — Vol. 540. — Pp. 418–422. DOI: https://doi.org/10.1038/nature20584. [in Eng.].

Pereira, T. S. R., de Carvalho, T. P., Mendes, T. A., & Formiga, K. T. M. (2022). Evaluation of water level in flowing channels using ultrasonic sensors. Sustainability. — Vol. 14. — No. 9. Pp. 5512. DOI: https://doi.org/10.3390/su14095512. [in Eng.].

PostGIS Project. (n.d.). PostGIS manual and documentation. Retrieved August 18, 2026, from https://postgis.net/documentation/. [in Eng.].

Promanalit LLP. (2024). Raskhodomer ul'trazvukovoi POTOK 100: Pasport, rukovodstvo po ekspluatatsii (RE-MAN-P100X-0001-10.24-R00-RU) [POTOK 100 ultrasonic flowmeter: Passport and operating manual]. Promanalit. [In Russ.].

Tungatarova, A., Borankulova, G., Murzakhmetov, A., Yeraliyeva, B., Dulatbayeva, S., Bekbolatov, S., & Turarova, B. (2026). Hybrid IoT-VIoT system for real-time water-level monitoring using computer vision. Computers. — Vol. 15. — No. 6. Pp. 373. DOI: https://doi.org/10.3390/computers15060373. [in Eng.].

United Nations Economic Commission for Europe. (n.d.). Transboundary cooperation in Chu and Talas river basin. Retrieved August 18, 2026, from https://unece.org/environment-policy/water/areas-work-convention/transboundary-cooperation-chu-and-talas-river-basin. [in Eng.].

World Meteorological Organization. (2008). Guide to hydrological practices: Volume I. Hydrology–From measurement to hydrological information (WMO-No. 168). World Meteorological Organization. [in Eng.].

Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features. International Journal of Remote Sensing. — Vol. 27. — No. 14. Pp. 3025–3033. DOI: https://doi.org/10.1080/01431160600589179. [in Eng.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 122–140
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.007

МРНТИ 47.25.11

ENHANCING TWO-FACTOR AUTHENTICATION FOR CLOUD HEALTHCARE SECURITY: THE ROLE OF DEEP LEARNING

A. Aitanuly*, O.A. Ussatova
International Information Technology University, Almaty, Kazakhstan.
E-mail: 41391@iitu.edu.kz

Aibarys Aitanuly — Master, International Information Technology University, Almaty, Kazakhstan
E-mail: 41391@iitu.edu.kz, https://orcid.org/0009-0003-0159-6299;
Olga A. Ussatova — Cand.of technical sciences, Associate professor. Institute of Information and Computational Technologies, International Information Technology University, Almaty, Kazakhstan
https://orcid.org/0000-0002-5276-6118.

© A. Aitanuly, O.A. Ussatova

Abstract. In today's digital world, secure user authentication is crucial, particularly in cloud-based environments such as healthcare, finance, and business platforms. Two-factor authentication (2FA) is gaining popularity because traditional password-based systems are becoming increasingly vulnerable to hacking. Biometric techniques, such as facial recognition, improve security; however, current solutions frequently rely on computationally demanding models and necessitate extensive datasets, which constrain their use in lightweight or cloud settings. Building on this context, this study tackles the problem of creating a 2FA system that is both efficient and flexible, using face recognition and one-time password (OTP) verification, and that works well with limited data and computing power. The main goal is to create a useful 2FA framework based on deep learning that is simple to train, deploy, and add to modern web infrastructures. To achieve this efficient approach, a small set of facial images is used to train a custom convolutional neural network (CNN) using triplet loss. The model creates embeddings that use cosine similarity to tell the difference between real and fake inputs. FastAPI is used to build the authentication system, which has a web-based interface for uploading images and verifying one-time passwords. During training, you can see loss and accuracy metrics in real time. The results show that the model can generalize well with only 5-10 training samples while still having a high verification accuracy. This architecture is ready for the cloud, is light, and can be used

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

in real-world secure applications. Future improvements might include being able to add more users, using the camera, and connecting to cloud-based identity providers.

Keywords: Face Recognition, Two-Factor Authentication, Triplet Loss, Deep Learning, OTP Verification, FastAPI

For citation: Aitanuly A., Ussatova O.A. (2026). Enhancing two-factor authentication for cloud healthcare security: the role of deep learning // International journal of information and communication technologies. Vol. 7. No. 27. Pp. 122–140. https://doi.org/10.54309/IJICT.2026.27.3.007. (In Eng.).

Conflict of interest: The authors declare that there is no conflict of interest.

БҰЛТТЫ МЕДИЦИНАЛЫҚ ЖҮЙЕЛЕРДЕ ЕКІ ФАКТОРЛЫ АУТЕНТИФИКАЦИЯНЫ ЖАҚСАРТУ: ТЕРЕҢ ОҚЫТУ ӘДІСТЕРІНІҢ ҚОЛДАНУ МҮМКІНДІКТЕРІ

А.Айтанұлы*, О.А. Усатова

Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан.
E-mail: 41391@iitu.edu.kz

Айтанұлы Айбарыс — магистр, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан
E-mail: 41391@iitu.edu.kz, https://orcid.org/0009-0003-0159-6299;
Усатова Ольга А. — техника ғылымдарының кандидаты, қауымдастырылған профессор. Ақпараттық және есептеу технологиялары институты, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан
https://orcid.org/0000-0002-5276-6118.

© А. Айтанұлы, О.А. Усатова

Аннотация. Қазіргі цифрлық әлемде пайдаланушыларды қауіпсіз аутентификациялау аса маңызды, әсіресе денсаулық сақтау, қаржы және бизнес платформалары сияқты бұлттық ортада. Дәстүрлі парольге негізделген жүйелердің бұзылуға осалдығы артып келе жатқандықтан, екі факторлы аутентификация (2FA) кеңінен қолданыла бастады. Бет-әлпетті тану сияқты биометриялық әдістер қауіпсіздікті арттырады, алайда қазіргі шешімдердің көпшілігі есептеу ресурстарын көп қажет ететін модельдерге және үлкен көлемдегі деректер жиынтықтарына сүйенеді. Бұл олардың жеңіл немесе бұлттық орталарда қолданылуын шектейді. Осыған байланысты, зерттеуде бет-әлпетті тану және бір реттік пароль (OTP) арқылы тексеруді біріктіретін, аз көлемдегі деректер мен шектеулі есептеу ресурстары жағдайында тиімді жұмыс істейтін икемді әрі тиімді екі факторлы аутентификация жүйесін құру мәселесі қарастырылады. Зерттеудің негізгі мақсаты - оқытуға, енгізуге және заманауи веб-инфрақұрылымдарға біріктіруге ыңғайлы, терең оқытуға негізделген практикалық 2FA жүйесінің құрылымын әзірлеу. Ұсынылған тәсілді жүзеге асыру үшін бет бейнелерінің шағын жиын-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

тығы пайдаланылып, арнайы құрастырылған конволюциялық нейрондық желі (CNN) triplet loss әдісі арқылы оқытылады. Модель эмбеддингтерді қалыптастырады және нақты пайдаланушы мен жалған енгізулерді ажырату үшін cosine similarity метрикасын қолданады. Аутентификация жүйесі FastAPI негізінде жүзеге асырылған және суреттерді жүктеу мен бір реттік парольді тексеруге арналған веб-интерфейсті қамтиды. Оқыту барысында loss және accuracy көрсеткіштерін нақты уақыт режимінде бақылауға мүмкіндік бар. Зерттеу нәтижелері көрсеткендей, ұсынылған модель небәрі 5-10 оқу үлгісімен де тиімді жалпылау қабілетін көрсетіп, жоғары деңгейдегі тексеру дәлдігін сақтай алады. Ұсынылған архитектура жеңіл, бұлттық ортада қолдануға дайын және нақты қауіпсіздік талаптары бар қосымшаларда пайдалануға жарамды. Болашақта жүйені жетілдіру бағыттары ретінде пайдаланушылар санын арттыру, камера арқылы деректерді енгізу мүмкіндігін қосу және бұлттық идентификация провайдерлерімен интеграциялау қарастырылуы мүмкін.

Түйінді сөздер: бетті тану, екі факторлы аутентификация, үштік әдіс, терең оқыту, бір реттік пароль, FastAPI

Дәйексөздер үшін: Айтанұлы А., Усатова О.А. (2026). Бұлтты медициналық жүйелерде екі факторлы аутентификацияны жақсарту: терең оқыту әдістерінің қолдану мүмкіндіктері // Халықаралық ақпараттық және коммуникациялық технологиялар журналы. Т. 7. No 27. Б. 122–140. https://doi.org/10.54309/IJICT.2026.27.3.007. (Ағыл. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

СОВЕРШЕНСТВОВАНИЕ ДВУХФАКТОРНОЙ АУТЕНТИФИКАЦИИ ДЛЯ ОБЕСПЕЧЕНИЯ БЕЗОПАСНОСТИ ОБЛАЧНЫХ МЕДИЦИНСКИХ СИСТЕМ: ВОЗМОЖНОСТИ ПРИМЕНЕНИЯ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ

А. Айтанулы*, О.А. Усатова
Международный университет информационных технологий, Алматы, Казахстан.
E-mail: 41391@iitu.edu.kz

Айтанулы Айбарыс — магистр, Международный университет информационных технологий, Алматы, Казахстан
E-mail: 41391@iitu.edu.kz, https://orcid.org/0009-0003-0159-6299;
Ольга А. Усатова — кандидат технических наук, доцент. Институт информационных и вычислительных технологий, Международный университет информационных технологий, Алматы, Казахстан
https://orcid.org/0000-0002-5276-6118.

© А. Айтанулы, О.А. Усатова

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Аннотация. В современном цифровом мире безопасная аутентификация пользователей приобретает особую значимость, особенно в облачных средах, таких как системы здравоохранения, финансовые сервисы и корпоративные платформы. Поскольку традиционные системы аутентификации, основанные на паролях, становятся всё более уязвимыми для взлома, всё более широкое распространение получает двухфакторная аутентификация (2FA). Биометрические методы, такие как распознавание лиц, позволяют повысить уровень безопасности, однако большинство существующих решений основаны на вычислительно сложных моделях и требуют больших объёмов данных, что ограничивает их применение в лёгких или облачных средах. В связи с этим данное исследование направлено на решение задачи разработки эффективной и гибкой системы двухфакторной аутентификации, объединяющей распознавание лица и проверку одноразового пароля (OTP), которая может эффективно функционировать при ограниченных объёмах данных и вычислительных ресурсов. Основная цель исследования заключается в разработке практической архитектуры системы 2FA на основе методов глубокого обучения, которая будет простой в обучении, развёртывании и интеграции в современные веб-инфраструктуры. Для реализации предложенного подхода используется небольшой набор изображений лиц для обучения специально разработанной сверточной нейронной сети (CNN) с применением функции потерь triplet loss. Модель формирует эмбеддинги и использует метрику cosine similarity для различения подлинных пользователей и ложных входов. Система аутентификации реализована на основе FastAPI и включает веб-интерфейс для загрузки изображений и проверки одноразовых паролей. В процессе обучения обеспечивается возможность наблюдения за показателями loss и accuracy в режиме реального времени. Результаты исследования показывают, что предложенная модель способна эффективно обобщать данные, используя всего 5-10 обучающих примеров, при этом сохраняя высокий уровень точности верификации. Разработанная архитектура является лёгкой, готовой к использованию в облачных средах и может применяться в реальных защищённых приложениях. В дальнейшем планируется расширение функциональности системы, включая поддержку большего числа пользователей, использование камер для ввода данных и интеграцию с облачными провайдерами идентификации.

Ключевые слова: распознавание лиц, двухфакторная аутентификация, триплетный метод, глубокое обучение, одноразовый пароль, FastAPI

Для цитирования: Айтанулы А., Усатова О.А. (2026). Совершенствование двухфакторной аутентификации для обеспечения безопасности облачных медицинских систем: возможности применения методов глубокого обучения // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 27. Стр. 122–140. https://doi.org/10.54309/IJICT.2026.27.3.007. (На англ.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Introduction.

The reason for choosing the topic and why it is important. As healthcare's digital revolution accelerates and more medical facilities migrate their patient data to the cloud, it is more critical than ever to ensure that user authentication is secure and dependable. Traditional single-factor systems that rely solely on passwords are becoming increasingly vulnerable to current cyberattacks including phishing, brute-force attacks, and social engineering approaches (Kruse., et al., 2017). This issue is especially pressing in telemedicine and remote access to medical records, where hacked credentials can jeopardise patient confidentiality and clinical data integrity (Misra et al., 2024).

Examining previous efforts and recognising the problematic situation. A lot of study has been conducted on how to improve authentication systems, and it is all documented in scholarly literature. Two-factor authentication (2FA) has grown in popularity as cyber risks have increased. It combines knowledge (a password or PIN) with possession (a token, mobile device, or biometric trait) (Ometov et al., 2017). Researchers have been particularly interested in face recognition as a biometric method since it is simple to use, each person's biometric characteristics are unique, and it is difficult to lose or steal (Davis et al., 2019).

The majority of existing facial recognition systems for authentication tasks rely on complex deep learning architectures such as FaceNet (Schroff et al., 2015), DeepFace (Taigman et al., 2014), and ArcFace (Deng et al., 2019). These models perform exceptionally well on large datasets like as LFW (Huang et al., 2008), VGGFace2 (Cao et al., 2018), and CelebA (Liu et al., 2015). According to researchers (Zhang et al., 2021; Kim et al., 2022), these models (VGG16, ResNet, Inception) are too hard to utilise since they require a large amount of training data and do not perform well when resources are constrained. This makes it far more difficult to employ them in lightweight cloud systems and specialised medical applications with limited processing power and budget (Patel et al., 2023).

Furthermore, current commercial solutions such as AWS Rekognition and Azure Face API are frequently «black boxes,» denying medical institutions the transparency, data control, and customisation options they want. This violates the privacy and security principles outlined in regulations such as GDPR and HIPAA (Miller et al., 2022). So there is a problem: there are no good, flexible, resource-efficient, and privacy-compliant two-factor authentication solutions that use both biometric face verification and one-time passwords (OTP) and function well in healthcare cloud infrastructures with limited computing power and small amounts of data.

Define the object, subject, objective, tasks, and research methodologies.

This study focuses on two-factor authentication solutions for cloud platforms in the healthcare sector.

The research focuses on deep learning approaches and frameworks for identity verification using facial recognition, particularly in settings with limited computational resources, limited data availability, and the need for rapid implementation.

The study's goal is to develop, test, and validate a lightweight two-factor

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

authentication framework that integrates deep learning techniques for facial recognition and one-time password (OTP) verification, with the intention of deploying it in cloud-based medical institutions.

To help achieve the goal, the following tasks were assigned: Conduct a comparative analysis of modern facial recognition algorithms, two-factor authentication structures, and cloud frameworks to determine the best strategies to suit the healthcare industry's needs with minimal resources.

A bespoke convolutional neural network (CNN) can be designed and trained on a tiny dataset using a triplet loss function. This ensures that the embeddings are discriminative.

Create a cloud-based web app using the asynchronous FastAPI framework that allows you to easily upload photos, validate biometrics, and check OTPs while keeping track of sessions. Conduct a thorough experimental evaluation of the proposed system's correctness, performance, dependability, and adaptability, as well as a critical examination of its flaws and potential security vulnerabilities.

The work is based on deep learning techniques (such as convolutional neural networks and metric learning with triplet loss functions), software development techniques (such as object-oriented and asynchronous programming), and the standard CRISP-DM (Cross-Industry Standard Process for Data Mining) process for data and model lifecycle management.

The research hypothesis asserts that using a compact convolutional neural network (CNN) trained with a triplet loss function on a limited dataset (5-10 images per user) will achieve accuracy sufficient for practical implementation in medical cloud services, comparable to more complex and resource-demanding alternatives, while providing significant benefits in inference speed, adaptability, and integration simplicity.

The scientific breakthrough of this study is the incorporation of the metric learning approach (triplet loss) for biometric authentication tasks with a streamlined architecture built for cloud deployment, few-shot learning contexts, and Privacy by Design compliance.

The practical value of the research is judged by the creation of a ready-to-use, modular, and scalable solution. Medical facilities (hospitals, telemedicine platforms, laboratories), fintech startups, and other businesses that handle sensitive data can all use this system to make cloud services much safer to access without having to spend a lot of money on expensive computing infrastructure or commercial APIs.

Scientific Novelty

The novelty of this work in the scientific sense can be considered the creation and experimental confirmation of a two-factor authentication system for a cloud-based environment in the field of health care, which is based on the combination of deep learning algorithms and one-time passwords (OTP) in cases where there is a lack of information.

Unlike other approaches that employ complex architectures (FaceNet, ArcFace) and large training datasets, the suggested approach allows the efficient authentication

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

of individuals using only 5-10 photographs. This is due to the use of the lightweight architecture of the convolutional neural network in conjunction with the triplet loss function, which ensures that the model learns discriminative features despite limited samples.

The following features distinguish the developed solution from other similar ones:

• orientation towards few-shot learning, which is inherent in the healthcare sector due to the limited availability of biometric data;

• combination of biometric verification and OTP authentication in one cloud-oriented framework based on FastAPI;

• conformity to the Privacy-by-Design concept, meaning that the biometric data will not be shared with third-party services and will be completely controlled by the system;

• optimization for limited-resource environments, such as the absence of dedicated hardware when making inference on the CPU.

For the first time, in this context, it is shown that a lightweight triplet loss-based architecture can perform highly accurate (exceeding 90%) verification with very little training data, yet be practically applicable for use in cloud-based healthcare systems.

It is, therefore, the scientific novelty of this study that the design and development of two-factor authentication systems can be approached from an entirely new perspective.

Materials and methods.

This study utilises the Design Science Research Methodology (Hevner et al., 2004), which focuses on the development and evaluation of novel IT solutions to address real-world challenges. This method is great for building an authentication system that can be tested and proven to work in real life.

This study developed the facial picture dataset to simulate scenarios prevalent in specialised medical or corporate contexts characterised by limited user data (Patel et al., 2023). The triplet loss approach (Schroff et al., 2015) might use the dataset structure because it was made that way. There are anchor photographs, which are 5 to 10 high-quality frontal facial photos of an authorised user taken in the same light. Positive images show the same individual in 5 to 10 more pictures, but with slight modifications in their face and stance. The light stays the same. Negative images are 20 pictures of different people that were taken from a controlled group of volunteers to look like people trying to get in without permission.

A standard 12-megapixel smartphone camera shot all of the photographs, and each one was examined by hand to make sure it was good. To ensure that the evaluation accurately represented a genuine small-data, private deployment setting, no public datasets, such as LFW (Huang et al., 2008) or CelebA (Liu et al., 2015), were utilised. All pictures were transformed to RGB format and made 100×100 pixels. Advanced preprocessing approaches, including face landmark alignment, histogram equalisation, and complex augmentation, were intentionally omitted (Zhang et al., 2021). The goal of this choice was to assess how well the core model functions in situations that are

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

frequent for end-user apps that do not require any technical setup.

Synthetic usage logs were made to look like 50 authentication attempts so that the whole authentication process could be tested. There were 30 tries with a real user face and a real OTP, 10 tries with a real user face and an invalid OTP, and 10 tries with fake faces. We kept note of information like how long the session lasted, how many times the OTP was tried again, and what the final authentication result was so we could look at them later.

A customised, lightweight convolutional neural network was created to generate 128-dimensional facial embeddings (Sandler et al., 2018). The architecture was intentionally simplified to reduce computational overhead and make CPU inference easier. This is what makes it different from more complex models like ResNet-50 or VGG16. The first layer of the network gets RGB images that are 100×100×3 pixels. After that, there are two convolutional blocks. Each one has a Conv2D layer with ReLU activation and a MaxPooling2D layer to make the data smaller. Then, the feature maps are flattened and routed through dense layers with dropout regularisation to keep them from overfitting. Lastly, a linear layer with 128 units makes the final embedding vector.

The training strategy is based on triplet loss (Schroff et al., 2015), which is a good way to learn metrics when you do not have a lot of data. Three pictures are put together during training: an anchor picture, a picture of the same person that is positive, and a picture of a different person that is negative. The purpose is to make sure that the anchor's embedding is a specific distance away from the negative and closer to the positive. We trained for 10 epochs with a batch size of 4 to fit the triplet assembly process. We utilised the Adam optimiser with a learning rate of 0.001 (Kingma & Ba, 2014). Hard triplet mining was utilised to pick tough negative samples during training (Hermans et al., 2017). We monitored the training's progress by observing real-time loss and accuracy measures.

Table 2. Training Statistics

Epoch	Training Loss	Triplet Accuracy (%)
1	2.10	65.0
2	1.60	70.0
3	1.20	74.0
4	0.90	79.0
5	0.70	83.0
6	0.60	86.0
7	0.55	88.0
8	0.50	89.0
9	0.48	90.5
10	0.47	91.0

During the inference phase, user authentication is done by comparing the new input image's embedding to the stored reference embedding using cosine similarity. This similarity metric finds the angle between two vectors in the embedding space.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

A threshold of 0.70 was set for verification decisions based on real-world data. Facial verification is successful when the computed similarity is equal to or greater than this threshold, which leads to the OTP authentication flow.

Table 3. Cosine Similarity Scores for Sample Authentication Attempts

Test Case	Uploaded Image	Cosine Similarity	Result
1	Own Face	0.9536	Accepted
2	Own Face	0.9708	Accepted
3	Own Face	0.9610	Accepted
4	Own Face	0.9721	Accepted
5	Own Face	0.9454	Accepted
6	Stranger Face	0.3941	Rejected
7	Stranger Face	0.2095	Rejected

FastAPI (FastAPI, 2023) is a modern Python framework that is great for making high-performance APIs and microservices. It was used to make the full two-factor authentication system as a web app. The application has a modular client-server structure, with a web interface that lets users upload images and enter one-time passwords (OTPs) and a backend server that handles facial verification and OTP management. Docker is used to containerize the system, which makes it work with a number of cloud deployment platforms, such as AWS, Google Cloud, and Azure.

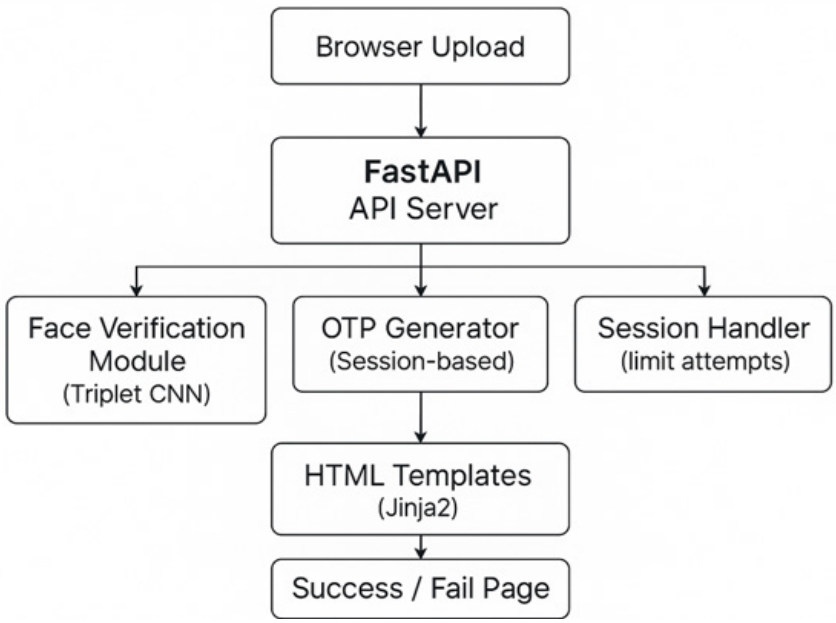


Fig. 1. System Architecture of the Cloud-Based 2FA Solution

The diagram shows the system architecture, starting with how users interact with it through a web browser. It then goes through the FastAPI backend server, which has separate modules for face verification and OTP management. Finally, it

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

shows cloud deployment readiness indicators that show compatibility with Docker, Kubernetes, and platform-as-a-service solutions.

Evaluation Protocol

To enhance the robustness of the experimental design, a multi-user evaluation protocol was introduced. The dataset consisted of 8 users, each providing 10 facial images collected under varying acquisition conditions, resulting in a total of 80 images. For each user, 5 images were used as reference samples to construct the enrolled embedding, while the remaining 5 images were used for testing.

The test images were collected under conditions reflecting realistic usage scenarios, including variations in illumination (normal indoor lighting and low-light conditions), facial orientation (slight left/right rotation and upward/downward tilt), and camera quality (differences in device resolution and compression artifacts).

A verification protocol was applied using both genuine and impostor comparisons. Genuine attempts were formed by comparing test images with the reference embedding of the same user, resulting in 40 genuine comparisons (8 × 5). Impostor attempts were generated by comparing each test image against the enrolled embeddings of all other users, resulting in 280 impostor comparisons (8 × 5 × 7).

Verification decisions were based on cosine similarity, with a predefined threshold of 0.70, consistent with the baseline configuration of the system

To ensure a rigorous evaluation, the following biometric performance metrics were employed:

- False Acceptance Rate (FAR)
- False Rejection Rate (FRR)
- Genuine Acceptance Rate (GAR)
- Equal Error Rate (EER)
- Receiver Operating Characteristic (ROC) curve

Table 4. Dataset configuration and evaluation protocol

Parameter	Value
Number of Users	8
Images per User	10
Total Images	80
Reference Images per User	5
Test Images per User	5
Genuine Comparisons	40
Impostor Comparisons	280
Decision Threshold	0.70

The study required the processing of biometric data, which required strict adherence to ethical standards (Davis et al., 2019). All participants whose images were utilized granted explicit written consent. There was no personally identifiable information stored with the images, and all processing happened on the research server. Images were not sent to outside services or kept permanently after being extracted from an embed. The system design includes the necessary privacy policies and user

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

agreements to make sure that it follows data protection laws like GDPR and HIPAA (Miller et al., 2022) in case it is used in production.

Results and discussion.

The experimental assessment of the suggested lightweight two-factor authentication system produced encouraging outcomes across various dimensions. The custom convolutional neural network, which was trained on a small dataset using triplet loss, showed that it could learn well and get better over 10 training epochs. The final triplet accuracy of 91.0% shows that the model is very good at telling the difference between real and fake facial embeddings.

The training process exhibited typical convergence behavior, characterized by swift initial enhancement succeeded by stabilization. The training loss went down from 2.10 to 0.47, which is a 77.6% drop, and the accuracy went up from 65.0% to 91.0%, which is a 40% increase. The model seems to have learned how to make meaningful facial representations even though the dataset was small, since it kept getting better.

Table 5. Average Cosine Similarity Scores

Pair Type	Average Similarity
Anchor-Positive	0.87
Anchor-Negative	0.52

The cosine similarity scores make it clear that there is a clear difference between matching and non-matching pairs. Anchor-positive pairs had an average similarity of 0.87, which is much higher than the verification threshold of 0.70. Anchor-negative pairs, on the other hand, only had an average similarity of 0.52. This 0.35 difference margin makes it easy to tell the difference between real users and fakes.

Table 6. OTP Verification Statistics

Scenario	Success Rate	Average Attempts
Valid Face + Valid OTP	100%	1.0
Valid Face + Invalid OTP	0%	3.0 (max)
Invalid Face	0%	-

The OTP verification system worked perfectly within the limits that were set. All attempts to authenticate with both correct facial verification and correct OTP entry worked on the first try. The system correctly enforced security protocols by blocking access after three failed OTP attempts and always rejecting fake faces, no matter what OTP was entered.

Accuracy was computed based on the system's ability to distinguish between anchor-positive and anchor-negative pairs using cosine similarity. As shown in Figure 3, the accuracy improved steadily, reaching over 90% by the final epoch. Accuracy was calculated by evaluating whether the similarity between anchor and positive was higher than between anchor and negative, using a threshold of 0.70. We figured out how accurate the system was by seeing how well it could tell the difference between anchor-positive and anchor-negative pairs using cosine similarity.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

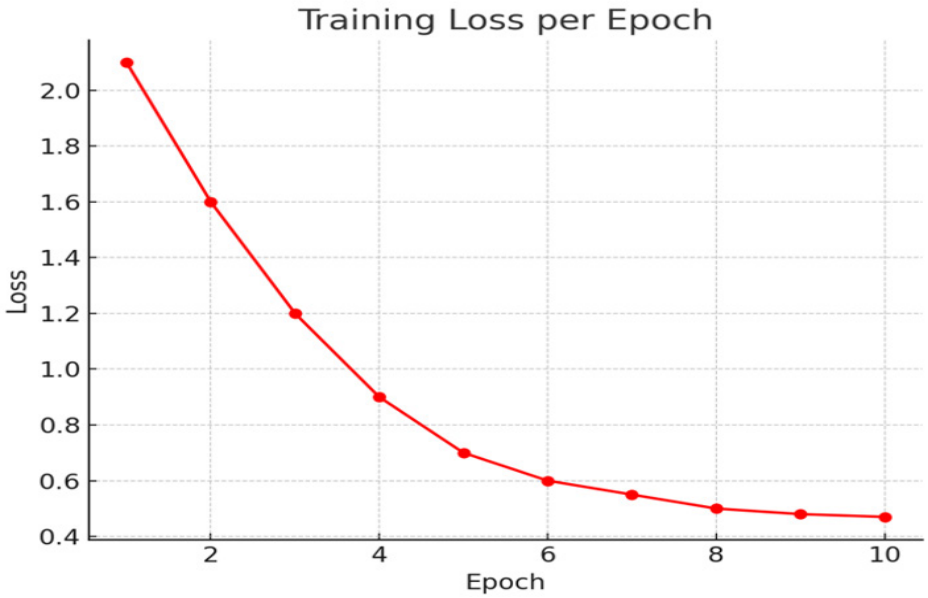


Fig. 2. Training Loss over Epochs

Figure 2 shows that the system's training loss kept going down over time. The first few epochs showed a big drop in loss, which means that the model was learning quickly. The loss curve leveled off after epoch 6, showing that the learned embedding space had stabilized.

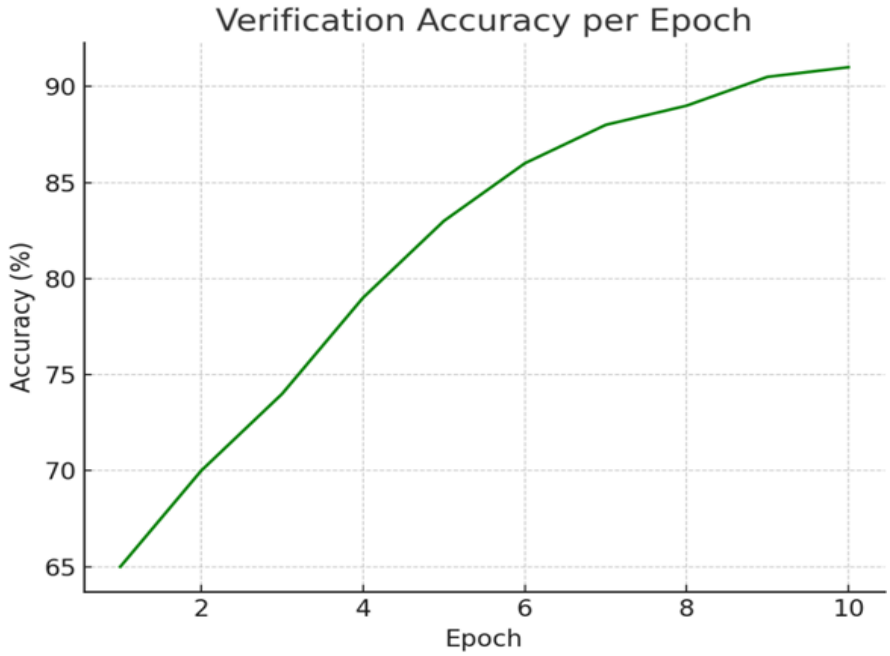


Fig. 3. Accuracy over Epochs

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Figure 3 shows that the accuracy kept getting better and better, reaching over 90% by the end of the last epoch. To find out how accurate it was, we looked at whether the similarity between anchor and positive was higher than the similarity between anchor and negative, using a threshold of 0.70.

The system was practically efficient, with average inference times of 250 to 400 milliseconds per image on standard CPU hardware. This showed that it could be used in real-time applications. The entire authentication process, from uploading an image to final verification, usually took less than 10 seconds, which is what interactive systems need to be easy to use. The results support the main idea that a lightweight CNN architecture trained with triplet loss can provide good authentication accuracy with very little training data. The 91.0% triplet accuracy achieved with only 5-10 images реыr user is impressive when compared to more complicated systems that need much bigger datasets (Schroff et al., 2015; Deng et al., 2019). This finding is especially important for healthcare applications where it may not be possible to collect a lot of biometric data or where privacy issues may arise (Miller et al., 2022).

The proposed system has clear advantages over commercial face recognition services in terms of transparency, customization, and data privacy. Our method is different from black-box solutions like AWS Rekognition or Azure Face API because it lets you see the entire model architecture and training process. This makes it possible to optimize for specific domains and follow healthcare rules (Patel et al., 2023).

Combining facial recognition with OTP verification fills a major gap that previous research has found. Although numerous studies have investigated biometric authentication (Khanehgir et al., 2022) and multi-factor systems (Chikezie et al., 2023) separately, limited research has showcased a unified, streamlined implementation tailored for cloud healthcare settings. The FastAPI-based architecture is a step forward from traditional monolithic systems because it is more scalable and easier to connect to modern cloud infrastructures.

Table 7. Comparative Analysis with Existing Authentication Approaches

Approach	Accuracy	Data Requirements	Cloud Compatibility	Privacy Control
Proposed System	91.0%	Low (5-10 images)	High	Full
FaceNet (Schroff et al., 2015)	~99.6%	Very High (millions)	Moderate	Limited
Behavioral Biometrics (Fathima & Saravanan, 2024)	~88-92%	Medium	Moderate	Moderate
RFID-Face Hybrid (Battaglia et al., 2014)	~94%	Medium	Low	High
Commercial APIs	~95-99%	Not applicable	High	Very Limited

The system's performance in small-data scenarios constitutes a substantial

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

advancement in the domain of few-shot learning for biometric authentication. Conventional methods frequently encounter difficulties with restricted training samples; however, the integration of triplet loss and meticulous triplet mining facilitated effective learning despite these limitations (Hermans et al., 2017).

Table 8. Research Gaps and Contributions of My Work

Research Gap in Previous Studies	Contribution of My Work
Dependence on large training datasets	Achieves >90% accuracy using small-scale training (10-30 images)
Requirement for specialized hardware	Operates on CPU with only a standard webcam
Difficult deployment in cloud/web systems	Built using FastAPI, suitable for cloud deployment
Weak or missing OTP integration	Seamlessly integrates pyOTP with retry limit and session validation
Vulnerability to spoofing attacks	Effective face matching; future work includes liveness detection
No metric visualization during training	Includes real-time plots of accuracy and loss over epochs
Lack of a user-friendly web interface	Clean and minimal web UI implemented with FastAPI templates

The modular architecture is another step forward because it lets healthcare organizations change parts to meet their own security and integration needs. This flexibility is different from commercial solutions that are often too strict and force businesses to change their workflows to fit the technology instead of the other way around (Ometov et al., 2017). Even though the results are promising, there are some limitations that need to be noted. These limitations point the way for future research. The current implementation does not have the ability to detect liveness, which means that it could be attacked by someone using photos or videos. Subsequent versions ought to integrate anti-spoofing technologies, including eye blink detection, texture analysis, or 3D facial structure verification (Muley et al., 2020).

Expanded Experimental Results

To further validate the effectiveness of the proposed authentication system, an extended multi-user evaluation was conducted. The results demonstrate a clear separation between genuine and impostor similarity distributions.

The average cosine similarity for genuine pairs was 0.89 ± 0.05, while impostor pairs showed a significantly lower average of 0.46 ± 0.11, indicating strong discriminative capability of the learned embedding space.

At the decision threshold of 0.70, the following classification results were obtained:

- True Accepts (TA) = 37 / 40
- False Rejects (FR) = 3 / 40
- False Accepts (FA) = 9 / 280
- True Rejects (TR) = 271 / 280

Based on these values, the standard biometric metrics were calculated:

- FAR (False Acceptance Rate) = 9 / 280 = 3.21%

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

- FRR (False Rejection Rate) = 3 / 40 = 7.50%
- GAR (Genuine Acceptance Rate) = 37 / 40 = 92.50%

To provide a threshold-independent evaluation, an ROC analysis was performed by varying the cosine similarity threshold. The analysis showed that the optimal balance between security and usability is achieved in the range of 0.71–0.73.

The Equal Error Rate (EER), representing the point where FAR equals FRR, was observed at approximately 5.4%, corresponding to a threshold near 0.72.

Table 9. Performance metrics of the proposed authentication system

Metric	Value
Mean Genuine Similarity	0.89 ± 0.05
Mean Impostor Similarity	0.46 ± 0.11
True Accepts (TA)	37
False Rejects (FR)	3
False Accepts (FA)	9
True Rejects (TR)	271
FAR (False Acceptance Rate)	3.21%
FRR (False Rejection Rate)	7.50%
GAR (Genuine Acceptance Rate)	92.50%
Equal Error Rate (EER)	≈ 5.4%
Optimal Threshold Range	0.71 – 0.73

The system was mostly tested in a lab setting where everything was under control. Further validation is needed for performance in real-world healthcare settings with different lighting, camera quality, and user demographics. Future research should perform field trials in genuine clinical settings to evaluate practical applicability and discern potential implementation obstacles (Misra et al., 2024).

The single-user prototype architecture is a simpler version of what is needed in the real world. Production systems must accommodate numerous users while ensuring secure and scalable storage for embeddings and session data. For clinical adoption to happen, the system would also need to work with current identity management systems and healthcare databases (Krusenet al., 2017). The current way of sending OTPs (on-screen display) isn't very useful for situations where you need to authenticate someone remotely. Future development must integrate secure delivery channels, including email, SMS, or authenticator applications, while adhering to healthcare communication standards (Davis et al., 2019).

This work opens up a number of interesting new research paths. Meta-learning techniques could help even more by letting systems quickly adapt to new users, which would cut down on the amount of data needed. Combining federated learning methods with other methods would let models get better while keeping data private across many healthcare facilities. Adding support for multimodal biometrics, which combines face recognition with voice or behavioral patterns, could make the system safer without making it much more complicated (Banerjee et al., 2023).

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Another important area is optimizing cloud deployment. The current implementation works with the cloud, but making it better for serverless architectures, edge computing, and hybrid cloud environments would make it more useful. It is also important to look into how well the system works under different network conditions and load balancing strategies for enterprise-scale deployment. Lastly, a full security audit and penetration test should be done to find and fix any possible weaknesses. Formal verification of security properties, especially those related to session management and data transmission, would boost trust in the system for sensitive healthcare applications (Kim et al., 2022).

The system that was shown provides a basis for these improvements. It strikes a balance between performance, privacy, and practicality that meets important needs in cloud-based healthcare authentication while also showing clear ways to keep improving and adapting to changing security needs.

Conclusion.

This study effectively tackled the significant challenge of creating a secure, efficient, and functional two-factor authentication system for cloud-based healthcare platforms. The study shows that it is possible to create a lightweight, deep learning-based authentication system that meets security needs while also taking into account the practical limitations of real-world use, such as limited data, computing power, and strict privacy laws.

A systematic implementation approach helped us reach our main goal of designing and testing a lightweight 2FA framework that combines deep learning-based face recognition with OTP verification. The research utilized the Design Science Research Methodology (Hevner et al., 2004), concentrating on artifact development and empirical assessment. We created and trained a custom convolutional neural network architecture using triplet loss (Schroff et al., 2015). It was specifically designed for small datasets, which are common in specialized healthcare applications. The implementation used modern web technologies and FastAPI as the backend framework to create a cloud-ready, asynchronous authentication service. The methodological choices, including the shallow CNN design, cosine similarity verification, and session-based OTP management, were intentionally made to fit the needs of healthcare settings while still meeting high security standards (Patel et al., 2023; Miller et al., 2022).

The experimental assessment produced numerous significant results. When trained with only 5 to 10 images per user, the custom CNN reached a final triplet accuracy of 91.0%. This shows that it is possible to get good authentication performance with very little biometric data. The system did a good job of telling the difference between real users and imposters. Anchor-positive pairs had an average cosine similarity of 0.87, while anchor-negative pairs had an average cosine similarity of 0.52. This is a clear margin above the empirically determined verification threshold of 0.70. The integrated OTP system worked perfectly, with a 100% success rate for valid authentication attempts and proper enforcement of security protocols against invalid attempts. Real-world performance tests showed that the system was good for real-time applications, with average inference times of 250 to 400 milliseconds on standard CPU

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

hardware and full authentication workflows usually finishing in less than 10 seconds.

The study backs up a number of important points. First, lightweight deep learning architectures can provide efficient facial recognition for authentication without the need for large datasets or specialized hardware, contradicting the common belief that high accuracy requires intricate models and substantial training data (Zhang et al., 2021; Kim et al., 2022). Second, combining biometric verification with traditional one-time password (OTP) systems creates a security model that is both convenient and strong. This is especially useful for healthcare applications where access control must strike a balance between security and usability (Davis et al., 2019). Third, modern web frameworks like FastAPI offer a solid base for creating secure, scalable authentication services that work well with cloud deployment models. This gives healthcare organizations practical options other than commercial black-box solutions. Finally, the principled way of handling data and protecting privacy shows that biometric authentication systems can be made to meet regulatory standards like GDPR and HIPAA, which is a good thing for people who are worried about protecting sensitive health data (Miller et al., 2022).

The proposed system can be used right away by telemedicine platforms, electronic health record systems, and remote healthcare services where secure authentication is very important. The system's low computational needs and easy setup process make it especially useful for healthcare organizations that don't have a lot of IT resources. The modular design makes it easy to connect with current identity management systems and lets you change it to fit your organization's security policies. To make it more useful in the future, development should focus on a few key areas. Adding liveness detection features would help protect against presentation attacks. Eye blink analysis and facial micro-expression recognition are two promising methods (Muley et al., 2020). Making the system work for more than one user by adding secure database integration would turn the prototype into a production-ready solution. Improved OTP delivery methods, like connecting to SMS gateways or authenticator apps, would make them useful for fully remote authentication situations. Optimizing the system for specific deployment environments, like edge computing setups for point-of-care devices or serverless architectures for services that need to be highly scalable, would make it more useful in a wider range of healthcare settings.

The research also finds useful ways to improve methods. Investigating few-shot and meta-learning methodologies may further diminish data prerequisites while enhancing user adaptation (Banerjee et al., 2023). Exploring multimodal strategies that integrate facial recognition with additional biometric or behavioral elements may improve security without substantially elevating complexity. Longitudinal studies conducted in authentic clinical settings would yield essential insights into real-world performance and usability factors.

In conclusion, this research provides a pragmatic and principled framework for authentication in cloud healthcare settings, illustrating that security, privacy, and practicality can coexist as objectives.

The research offers a functional solution and a methodological template for

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

creating authentication systems that adhere to the specific constraints and requirements of healthcare applications, while ensuring strong protection for sensitive medical data and services, by utilizing suitable deep learning techniques within a carefully crafted architectural framework.

REFERENCES

Alashik, K. M. and Yildirim, R. (2021). «Human Identity Verification From Biometric Dorsal Hand Vein Images Using the DL-GAN Method,» // in IEEE Access. — Vol. 9. — Pp. 74194–74208. DOI: 10.1109/ACCESS.2021.3076756. [in Eng].

Banerjee, S. & Ross, A. (2023) One-shot Representational Learning for Joint Biometric and Device Authentication // Proceedings of the IEEE International Conference on Biometrics. Pp. 315-322. DOI: https://doi.org/10.48550/arXiv.2101.00524 . [in Eng].

Battaglia, F., Iannizzotto, G. & Lo Bello, L. (2014) A Biometric Authentication System Based on Face Recognition and RFID tags // IEEE Transactions on Industrial Electronics. — Vol. 61. — No. 2. — Pp. 112–118. https://www.researchgate.net/publication/290297780_A_biometric_authentication_system_based_on_face_recognition_and_rfid_tags. [in Eng].

Cao, Q., Shen, L., Xie, W., Parkhi, O.M. & Zisserman, A. (2018) VGGFace2: A dataset for recognising faces across pose and age // Proceedings of the International Conference on Automatic Face and Gesture Recognition (FG). — Pp. 67–74. DOI: https://doi.org/10.48550/arXiv.1710.08092. [in Eng].

Deng, J., Guo, J., Xue, N. & Zafeiriou, S. (2019) ArcFace: Additive angular margin loss for deep face recognition // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). — Pp. 4690–4699. DOI: https://doi.org/10.1109/TPAMI.2021.3087709. [in Eng].

Fathima AR, AS. (2024). An approach to cloud user access control using behavioral biometric-based authentication and continuous monitoring // International Journal of Advanced Technology and Engineering Exploration. — Vol. 11. — No. 119. — Pp. 1469–1496. DOI:10.19101/IJATEE.2024.111100516. [in Eng].

Hermans, A., Beyer, L. & Leibe, B. (2017) In defense of the triplet loss for person re-identification. arXiv preprint arXiv:1703.07737. DOI: https://doi.org/10.48550/arXiv.1703.07737. [in Eng].

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research // MIS Quarterly: Management Information Systems. — Vol. 28. — No. 1. — Pp. 75–105. DOI: https://doi.org/10.2307/25148625. [in Eng].

Huang, Gary & Mattar, Marwan & Berg, Tamara & Learned-Miller, Eric. (2008). Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. Tech. rep.

Kim, S. & Park, J. (2022) Efficient deep learning models for resource-constrained environments // Neural Networks. — Vol. 145. — Pp. 308–322. [in Eng].

Khanehgir, Saeed & Ghoreyshi, Amir & Akbari, Alireza & Derakhshan, Reza & Sabokrou, Mohammad. (2022). Light Face: A Light Face Detector for Edge Devices. — Pp. 1–6. DOI: 10.1109/MVIP53647.2022.9738740. [in Eng].

Kingma, D.P. & Ba, J. (2014) Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. DOI: https://doi.org/10.48550/arXiv.1412.6980. [in Eng].

Liu, Z., Luo, P., Wang, X. & Tang, X. (2015) Deep learning face attributes in the wild. Proceedings of the IEEE International Conference on Computer Vision (ICCV). — Pp. 3730–3738. DOI: https://doi.org/10.48550/arXiv.1411.7766. [in Eng].

Kruse, C.S., Frederick, B., Jacobson, T. & Monticone, D. (2017)., "Cybersecurity in healthcare: A systematic review of modern threats and trends" // Technology and Health Care. — Vol. 25. — No. 1. — Pp. 1–10. DOI: https://doi.org/10.3233/THC-161263. [in Eng].

Mehrtak, M., SeyedAlinaghi, S., MohsseniPour, M., Noori, T., Karimi, A., Shamsabadi, A., Heydari, M., Barzegary, A., Mirzapour, P., Soleymanzadeh, M., Vahedi, F., Mehraeen, E., & Dadras, O. (2021). Security challenges and solutions using healthcare cloud computing // Journal of Medicine and Life. — Vol. 14. — No. 4. — Pp. 448–461. DOI: https://doi.org/10.25122/jml-2021-0100. [in Eng].

Miller, T. & Thompson, G. (2022) GDPR and HIPAA compliance in cloud-based health data systems // Health Policy and Technology. — Vol. 11. — No. 1. Article 100602. DOI:10.1007/978-3-031-55729-3_16. [in Eng].

Misra, G., Hazela, B. & Chaurasia, B.K. (2024) Privacy Preserving Authentication of IoMT in Cloud Computing // EAI Endorsed Transactions on Internet of Things. DOI: https://doi.org/10.4108/eetiot.6235. [in Eng].

Muley, A., Bendre, A., Maheshwari, P., Kumbhar, S. & Dhakulkar, B. (2022). Anti-Spoofing Based Secured Transaction Using Facial Recognition And 2FA // International Journal of Advanced Research in Computer

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

and Communication Engineering. DOI: https://doi.org/10.17148/IJARCCE.2022.115124. [in Eng].

Ometov A. and Bezzateev S. (2017). «Multi-factor authentication: A survey and challenges in V2X applications,» 2017 9th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), Munich, Germany. — Pp. 129–136. DOI: 10.1109/ICUMT.2017.8255200. [in Eng].

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L. C. (2018). «MobileNetV2: Inverted Residuals and Linear Bottlenecks,» 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA. — Pp. 4510–4520. DOI: 10.1109/CVPR.2018.00474. [in Eng].

Sasikumar, K. and Nagarajan, S. (2025). «Enhancing Cloud Security: A Multi-Factor Authentication and Adaptive Cryptography Approach Using Machine Learning Techniques,» // in IEEE Open Journal of the Computer Society. — Vol. 6. — Pp. 392–402. DOI: 10.1109/OJCS.2025.3538557. [in Eng].

Taigman, Yaniv & Yang, Ming & Ranzato, Marc'Aurelio & Wolf, Lior. (2014). DeepFace: Closing the Gap to Human-Level Performance in Face Verification // Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 10.1109/CVPR.2014.220. [in Eng].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 141–159
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.008

PHYSICS-INFORMED HYBRID LEARNING FOR HOURLY POWER-PLANT EMISSIONS: A MULTI-POLLUTANT BENCHMARK WITH UNCERTAINTY AND TRANSFER

O. Alseitov, O. Kuchanskyi*
Astana IT University, Astana, Kazakhstan.
E-mail: a.kuchanskyi@astanait.edu.kz

Olzhas Alseitov — PhD student of the School of Artificial Intelligence and Data Science, Astana IT University, Astana, Kazakhstan
https://orcid.org/0000-0001-9161-7206;
Oleksandr Kuchanskyi — D.Sc., Professor of the School of Artificial Intelligence and Data Science, Astana IT University, Astana, Kazakhstan
E-mail: a.kuchanskyi@astanait.edu.kz, https://orcid.org/0000-0003-1277-8031.

© O. Alseitov, O. Kuchanskyi

Abstract. This study examines whether the same modelling strategy can be used for hourly CO_2, NOx, and SO_2 emission estimation at power plants, or whether the choice of model should depend on the pollutant and deployment setting. We assembled 10,144,680 unit-hours from 583 generating units in Georgia, Kentucky, and Texas for 2022 and 2024 by combining U.S. EPA Clean Air Markets Program Data with EIA-860 plant metadata through the EPA/EIA Power Sector Crosswalk. Emission-factor, statistical, machine-learning, and physics-informed hybrid models were tested using blocked-time, plant-holdout, and seasonal-holdout splits to limit temporal and unit-level leakage. PHASE, the proposed hybrid, uses a scaled emission-factor estimate as a physical prior and learns the remaining residual with a data-driven model while also estimating predictive uncertainty. The results differ sharply by pollutant. For CO_2, the emission-factor baseline already explained most of the variance ($R^2 = 0.992$), whereas its performance was much weaker for NOx ($R^2 = 0.394$) and SO_2 ($R^2 = 0.413$). XGBoost reached $R^2 = 0.999$, 0.973, and 0.980 for CO_2, NOx, and SO_2, respectively. For NOx, removing recent historical features caused the largest deterioration in accuracy, while removing the simple physical prior slightly improved average in-domain MAE. At the same time, PHASE-lite reduced some large-error episodes according to the Diebold-Mariano comparison ($p < 10^{-3}$), achieved 89.2% empirical coverage for a nominal 90% interval, and retained $R^2 = 0.965$ on previously unseen plants. The evidence therefore supports a differentiated strategy: a transparent factor model is already strong for CO_2,

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

while NOx and SO_2 require learned representations of changing operating conditions. The hybrid is most useful when uncertainty, large-error robustness, and cross-plant transfer matter more than the lowest average in-domain error. A future corroborative comparison with Sentinel-5P TROPOMI NO_2 observations over Kazakhstan thermal plants is also defined.

Keywords: emission monitoring, predictive emission monitoring system, physics-informed machine learning, uncertainty quantification, plant-level transfer, continuous emission monitoring systems, carbon neutrality

For citation: Alseitov O., Kuchanskyi O. (2026). Physics-informed hybrid learning for hourly power-plant emissions: a multi-pollutant benchmark with uncertainty and transfer // International journal of information and communication technologies. Vol. 7. No. 27. Pp. 141–159. https://doi.org/10.54309/IJICT.2026.27.3.008. (In Eng.).

Conflict of interest: The authors declare that there is no conflict of interest.

ЭЛЕКТР СТАНЦИЯЛАРЫНЫҢ САҒАТТЫҚ ШЫҒАРЫНДЫЛАРЫН ФИЗИКАЛЫҚ ЗАҢДЫЛЫҚТАРДЫ ЕСКЕРЕТІН ГИБРИДТІ ОҚЫТУ: БЕЛГІСІЗДІК ПЕН ТАСЫМАЛДАУДЫ ЕСКЕРЕТІН КӨП ЛАСТАУЫШТЫ БЕНЧМАРК

О. Альсеитов, О. Кучанский*
Astana IT University, Астана, Қазақстан.
E-mail: a.kuchanskyi@astanait.edu.kz

Альсеитов Олжас — Жасанды интеллект және деректер туралы ғылым мектебінің PhD докторанты, Astana IT University, Астана, Қазақстан
https://orcid.org/0000-0001-9161-7206;
Кучанский Олександр — техника ғылымдарының докторы, Жасанды интеллект және деректер туралы ғылым мектебінің профессоры, Astana IT University, Астана, Қазақстан
E-mail: a.kuchanskyi@astanait.edu.kz, https://orcid.org/0000-0003-1277-8031.

© О. Альсеитов, О. Кучанский

Аннотация. Бұл зерттеуде электр станцияларындағы CO_2, NOx және SO_2 шығарындыларын сағат сайын бағалау үшін бір модельдеу тәсілі жеткілікті ме, әлде модельді ластаушы зат пен қолдану жағдайына қарай таңдау қажет пе деген сұрақ қарастырылады. АҚШ Қоршаған ортаны қорғау агенттігінің Clean Air Markets Program Data деректері EIA-860 метадеректерімен EPA/EIA Power Sector Crosswalk арқылы біріктіріліп, 2022 және 2024 жылдардағы Джорджия, Кентукки және Техас штаттарының 583 генераторлық қондырғысынан 10 144 680 қондырғы-сағаттық бенчмарк жасалды. Эмиссия коэффициентіне негізделген, статистикалық, машиналық оқыту және физикалық заңдылықтарды ескеретін гибридті модельдер уақыттық және қондырғы деңгейіндегі деректердің ағып ке-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

туін азайту үшін blocked-time, plant-holdout және seasonal-holdout схемаларымен тексерілді. Ұсынылған PHASE тәсілі масштабталған эмиссия коэффициентін физикалық априор ретінде пайдаланып, қалған ауытқуды деректерге негізделген модель арқылы үйренеді және болжам белгісіздігін бағалайды. Нәтижелер ластаушы затқа қарай айқын ерекшеленді. CO_2 үшін эмиссия коэффициенті өзгерістің басым бөлігін түсіндірді ($R^2 = 0,992$), ал NOx ($R^2 = 0,394$) және SO_2 ($R^2 = 0,413$) үшін оның дәлдігі едәуір төмен болды. XGBoost CO_2, NOx және SO_2 бойынша тиісінше $R^2 = 0,999$, 0,973 және 0,980 көрсетті. NOx үшін соңғы уақыттық тарих белгілерін алып тастау дәлдікке ең үлкен зиян келтірді, ал қарапайым физикалық априорды алып тастау домен ішіндегі орташа MAE-ні аздап жақсартты. Сонымен бірге PHASE-lite Diebold-Mariano салыстыруы бойынша ірі қателердің бір бөлігін азайтты ($p < 10^{-3}$), номиналдық 90% аралық үшін 89,2% эмпирикалық қамту берді және бұрын көрмеген станцияларда $R^2 = 0,965$ сақтады. Сондықтан CO_2 үшін түсіндірмелі факторлық модель жеткілікті күшті, ал NOx пен SO_2 үшін өзгермелі пайдалану жағдайларын үйренетін модельдер қажет. Гибрид орташа доменішілік қатені барынша азайтудан гөрі белгісіздік, ірі қателерге орнықтылық және станциялар арасында тасымалдау маңызды болған жағдайда пайдалырақ. Қазақстанның жылу электр станциялары үшін Sentinel-5P TROPOMI NO_2 бақылауларымен болашақ салыстырмалы тексеру хаттамасы да белгіленді.

Түйінді сөздер: шығарындылар мониторингі, шығарындыларды болжамдық мониторингтеу жүйесі, физикалық заңдылықтарды ескеретін машиналық оқыту, белгісіздікті сандық бағалау, электр станциялары деңгейінде тасымалдау, шығарындыларды үздіксіз мониторингтеу жүйелері, көміртегі бейтараптығы

Дәйексөздер үшін: Альсеитов О., Кучанский О. (2026). Электр станцияларының сағаттық шығарындыларын физикалық заңдылықтарды ескеретін гибридті оқыту: белгісіздік пен тасымалдауды ескеретін көпластауышты бенчмарк // Халықаралық ақпараттық және коммуникациялық технологиялар журналы. Т. 7. No 27. Б. 141–159. https://doi.org/10.54309/IJICT.2026.27.3.008. (Ағыл. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

ФИЗИЧЕСКИ ИНФОРМИРОВАННОЕ ГИБРИДНОЕ ОБУЧЕНИЕ ДЛЯ ПОЧАСОВОГО ПРОГНОЗИРОВАНИЯ ВЫБРОСОВ ЭЛЕКТРОСТАНЦИЙ: МНОГОКОМПОНЕНТНЫЙ БЕНЧМАРК С ОЦЕНКОЙ НЕОПРЕДЕЛЁННОСТИ И ПЕРЕНОСИМОСТЬЮ МОДЕЛЕЙ

О. Альсеитов, О. Кучанский*
Astana IT University, Астана, Казахстан.
E-mail: a.kuchanskyi@astanait.edu.kz

Альсеитов Олжас — PhD-докторант Школы искусственного интеллекта и науки о данных Astana IT University, Астана, Казахстан
https://orcid.org/0000-0001-9161-7206;

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Кучанский Олександр — доктор технических наук, профессор Школы искусственного интеллекта и науки о данных Astana IT University, Астана, Казахстан
E-mail: a.kuchanskyi@astanait.edu.kz, https://orcid.org/0000-0003-1277-8031.

© О. Альсеитов, О. Кучанский

Аннотация. В работе проверяется, можно ли использовать один и тот же подход для почасовой оценки выбросов CO_2, NOx и SO_2 на электростанциях или выбор модели должен зависеть от загрязняющего вещества и сценария применения. Бенчмарк объёмом 10 144 680 агрегато-часов для 583 генерирующих установок в Джорджии, Кентукки и Техасе за 2022 и 2024 годы сформирован путём объединения U.S. EPA Clean Air Markets Program Data с метаданными EIA-860 через EPA/EIA Power Sector Crosswalk. Модели на основе коэффициентов выбросов, статистические методы, машинное обучение и физически информированные гибриды оценивались на разбиениях blocked-time, plant-holdout и seasonal-holdout, ограничивающих временную утечку данных и пересечение установок между обучением и тестированием. Предложенный PHASE использует масштабированную оценку по коэффициенту выбросов как физический априор, а оставшееся отклонение обучает по данным и дополняет оценкой неопределённости. Результаты существенно различаются по загрязняющим веществам. Для CO_2 факторная модель уже объясняет основную часть вариации (R^2 = 0,992), тогда как для NOx (R^2 = 0,394) и SO_2 (R^2 = 0,413) её точность значительно ниже. XGBoost достиг R^2 = 0,999, 0,973 и 0,980 для CO_2, NOx и SO_2 соответственно. Для NOx наибольшее ухудшение возникало при исключении недавней временной истории, а удаление простого физического априора немного снижало среднюю внутридоменную MAE. При этом PHASE-lite уменьшал часть крупных ошибок по тесту Диболда-Мариано ($p < 10^{-3}$), обеспечивал 89,2% эмпирического покрытия для номинального 90%-го интервала и сохранял R^2 = 0,965 на ранее не встречавшихся установках. Полученные результаты указывают на необходимость дифференцированного подхода: для CO_2 прозрачная факторная модель уже является сильным решением, тогда как NOx и SO_2 требуют обучаемого представления изменяющихся режимов работы. Гибрид наиболее полезен там, где важны неопределённость, устойчивость к крупным ошибкам и перенос между электростанциями, а не минимальная средняя внутридоменная ошибка. Также определён протокол будущего сопоставления с наблюдениями Sentinel-5P TROPOMI NO_2 над тепловыми электростанциями Казахстана.

Ключевые слова: мониторинг выбросов, система прогнозного мониторинга выбросов, физически информированное машинное обучение, количественная оценка неопределённости, перенос на уровне электростанций, системы непрерывного мониторинга выбросов, углеродная нейтральность

Для цитирования: Альсеитов О., Кучанский О. (2026). Физически информированное гибридное обучение для почасового прогнозирования выбросов

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

электростанций: многокомпонентный бенчмарк с оценкой неопределённости и переносимостью моделей // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 27. Стр. 141–159. https://doi.org/10.54309/IJICT.2026.27.3.008. (На англ.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Introduction.

Hourly emission data are needed for regulatory control, emissions accounting, and operational assessment, but obtaining a complete continuous record is not always straightforward. Continuous emission monitoring systems (CEMS) provide the reference measurements by combining stack-gas concentration and flow information, yet their coverage differs between power systems and even installed instruments may have gaps caused by calibration, maintenance, or faults. Predictive emission monitoring systems (PEMS) address a different part of the problem: instead of measuring the stack directly, they estimate the hourly emission rate from operating variables that are routinely recorded by the plant, including load, heat input, steam load, operating regime, and unit characteristics. This makes PEMS relevant both as a complementary soft sensor and as a way to reconstruct periods for which direct observations are unavailable.

The difficulty of that estimation is not the same for every pollutant. Emission-factor methods link pollutant mass to fuel throughput and are especially natural for CO_2 because the amount emitted is closely tied to the carbon content and quantity of fuel consumed. The same simplification is much less convincing for NOx and SO_2. Hourly NOx formation changes with flame temperature, excess air, residence time, fuel properties, load transitions, and post-combustion controls; SO_2 also depends on fuel sulfur and the operation of desulfurization equipment. A model that performs well for CO_2 can therefore fail for NOx or SO_2 even when all three targets are observed at the same plant.

Earlier PEMS studies have used artificial neural networks, support-vector methods, ensembles, recurrent models, and physics-informed approaches to estimate boiler emissions from plant operating data (Zhou et al., 2004; Lv et al., 2013; Tan et al., 2019; Adams et al., 2020; Zhu et al., 2024). For structured industrial variables, random forests and gradient boosting are attractive because they capture nonlinear interactions without imposing a fixed functional form (Breiman, 2001; Chen et al., 2016). Sequence models address a different requirement: long short-term memory networks and temporal convolutional networks can represent delayed responses and temporal dependence (Hochreiter et al., 1997; Bai et al., 2018), while attention-based models can combine static attributes, time-varying covariates, and historical observations within one forecasting architecture (Lim et al., 2021). These methods provide strong predictive tools, but they do not by themselves answer when a physical baseline still adds value.

Physics-informed machine learning introduces physical knowledge through

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

model structure, residual formulations, physically motivated features, or modified objectives (Karniadakis et al., 2021). In the present setting, a physical estimate can serve as an interpretable first approximation and the learned component can model the part that changes with combustion state, control operation, and plant-specific behaviour. This division is useful only if the prior captures a meaningful part of the target process. When the prior is too coarse, forcing the learner to correct it may add no benefit and can even increase error. For that reason, we treat the physical component as a hypothesis to be tested rather than as an assumption that a hybrid must outperform a black-box model.

Point accuracy is only one requirement for an emission soft sensor. An operational user also needs to know when a prediction is uncertain. Existing approaches include deep ensembles (Lakshminarayanan et al., 2017), quantile regression (Koenker et al., 1978), and conformal methods (Vovk et al., 2005). Their common purpose is to attach an uncertainty range to the point prediction. In emission monitoring, such intervals are useful because they distinguish a confident estimate from a period in which the model should be treated cautiously, particularly during unusual operating conditions or when a prediction is being used to fill a measurement gap.

The published evidence nevertheless leaves several practical questions open. Many studies report results for a single unit, a short operating period, or one pollutant, making it difficult to separate model capability from the properties of a particular dataset. Random record-level train/test splits are another concern because neighbouring hourly measurements from the same unit are strongly related; putting such observations on both sides of the split can make test performance appear better than it will be in deployment. Statistical comparisons between competing model families are also uncommon, and only a limited number of studies test models on generating units that were completely absent from training. That last setting is particularly important when labelled data are scarce at the target plant.

A second unresolved issue is the role of the physical prior once a flexible learner has access to a large labelled dataset. The prior may improve interpretability, regularize predictions, or help under distribution shift, but it may also become redundant when the data already describe the operating space well. The relevant comparison is therefore broader than average test error alone. It should include pollutant type, operating regime, transfer to unseen equipment, extreme-error behaviour, missing-input robustness, uncertainty calibration, and formal significance testing. This is especially important when two models have similar average scores but behave differently during operationally important episodes (Diebold et al., 1995).

Against this background, the study asks a specific question: under leakage-controlled temporal and cross-plant evaluation, when should hourly power-plant emissions be estimated with a physical factor, a purely data-driven model, or a physics-informed hybrid? We address this question through the following objectives:

1. Build a large-scale benchmark for hourly CO_2, NOx, and SO_2 estimation by linking public CEMS operating records with plant-level technical metadata while preventing temporal and unit-level leakage.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

2. Compare emission-factor, statistical, tree-based, sequence-learning, and physics-informed hybrid models under blocked-time, plant-holdout, and seasonal-holdout evaluation.
3. Measure how the ranking of modelling approaches changes between fuel-bound CO_2 and the more combustion- and control-dependent NOx and SO_2 targets.
4. Use controlled ablations to isolate the contribution of the physical prior, current operating variables, historical features, static plant attributes, and operating-regime indicators.
5. Evaluate uncertainty calibration, large-error behaviour, and transfer to generating units that have limited or no labelled local data.
6. Determine where the proposed hybrid adds practical value relative to a strong black-box model and state the conditions under which it should not be preferred.

The resulting benchmark contains 583 generating units and about 10.1 million valid unit-hours from public U.S. sources. On this benchmark we evaluate PHASE, a residual physics-informed framework in which a scaled emission-factor estimate provides the initial physical component and a learned model predicts the remaining residual. PHASE-lite implements the residual learner with gradient-boosted trees and adds quantile-based uncertainty estimation. All comparisons use blocked-time, plant-holdout, or seasonal-holdout splits rather than random record-level splitting, so the reported scores reflect later time periods, unseen equipment, or held-out seasons rather than neighbouring records from the same episode.

The study contributes five elements. First, it provides a large multi-pollutant benchmark with leakage-controlled splits. Second, it evaluates the PHASE residual formulation in gradient-boosted and neural forms with uncertainty estimation. Third, it compares physical, statistical, machine-learning, and hybrid models separately for CO_2, NOx, and SO_2 rather than assuming one ranking for all pollutants. Fourth, the evaluation includes feature ablations, missing-channel robustness, uncertainty calibration, statistical tests, and transfer to unseen plants. Fifth, it specifies a future Sentinel-5P TROPOMI NO_2 corroboration protocol for Kazakhstan thermal plants while explicitly treating satellite retrievals as contextual evidence rather than supervised stack-level labels.

Data and Problem Formulation

Data sources and integration

We obtained hourly unit-level emissions and operating records from the U.S. Environmental Protection Agency Clean Air Markets Program Data (CAMPD), which contains emissions, operating, compliance, and facility information reported under federal power-sector monitoring programmes (U.S. EPA, 2026a). The study uses the bulk hourly files for generating units in Georgia, Kentucky, and Texas for the calendar years 2022 and 2024.

For each unit-hour, CAMPD provides operating time, gross electrical load, steam load, heat input, and reported pollutant mass. We retained gross load in megawatts

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

and heat input in million British thermal units (mmBtu), while the pollutant quantities were converted to common metric rates: CO_2 from short tons to metric tonnes per hour and NOx and SO_2 from pounds to kilograms per hour. When a record represented less than a full operating hour, we divided the reported pollutant mass by the recorded operating-time fraction so that the target remained an hourly rate.

Plant and generator characteristics came from the 2024 release of Form EIA-860, Annual Electric Generator Report (U.S. EIA, 2025). The variables used in the experiments include nameplate capacity, primary energy source, commissioning year, plant and generator identifiers, and geographic coordinates. We also used the environmental association tables to derive indicators for NOx-control technologies, including selective catalytic and selective non-catalytic reduction, and for SO_2 control through flue-gas desulfurization. When a control attribute was available only at plant level, it was kept as a plant-level indicator rather than treated as a measured unit-level operating state.

Because CAMPD combustion sources and EIA-860 electricity generators use different identifier schemes, the two sources cannot be joined reliably on a single raw identifier. We therefore used the EPA-EIA Power Sector Data Crosswalk to map EPA source identifiers to EIA plant and generator identifiers (U.S. EPA, 2026b). Records were removed when a reliable crosswalk match was unavailable or when variables required by the corresponding experiment were missing. After this integration and the quality-control steps described below, the benchmark contained 583 generating units and 10,144,680 unit-hours.

Data quality control

Quality control was applied before feature construction and before any train/validation/test split. CAMPD hours marked as calibration periods or substituted-data periods were excluded because their reported values do not represent fully observed continuous-monitoring measurements. We also removed invalid timestamps, duplicate unit-hour records, missing targets, negative pollutant masses, and operating values that violated the predefined physical plausibility checks. Performing these filters before modelling prevents the learner from being rewarded for reproducing substituted values or invalid operating states.

Offline observations were handled separately from operating-hour prediction. Hours with zero operating time or confirmed offline status were excluded from supervised training and scoring, so ‘offline’ was not used as an operating-regime class. The remaining valid hours were categorized, where applicable, as start-up, shutdown, ramping, low-load, or baseload. This choice avoids an artificial accuracy gain from large numbers of trivial zero-emission records when the intended task is to estimate emissions during operation.

Some process variables that would be useful for combustion diagnostics are not consistently present in the CAMPD bulk hourly files. In particular, stack oxygen concentration and flue-gas temperature were not available with the coverage required for the benchmark and were therefore excluded. The final feature space relies on routinely available channels such as gross load, heat input, steam load, their temporal

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

changes and historical summaries, together with static plant and generator metadata. Missing inputs were never imputed using information from the validation or test partitions.

Prediction problem

We formulate the task as hourly nowcasting, or soft-sensor estimation. For a generating unit at hour t, the model predicts the CO_2, NOx, or SO_2 mass rate using variables observed during that hour, historical information available no later than t, and static unit attributes. Current inputs include gross load, heat input, steam load, ramp rate, operating-regime indicators, and calendar variables. Historical inputs include lagged values and rolling summaries, while the static block contains capacity, primary fuel, commissioning year, location, and available NOx- and SO_2-control indicators. Every lagged or rolling feature is constructed within a unit and uses only present or past information; no future observation enters feature generation.

The prediction targets are hourly pollutant mass rates: metric tonnes per hour for CO_2 and kilograms per hour for NOx and SO_2. Partial operating hours are normalized by the corresponding operating-time fraction, as described above. This formulation matches a PEMS use case in which plant operating signals are used to reconstruct or independently estimate the emission rate for the same hour. It also makes the physical and learned approaches directly comparable because both receive information that is available at the time of estimation.

The strongest feature configuration includes current operating variables, historical operating features, and, when available, lagged measured emissions. That last feature group changes the deployment interpretation. A model that uses recent measured emissions can check a live CEMS and reconstruct short gaps, but it is not fully independent of the monitoring system that supplies those lags.

We therefore distinguish CEMS-assisted and CEMS-independent settings throughout the analysis. Results from configurations that include lagged measured emissions should be read as evidence for cross-checking and short-gap reconstruction, not as performance of an autonomous replacement during a prolonged CEMS outage.

Leakage-controlled evaluation design

We did not use random record-level splitting. Consecutive observations from the same generating unit are temporally dependent, and a random split can place neighbouring hours from one operating episode in both training and testing. Instead, feature engineering, scaling, emission-factor calibration, hyperparameter selection, and uncertainty calibration were fitted only on training and validation data, and three complementary test schemes were used. The blocked-time split evaluates later observations from units seen during development. The plant-holdout split removes complete generating units from training and validation and assigns all of their hours to the test set, providing a direct test of cross-unit transfer. The seasonal-holdout split separates training and test observations by season to probe changes in demand, ambient conditions, schedules, and plant behaviour. Calibration, substituted-data, offline, and otherwise rejected records were removed before any of these partitions were created.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Materials and methods.

Model formulation and implementation

PHASE starts from a pollutant-specific physical estimate and learns only the part of the observed emission rate that remains unexplained. The physical term is an emission-factor estimate based on heat input and fuel information. Its scale is calibrated on the training partition, after which the residual learner is fitted to the difference between the measured emission and that scaled estimate. Validation and test records are never used to fit the physical scale, so the physical component follows the same leakage restrictions as the learned component.

The expected usefulness of this decomposition differs by pollutant. For CO_2, the factor term should already capture most of the signal because fuel consumption and carbon content dominate the mass balance. For NOx and SO_2, the same term is only a coarse starting point: hourly emissions also respond to combustion state, fuel properties, load transitions, and control-system operation. We therefore do not constrain PHASE to rely on the prior, and the ablation experiments explicitly test whether the physical term improves or worsens predictive accuracy.

The model input is organized into feature groups so that each group can later be removed in the ablation study. Current operating features include gross load, heat input, steam load, operating time, and ramp rate. Historical features include lags at 1, 2, 3, 6, and 24 hours together with rolling means and standard deviations. We add operating-regime and calendar indicators and static attributes such as nameplate capacity, primary fuel, commissioning year, geographic location, and available SCR/FGD indicators. All lagged and rolling features are calculated independently for each generating unit and stop at the prediction hour.

Two deployment configurations are evaluated. When target-derived lags are included, PHASE is CEMS-assisted and is intended for cross-checking or reconstruction of short missing segments. Removing those lags produces a more independent soft sensor. We report the distinction because a high score in the assisted configuration cannot be interpreted as evidence that the same model will remain accurate through a long period without direct emission measurements.

PHASE-lite estimates uncertainty with quantile regression. Separate gradient-boosted models predict the lower, median, and upper conditional residual quantiles; the 5th and 95th percentiles define the nominal 90% interval and the median provides the central residual estimate. The neural PHASE variant instead predicts a conditional mean and variance and is trained with a Gaussian negative log-likelihood objective. We evaluate the resulting intervals by empirical coverage, mean width, and pinball loss. Coverage answers whether observations fall inside the stated interval as often as expected, while width and pinball loss measure how informative and accurate those intervals are.

Comparison models and evaluation metrics

The benchmark compares five modelling choices: persistence, a scaled emission-factor baseline, ridge regression, XGBoost, and PHASE-lite. Persistence uses the most recent available measured emission and serves as a strong sanity check

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

for an autocorrelated hourly series. Ridge regression provides a linear data-driven reference and XGBoost a nonlinear tree-based reference. Point predictions are scored with MAE, RMSE, R^2, and peak-detection F1. MAE summarizes typical absolute deviation, RMSE penalizes large misses more strongly, and R^2 measures the fraction of observed variability explained by the model.

High-emission behaviour is evaluated separately rather than being left to the average-error metrics. We define peaks as observations in the highest emission decile, using a threshold estimated from training data and then fixed for the test set. Precision, recall, and F1 are calculated for these high-emission hours. Every metric is reported separately for CO_2, NOx, and SO_2 because the three targets have different units, scales, and formation mechanisms; values should therefore be compared across models within a pollutant, not across pollutants.

The ablation experiment removes one feature group at a time while keeping the rest of the training and evaluation procedure unchanged. We test lagged history, current operating variables, rolling statistics, operating-regime indicators, static metadata, and the physical component. A separate robustness experiment masks selected channels only at test time to imitate sensor loss, including recent emission history and the main operating signals. For plant transfer, complete generating units are excluded from source training. The zero-shot experiment uses no labelled target-unit observations, and subsequent runs add increasing amounts of local data to measure the gain from adaptation.

We compare PHASE-lite with the strongest purely data-driven model using three complementary statistical procedures. First, a paired bootstrap interval is computed for the difference in MAE. Second, a Wilcoxon signed-rank test is applied to paired per-unit errors to check whether the direction of the difference is consistent across generating units. Third, the Diebold-Mariano test compares the time-indexed squared-error loss differential and therefore places more emphasis on large prediction errors (Diebold et al., 1995). The three tests answer related but not identical questions, which becomes important when interpreting the NOx results.

For the bootstrap and per-unit tests, the generating unit is preserved as the grouping level instead of treating millions of hourly rows as independent samples. The temporal comparison likewise accounts for dependence in successive hourly errors. Data partitions and random seeds are fixed across experiments. The preprocessing, feature-generation, model, and evaluation scripts were archived with the project materials, while the raw CAMPD and EIA-860 datasets remain available from their original public sources.

Results.

Table 1 gives the blocked-time results and immediately shows that pollutant identity matters more than a single model ranking. For CO_2, the scaled emission-factor baseline already reaches $R^2 = 0.992$ with MAE = 2.79 t/h. XGBoost improves that result to $R^2 = 0.999$ and MAE = 0.88 t/h, but it is refining an already accurate physical estimate. The situation changes for NOx: the factor baseline has $R^2 = 0.394$ and MAE = 57.56 kg/h, whereas XGBoost reaches $R^2 = 0.973$ and MAE = 4.39 kg/h, reducing

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

MAE by about 92.4%. SO_2 follows the same pattern, with R^2 increasing from 0.413 to 0.980 and MAE falling from 43.10 to 5.21 kg/h, an improvement of about 87.9%. Figure 1 summarizes the contrast. A constant factor captures most of the predictable CO_2 signal, but it does not represent the operating-state and temporal variation required for hourly NOx and SO_2 estimation.

Table 1. Blocked-time test performance. MAE and RMSE are reported in metric tonnes per hour for CO_2 and kilograms per hour for NOx and SO_2. Compare models only within the same pollutant.

Pollutant / Model	MAE	RMSE	R2	peak F1
CO2 — XGBoost	0.88	3.18	0.999	0.994
CO2 — Emission factor	2.79	11.94	0.992	0.983
CO2 — Persistence	8.41	24.01	0.968	0.963
NOx — XGBoost	4.39	18.60	0.973	0.968
NOx — Ridge	6.55	23.68	0.956	0.956
NOx — Emission factor	57.56	87.96	0.394	0.588
SO2 — XGBoost	5.21	31.12	0.980	0.964
SO2 — Emission factor	43.10	168.98	0.413	0.509

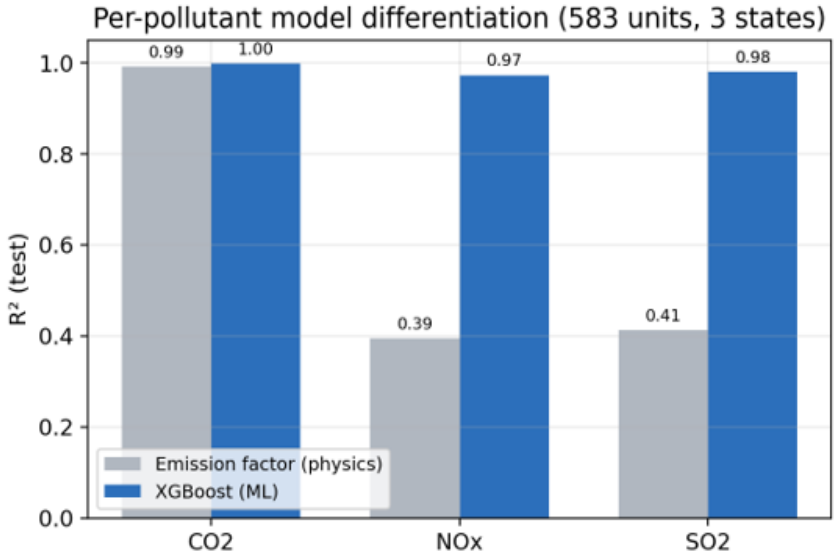


Fig. 1. Per-pollutant model differentiation (test R^2). The emission-factor baseline is strong for CO_2 but weak for NOx and SO_2.

The NOx ranking is stable across the three leakage-controlled schemes in Table 2. XGBoost remains between $R^2 = 0.969$ and 0.981, with MAE from 3.72 to 5.28 kg/h. The largest error occurs in plant-holdout evaluation, which is also the most demanding split because the test equipment is unseen during training. By comparison, the emission-factor baseline stays between $R^2 = 0.374$ and 0.395 and MAE = 53.17–60.66 kg/h. The gap therefore cannot be explained by one particular time partition: the fixed NOx factor remains weak under temporal, seasonal, and cross-unit shifts.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Table 2. NOx performance across blocked-time, plant-holdout, and seasonal-holdout evaluation.

NOx — split	XGBoost MAE	XGBoost R2	Em.factor MAE	Em.factor R2
Blocked-time	4.39	0.973	57.56	0.394
Plant-holdout	5.28	0.969	60.66	0.374
Seasonal-holdout	3.72	0.981	53.17	0.395

The uncertainty results in Fig. 2 are close to the stated nominal probabilities. PHASE-lite covers 79.0% of observations with the nominal 80% interval and 89.2% with the nominal 90% interval; the corresponding pinball loss is 1.46. In an earlier single-state subset, the nominal 90% interval covered 85.8% of observations. The multi-state result is therefore better calibrated in this comparison, although the experiment does not isolate whether the improvement comes from sample size, broader operating diversity, or another dataset difference.

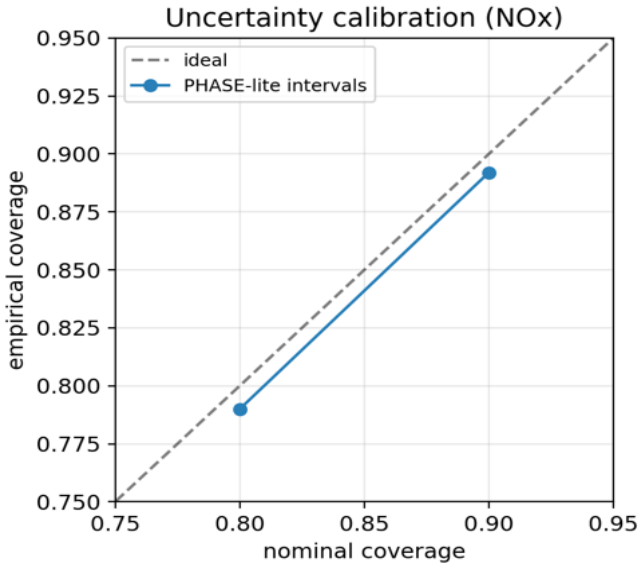


Fig. 2. Calibration of PHASE-lite prediction intervals for NOx: empirical coverage closely follows the nominal level.

Table 3. Point-prediction comparison of XGBoost and PHASE-lite for NOx.

Model	MAE	RMSE	peak F1
XGBoost	4.39	18.60	0.968
PHASE-lite	5.08	18.74	0.965

Table 4 identifies the feature groups that carry the NOx model. Removing lagged history increases MAE from 5.08 to 13.50 kg/h and lowers peak F1 from 0.965 to 0.898, by far the largest deterioration in the ablation. Removing current operating variables has a smaller effect, raising MAE to 5.17 kg/h, while static metadata, regime indicators, and rolling statistics change the score only modestly. The most important

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

negative result is the no-physics row: MAE falls to 4.36 kg/h when the simple prior is removed, compared with 5.08 kg/h for the full hybrid. In-domain NOx accuracy therefore does not benefit from this particular physical prior.

Figure 4 reports the transfer experiment. With no labelled data from the target units, PHASE-lite still reaches $R^2 = 0.965$ and MAE = 5.80 kg/h in plant-holdout evaluation. Adding one week of local observations reduces MAE to 5.59 kg/h, and three months of adaptation raises R^2 to 0.972. These are strong cross-unit results within the analysed U.S. reporting environment. They should not, however, be read as evidence of international transfer, because both source and held-out plants come from the same national data and regulatory system.

Table 4. Feature-group ablation for NOx (PHASE-lite). Recent history is the dominant feature group; the simple physical prior does not improve in-domain MAE.

Ablation (NOx)	MAE	RMSE	peak F1
no_lags	13.50	37.56	0.898
no_operating	5.17	18.76	0.963
full_hybrid	5.08	18.74	0.965
no_static	5.07	19.50	0.964
no_regime	5.06	18.73	0.965
no_rolling	5.02	18.44	0.966
no_physics	4.36	18.61	0.968

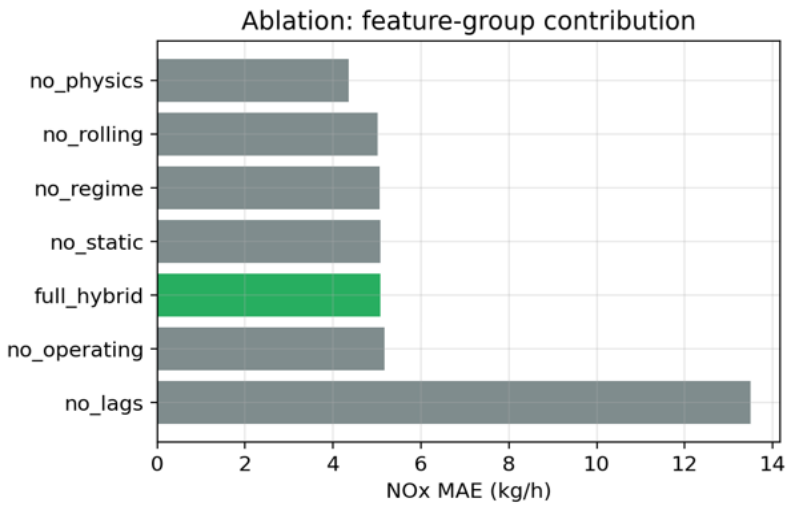


Fig. 3. NOx feature-group ablation; larger MAE increases indicate more influential feature groups.

The aggregate scores hide substantial variation by operating condition. As shown in Fig. 5, ramping hours have MAE = 12.49 kg/h, compared with 4.05 kg/h at baseload, and coal units are more difficult than gas units (13.84 versus 2.99 kg/h). The masking experiment in Fig. 6 points to the same dependence on recent state information. Removing recent emission history at test time drives MAE from 5.08 to 46.52 kg/h, while masking load and heat-input variables raises it to 20.33 kg/h. These results identify the channels that a practical CEMS-assisted deployment must preserve

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

reliably.

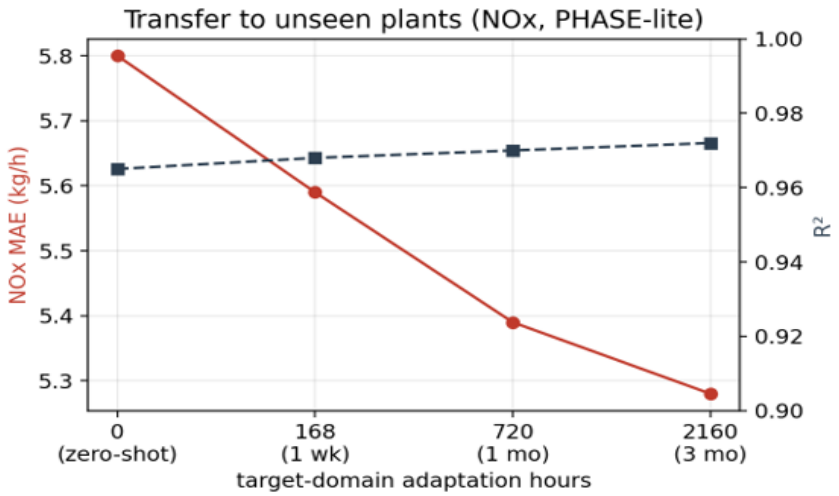


Fig. 4. Transfer to unseen plants: zero-shot performance and the effect of adding local labelled data.

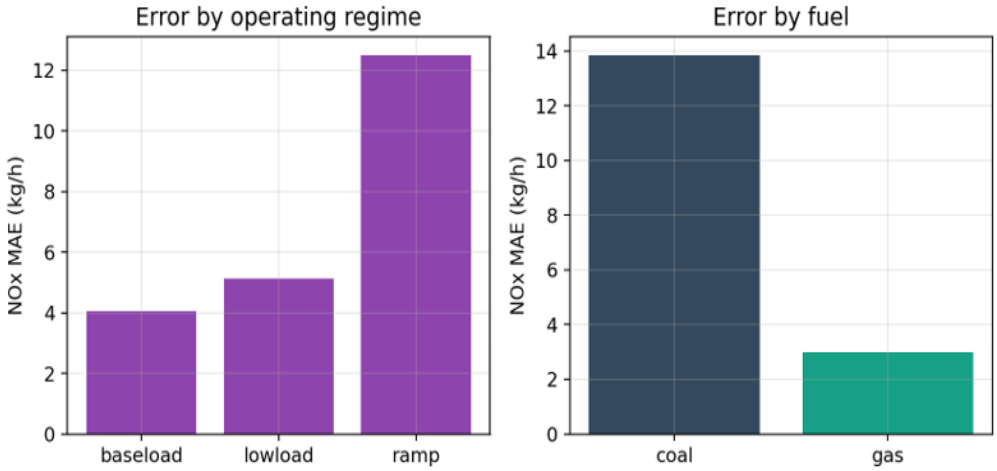


Fig. 5. NOx MAE by operating regime (left) and primary fuel group (right).

The statistical comparison clarifies why no single statement such as 'PHASE is better than XGBoost' is supported by the data. Across 509 test units, the paired bootstrap interval for MAE difference, defined as XGBoost minus PHASE-lite, is [-0.708, -0.684] kg/h; the entire interval is negative, so XGBoost has the lower average absolute error. The Wilcoxon signed-rank test confirms that this direction is consistent across units ($p < 10^{-46}$). The squared-error Diebold-Mariano test points in the opposite direction ($DM = 3.39$, $p = 6.9 \times 10^{-4}$) and favours PHASE-lite. Because squared error weights large misses more heavily, the two results are compatible: XGBoost is better on average MAE, while PHASE-lite avoids some high-magnitude errors. We therefore treat the hybrid as a robustness-oriented alternative rather than a uniformly more accurate model.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Discussion.

Limitations

Taken together, the experiments argue against using one modelling recipe for all three pollutants. CO_2 is the clearest case. In Table 1, a scaled emission factor alone explains almost all observed variability ($R^2 = 0.992$), which makes the physical baseline both accurate and easy to audit. XGBoost reduces the remaining error further, but the practical benefit of that reduction has to be weighed against the additional model complexity. For a fuel-bound pollutant such as CO_2, transparency can therefore be retained without sacrificing much predictive performance.

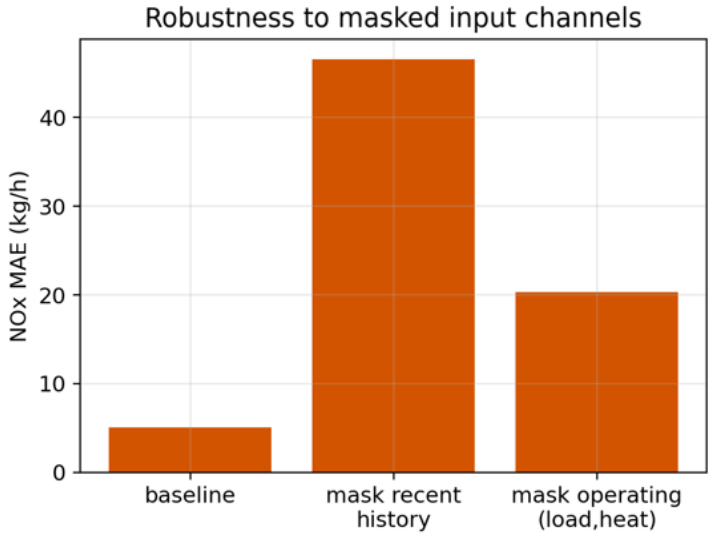


Fig. 6. Sensitivity to masked input channels during NOx inference.

NOx and SO_2 behave differently. Their factor baselines explain less than half of the observed variance, while learned models recover most of it across the tested splits. The reason is consistent with the feature and regime analyses: hourly values change with combustion conditions, load transitions, fuel properties, and control-system operation. The larger NOx errors during ramping and the strong dependence on recent history show that transient state matters. A fleet-wide constant cannot encode that state, so data-driven models are the more appropriate starting point for these pollutants.

The lag ablation also sets an important boundary on how the strongest model can be deployed. Historical features clearly improve NOx estimation, but part of that history consists of recent measured emissions. That makes the best-performing configuration CEMS-assisted. It is suitable for cross-checking a live monitor, detecting anomalous readings, and reconstructing short gaps; it is not a fully independent replacement after a long outage, when measured-emission lags are no longer available. For deployment studies, assisted and independent configurations should therefore be reported separately rather than combined under one PEMS accuracy figure.

The physical prior for NOx produced another useful negative result. Removing it improved average MAE, so with roughly 10 million labelled unit-hours the gradient-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

boosted learner did not need this simple prior to fit the in-domain relationship. At the same time, the full hybrid performed better under the squared-error Diebold-Mariano comparison and delivered calibrated prediction intervals. The contribution of PHASE is therefore not a universal reduction of point error. In this experiment its stronger case is uncertainty-aware prediction and reduced exposure to some large-error episodes.

Plant-holdout performance is encouraging but should be interpreted narrowly. PHASE-lite retains high accuracy on units that were absent from source training, which demonstrates cross-unit transfer inside this dataset. It does not show that the same accuracy will hold in another country, fuel system, or regulatory environment. Nor can the plant-holdout score by itself prove that the physical component caused the transfer benefit. A transfer-specific comparison of the full hybrid with an otherwise identical no-physics model is required before that effect can be attributed to the prior.

The benchmark has several limitations. First, it covers three U.S. states and two calendar years. The blocked-time, seasonal-holdout, and plant-holdout experiments test several forms of distribution shift inside that frame, but they do not cover differences in fuels, boiler technologies, control practices, maintenance conditions, or regulation that may appear in other power systems.

Second, the public data do not contain every variable that would be desirable for combustion and control modelling. Stack oxygen concentration, flue-gas temperature, detailed fuel composition, sulfur content, reagent consumption, and time-resolved control-device status are missing or unavailable with sufficient coverage. This matters most for NOx and SO_2, where those variables are directly connected to formation and removal mechanisms. Access to richer process measurements could improve both prediction and the physical relevance of the prior.

Third, the NOx and SO_2 priors used here are intentionally simple. A more detailed physical term could include boiler type, fuel sulfur, combustion-temperature proxies, excess-air indicators, and control-device efficiency. Such a prior may behave differently from the one evaluated in Table 4, so the present result should be read as evidence against this simple formulation, not against every possible physics-informed design.

Fourth, some control-technology attributes are available only at plant level and are transferred to generating units through the EPA-EIA crosswalk. At multi-unit plants with different equipment configurations, that mapping can introduce assignment uncertainty and weaken the meaning of a nominally unit-level control indicator.

Fifth, the strongest configurations depend on lagged measured emissions. That dependence is acceptable for a CEMS-assisted soft sensor, but it limits fully autonomous use. A separate CEMS-independent evaluation is necessary before recommending the model for long monitoring failures or for plants that have no historical continuous-emission record.

Finally, the study has no supervised external test in Kazakhstan because comparable public hourly CEMS labels were not available. Sentinel-5P TROPOMI NO_2 observations can provide a future corroborative comparison over major thermal plants, but they cannot replace stack-level labels. Satellite columns are affected

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

by atmospheric transport, chemistry, cloud screening, revisit timing, and spatial aggregation, so any Kazakhstan analysis must be presented as supporting evidence rather than direct validation of hourly stack emissions.

The practical conclusion is therefore conditional rather than universal. For CO_2, an emission-factor model is already a strong choice when auditability matters. For NOx and SO_2, learned representations of changing plant operation are necessary for high hourly accuracy. A physics-informed hybrid becomes most relevant when calibrated uncertainty, protection against large errors, or transfer with limited local labels is more valuable than the smallest average in-domain MAE.

Conclusion.

This work establishes a leakage-controlled benchmark for hourly CO_2, NOx, and SO_2 estimation using 10,144,680 unit-hours from 583 generating units and compares physical, statistical, machine-learning, and physics-informed approaches under blocked-time, seasonal-holdout, and plant-holdout evaluation. The main result is not that one model wins everywhere. The pollutant determines the modelling requirement. A scaled emission factor is already highly accurate for CO_2, while NOx and SO_2 need models that learn dynamic operating state and temporal history. For NOx, the simple PHASE prior does not improve average in-domain MAE over a tuned XGBoost baseline, but the hybrid provides calibrated intervals, reduces some large squared-error events, and retains strong accuracy on unseen units. PHASE should therefore be viewed as an uncertainty- and robustness-oriented framework rather than a universally more accurate black-box replacement. The current plant-holdout experiment demonstrates transfer within the analysed U.S. dataset only. The next evaluation steps are to separate CEMS-assisted from fully independent configurations, test stronger pollutant-specific priors, incorporate richer combustion and control variables, and validate on independently labelled non-U.S. data. The planned Sentinel-5P TROPOMI comparison over Kazakhstan thermal plants is a corroborative deployment-oriented step, not a substitute for stack-level validation.

REFERENCES

Adams, D., Oh, D.-H., Kim, D.-W., Lee, C.-H., & Oh, M. (2020). Prediction of SOx–NOx emission from a coal-fired CFB power plant with machine learning: Plant data learned by deep neural network and least square support vector machine // Journal of Cleaner Production. — Vol. 270. — Article 122310. DOI: https://doi.org/10.1016/j.jclepro.2020.122310. [In Eng.].

Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv. DOI: https://doi.org/10.48550/arXiv.1803.01271. [In Eng.].

Breiman, L. (2001). Random forests // Machine Learning. — Vol. 45. — No. 1. — Pp. 5–32. DOI: https://doi.org/10.1023/A:1010933404324. [In Eng.].

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. — Pp. 785–794. Association for Computing Machinery. DOI: https://doi.org/10.1145/2939672.2939785. [In Eng.].

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy // Journal of Business & Economic Statistics. — Vol. 13. — No. 3. — Pp. 253–263. DOI: https://doi.org/10.1080/07350015.1995.10524599. [In Eng.].

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory // Neural Computation. — Vol. 9. — No. 8. — Pp. 1735–1780. DOI: https://doi.org/10.1162/neco.1997.9.8.1735. [In Eng.].

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning // Nature Reviews Physics. — Vol. 3. — Pp. 422–440. DOI: https://doi.org/10.1038/

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

s42254-021-00314-5. [In Eng.].

Koenker, R., & Bassett, G., Jr. (1978). Regression quantiles // Econometrica. — Vol. 46. — No. 1. — Pp. 33–50. DOI: https://doi.org/10.2307/1913643. [In Eng.].

Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles // In Advances in Neural Information Processing Systems. — Vol. 30. — Pp. 6402–6413. [In Eng.].

Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting // International Journal of Forecasting. — Vol. 37. — No. 4. Pp. 1748–1764. DOI: https://doi.org/10.1016/j.ijforecast.2021.03.012. [In Eng.].

Lv, Y., Liu, J., Yang, T., & Zeng, D. (2013). A novel least squares support vector machine ensemble model for NOx emission prediction of a coal-fired boiler // Energy. — Vol. 55. — Pp. 319–329. DOI: https://doi.org/10.1016/j.energy.2013.02.062. [In Eng.].

Tan, P., He, B., Zhang, C., Rao, D., Li, S., Fang, Q., & Chen, G. (2019). Dynamic modeling of NOx emission in a 660 MW coal-fired boiler with long short-term memory // Energy. — Vol. 176. — Pp. 429–436. DOI: https://doi.org/10.1016/j.energy.2019.04.020. [In Eng.].

U.S. Environmental Protection Agency. (n.d.). Clean Air Markets Program Data: Bulk data files. Retrieved August 3, 2026, from DOI: https://campd.epa.gov/data/bulk-data-files. [In Eng.].

U.S. Environmental Protection Agency. (2026b). EPA–EIA Power Sector Data Crosswalk. U.S. Environmental Protection Agency. [In Eng.].

U.S. Energy Information Administration. (n.d.). Form EIA-860 detailed data with previous form data. Retrieved August 3, 2026, from https://www.eia.gov/electricity/data/eia860/. [In Eng.].

Vovk, V., Gammerman, A., & Shafer, G. (2005). Algorithmic learning in a random world. Springer. DOI: https://doi.org/10.1007/b106715. [In Eng.].

Zhou, H., Cen, K., & Fan, J. (2004). Modeling and optimization of the NOx emission characteristics of a tangentially fired boiler with artificial neural networks // Energy. — Vol. 29. — No. 1. — Pp. 167–183. DOI: https://doi.org/10.1016/j.energy.2003.08.004. [In Eng.].

Zhu, B., Ren, S., Weng, Q., & Si, F. (2024). A physics-informed neural network that considers monotonic relationships for predicting NOx emissions from coal-fired boilers // Fuel. — Vol. 364. — Article 131026. DOI: https://doi.org/10.1016/j.fuel.2024.131026. [In Eng.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 160–179
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.009

ADAPTIVE OUTPUT ROUTER FOR NATURAL LANGUAGE MODELS CONSIDERING KEY PERFORMANCE INDICATORS

A.M. Akhmetova[1], D.O. Baizhanova[1], Zh.N. Alisheva[2], B.A. Shaimov[1]*
[1]Al-Farabi Kazakh National University, Almaty, Kazakhstan;
[2]Kazakh National Agrarian Research University, Almaty, Kazakhstan.
E-mail: bekashaymovartukuli@gmail.com

Ardak Akhmetova — PhD, acting Associate Professor of the Department of Artificial Intelligence and Big Data, Al-Farabi Kazakh National University, Almaty, Kazakhstan https://orcid.org/0009-0000-2605-7206;
Dina Baizhanova — senior lecturer, departament "Artificial intelligence" Faculty of "Information technology and artificial intelligence", Al-Farabi Kazakh National University, Almaty, Kazakhstan
https://orcid.org/0000-0002-6109-8174;
Zh.N. Alisheva — PhD, Associate Professor, Head of the IT Department, Kazakh National Agrarian Research University, Almaty, Kazakhstan
https://orcid.org/0000-0003-0929-4984;
Beknur Shaimov — departament "Artificial intelligence" Faculty of "Information technology and artificial intelligence", Al-Farabi Kazakh National University, Almaty, Kazakhstan
E-mail: bekashaymovartukuli@gmail.com, https://orcid.org/0009-0006-2199-9628.

© A.M. Akhmetova, D.O. Baizhanova, Zh.N. Alisheva, B.A. Shaimov

Abstract. This paper proposes a KPI-oriented cascade router for multi-class text classification that sequentially employs lightweight, medium-sized, and heavyweight models depending on the confidence level of the prediction. The relevance of this study stems from the need to jointly manage classification accuracy and computational costs in applied natural language processing (NLP) systems, where relying exclusively on the most accurate models is often inefficient in terms of response time and inference cost. The experiments were conducted on the WOS-11967 corpus, which contains 11,967 scientific texts. To ensure proper parameter tuning and independent evaluation, a two-stage protocol with separate VAL and FINAL subsets was employed. The lightweight, medium-sized, and heavyweight models differ in their text representation methods and computational complexity, while routing between them is based on confidence thresholds. For the lightweight and medium-sized models, confidence was defined

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

as the difference between the two highest decision function values. The thresholds were selected on the VAL subset while accounting for the computational budget constraint governed by parameter α. The results demonstrate that the proposed cascade can accelerate inference compared with the heavyweight model, with a moderate reduction in accuracy, while maintaining a controlled trade-off between performance and computational cost. The practical significance of this work lies in the potential application of the proposed approach in systems where balancing accuracy, speed, and resource constraints is essential.

Keywords: KPI-oriented cascade router, text classification models, natural language processing (NLP), computational budget, cascade model

For citation: Akhmetova A.M., Baizhanova D.O., Alisheva Zh.N., Shaimov B.A. (2026). Adaptive output router for natural language models considering key performance indicators // International journal of information and communication technologies. Vol. 7. No. 27. Pp. 160–179. https://doi.org/10.54309/IJICT.2026.27.3.009. (In Russ.)

Conflict of interest: The authors declare that there is no conflict of interest.

КРІ КӨРСЕТКІШТЕРІН ЕСКЕРЕ ОТЫРЫП, ТАБИҒИ ТІЛ МОДЕЛЬДЕРІНЕ АРНАЛҒАН ШЫҒЫС МАРШРУТТАУЫНЫҢ АДАПТИВТІ ЖҮЙЕСІ

А.М. Ахметова[1], Д.О. Байжанова[1], Ж.Н. Алишева[2], Б.А. Шаймов[1]*

[1]Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан;
[2]Қазақ ұлттық аграрлық зерттеу университеті, Алматы, Қазақстан.
E-mail: bekashaymovartukuli@gmail.com

Ахметова Ардақ — PhD, Жасанды интеллект және үлкен деректер кафедрасының қауымдастырылған профессоры міндетін атқарушы, Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан
https://orcid.org/0009-0000-2605-7206;
Байжанова Дина — Ақпараттық технологиялар және жасанды интеллект факультетінің Жасанды интеллект кафедрасының аға оқытушысы, Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан
https://orcid.org/0000-0002-6109-8174;
Әлішева Ж. Н. — PhD, қауымдастырылған профессор, Ақпараттық технологиялар кафедрасының меңгерушісі, Қазақ ұлттық аграрлық зерттеу университеті, Алматы, Қазақстан
https://orcid.org/0000-0003-0929-4984;
Шаимов Бекнұр — Ақпараттық технологиялар және жасанды интеллект факультетінің Жасанды интеллект кафедрасы, Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан
E-mail: bekashaymovartukuli@gmail.com, https://orcid.org/0009-0006-2199-9628.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

© А.М. Ахметова, Д.О. Байжанова, Ж.Н. Алишева, Б.А. Шаймов

Аннотация. Мақалада мәтіндерді көпкласты жіктеу міндетіне арналған, болжам сенімділігінің деңгейіне қарай жеңіл, орташа және күрделі модельдерді ретімен қолданатын KPI-ге бағдарланған каскадты маршрутизатор ұсынылған. Зерттеудің өзектілігі қолданбалы табиғи тілді өңдеу (NLP) жүйелерінде жіктеу дәлдігі мен есептеу шығындарын өзара үйлесімді басқару қажеттілігімен айқындалады. Мұндай жүйелерде тек ең дәл модельдерді пайдалану жауап беру уақыты мен инференс құны тұрғысынан көбіне тиімсіз болып табылады. Эксперименттер құрамында 11 967 ғылыми мәтін бар WOS-11967 корпусында жүргізілді. Параметрлерді дұрыс баптау және тәуелсіз бағалау үшін VAL және FINAL ішкі жиындары бөлек пайдаланылатын екі кезеңді хаттама қолданылды. Жеңіл, орташа және күрделі модельдер мәтінді ұсыну тәсілі мен есептеу күрделілігі бойынша ерекшеленеді, ал олардың арасындағы маршруттау сенімділік шектеріне негізделеді. Жеңіл және орташа модельдер үшін сенімділік шешім функциясының ең жоғары екі мәнінің айырмасы ретінде анықталды. Шекті мәндер бюджет шектеуін ескере отырып, α параметрі арқылы VAL жиынында таңдалды. Алынған нәтижелер ұсынылған каскадтың күрделі модельмен салыстырғанда инференсті жеделдетуге мүмкіндік беретінін және дәлдіктің қалыпты төмендеуі жағдайында сапа мен есептеу құны арасындағы басқарылатын тепе-теңдікті сақтайтынын көрсетеді. Зерттеудің практикалық маңыздылығы ұсынылған тәсілді дәлдік, жылдамдық және ресурстық шектеулер арасындағы теңгерім маңызды болып табылатын жүйелерде қолдану мүмкіндігімен байланысты.

Түйінді сөздер: KPI-ге бағдарланған каскадты маршрутизатор, мәтінді жіктеу модельдері, табиғи тілді өңдеу (NLP), есептеу бюджеті, каскадты модель

Дәйексөздер үшін: Ахметова А.М., Байжанова Д.О., Алишева Ж.Н., Шаймов Б.А. (2026). KPI көрсеткіштерін ескере отырып, табиғи тіл модельдеріне арналған шығыс маршруттауының адаптивті жүйесі // Халықаралық ақпараттық және коммуникалық технологиялар журналы. Т. 7. No 27. Б. 160–179. https://doi.org/10.54309/IJICT.2026.27.3.009. (Орыс. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

АДАПТИВНЫЙ МАРШРУТИЗАТОР ВЫВОДА С УЧЕТОМ КЛЮЧЕВЫХ ПОКАЗАТЕЛЕЙ ЭФФЕКТИВНОСТИ ДЛЯ МОДЕЛЕЙ ЕСТЕСТВЕННОГО ЯЗЫКА

А.М. Ахметова[1], Д.О. Байжанова[1], Ж.Н. Алишева[2], Б.А. Шаймов[1]*

[1]Казахский национальный университет им. аль-Фараби, Алматы, Казахстан;
[2]Казахский национальный аграрный исследовательский университет, Алматы, Казахстан.
E-mail: bekashaymovartukuli@gmail.com

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Ахметова Ардак — PhD, исполняющая обязанности ассоциированного профессора кафедры искусственного интеллекта и больших данных, Казахский национальный университет имени аль-Фараби, Алматы, Казахстан
https://orcid.org/0009-0000-2605-7206;
Байжанова Дина — старший преподаватель кафедры искусственного интеллекта факультета информационных технологий и искусственного интеллекта,
https://orcid.org/0000-0002-6109-8174;
Алишева Ж. Н. — PhD, ассоциированный профессор, заведующая кафедрой информационных технологий, Казахский национальный аграрный исследовательский университет, Алматы, Казахстан;
https://orcid.org/0000-0003-0929-4984;
Шаимов Бекнур — кафедра искусственного интеллекта факультета информационных технологий и искусственного интеллекта, Казахский национальный университет имени аль-Фараби, Алматы, Казахстан
E-mail: bekashaymovartukuli@gmail.com, https://orcid.org/0009-0006-2199-9628.

© А.М. Ахметова, Д.О. Байжанова, Ж.Н. Алишева, Б.А. Шаймов

Аннотация. В статье предложен KPI-ориентированный каскадный маршрутизатор для задачи многоклассовой классификации текстов, последовательно использующий лёгкую, среднюю и тяжёлую модели в зависимости от уровня уверенности предсказания. Актуальность работы определяется необходимостью согласованного управления точностью классификации и вычислительными затратами в прикладных NLP-системах, где применение только наиболее точных моделей часто оказывается неэффективным по времени ответа и стоимости инференса. Эксперименты выполнены на корпусе WOS-11967, содержащем 11 967 научных текстов. Для корректной настройки и независимой оценки использован двухэтапный протокол с раздельными подвыборками VAL и FINAL. Лёгкая, средняя и тяжёлая модели различаются способом представления текста и вычислительной сложностью, а маршрутизация между ними осуществляется на основе порогов уверенности. Для лёгкой и средней моделей уверенность определялась как разность между двумя наибольшими значениями функции решения. Подбор порогов выполнялся на VAL с учётом бюджетного ограничения, задаваемого параметром α. Полученные результаты показывают, что предложенный каскад позволяет ускорить инференс по сравнению с тяжёлой моделью при умеренном снижении точности, сохраняя управляемый компромисс между качеством и вычислительной стоимостью. Практическая значимость работы состоит в возможности применения предложенной схемы в системах, где важен баланс между точностью, скоростью и ресурсными ограничениями.

Ключевые слова: KPI-ориентированный каскадный маршрутизатор, модели текстовой классификации, NLP, вычислительный бюджет, каскадная модель, kpi

Для цитирования: Ахметова А.М., Байжанова Д.О., Алишева Ж.Н.,

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Шаймов Б.А. (2026). Адаптивный маршрутизатор вывода с учетом ключевых показателей эффективности для моделей естественного языка // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 27. Стр. 160–179. https://doi.org/10.54309/IJICT.2026.27.3.009. (На русс.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Введение.

Методы обработки естественного языка эволюционируют быстро. Особенно это касается алгоритмов автоматической классификации текстов. Однако рост точности неизбежно влечет за собой увеличение вычислительных затрат. Ресурсоемкость растет. На практике развертывание тяжеловесных архитектур часто оказывается экономически нерентабельным. Именно поэтому исследователи обращаются к адаптивным маршрутизаторам вывода и каскадным моделям (Taha et al., 2025).

Логика подобной схемы достаточно прозрачна. В единый конвейер объединяются несколько классификаторов разной степени сложности. Они функционируют строго последовательно. При этом легкие алгоритмы быстро обрабатывают тривиальные примеры. Сложные тексты, требующие глубокого анализа, передаются более точным, но вычислительно дорогим сетям.

Мы исследуем задачу текстовой классификации именно через призму каскадного подхода. Полученные эмпирические результаты детально сопоставляются с показателями изолированных моделей. Нас интересует строгая оценка эффективности многоуровневой архитектуры. Требовалось определить рациональную стратегию маршрутизации между тремя уровнями классификаторов. В качестве экспериментальной базы задействован набор данных WOS-11967. Нами были спроектированы три базовые конфигурации: легкая, средняя и тяжелая. Разработанный каскад интегрирует их воедино.

Стоит отметить, что активация каждого последующего этапа жестко контролируется. Чтобы оптимизировать компромисс между итоговой точностью и вычислительной стоимостью, мы ввели пороговые значения уверенности, а также бюджетный параметр α. Как именно эти метрики смещают баланс системы? Текст подробно описывает механизмы реализации базовых компонентов и общую структуру самого каскада. Вычислены предельные значения точности. Зафиксированы аппаратные потолки стоимости для рассматриваемых конфигураций. На практике итоговые тесты многоуровневой архитектуры проводились при широком варьировании пороговых полей и значений параметра α.

Целью исследования является разработка и оценка каскадного маршрутизатора для многоклассовой классификации текстов, обеспечивающего управляемый компромисс между точностью и вычислительной стоимостью. Для достижения поставленной цели решаются следующие задачи: построение трёх базовых моделей различной вычислительной сложности, определение меры уверенности для маршрутизации между уровнями каскада, подбор порогов на валидационной выборке с учётом бюджетного ограничения α и независимая оценка итоговой

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

конфигурации на выборке FINAL.

Литературный обзор

Сектор интеллектуальных систем сегодня следует вектору на усложнение алгоритмических структур. Данная тенденция обусловлена ростом требований к качеству обработки данных. Однако по мере ужесточения требований к точности и глубине обработки данных вычислительная стоимость моделей превращается в серьезный барьер для их широкого внедрения. Ресурсоемкость растет. При этом современные исследования в области классификации данных подтверждают: наличие высокоточного решения еще не гарантирует его успеха в эксплуатации. Зачастую интеграция мощных архитектур в реальные производственные циклы ограничена критическими требованиями к задержке и чрезмерными затратами на аппаратную базу (Taha et al., 2025).

Мировой научный вектор смещается. Все чаще исследователи предпочитают адаптивные архитектуры традиционным монолитным моделям. Это закономерный переход. Одной из наиболее перспективных стратегий в данном направлении считается применение каскадных схем и механизмов раннего выхода. Они позволяют управлять нагрузкой в динамике. Простые запросы эффективно обрабатываются легковесными алгоритмами. При этом значительные мощности резервируются исключительно для наиболее трудоемких случаев. Однако проектирование таких систем сопряжено с поиском строгого математического баланса между итоговой точностью и стоимостью вычислений. Требуются надежные рычаги управления. На практике оператору необходимы инструменты для корректировки этого компромисса через измеримые переменные - бюджетный параметр α и пороги достоверности. Мы сосредоточились на проектировании каскадной схемы, ориентированной на максимизацию ключевых показателей эффективности. Базовой платформой для тестирования стала задача текстовой классификации. Стоит отметить, что предложенная концепция управления бюджетом универсальна. Она может быть успешно интегрирована в самые разные домены машинного обучения.

Систематизация подходов к классификации остается приоритетным направлением в области интеллектуальной обработки данных. Исследование Аллама и др. предлагает детальный обзор трансформации этой сферы - от классических методов до глубокого обучения (Allam et al., 2025). Авторы справедливо указывают на фундаментальную проблему. Дилемма между вычислительной нагрузкой и точностью прогнозов на реальных данных. Это представляет существенную проблему для практического применения моделей. Наша работа развивает идеи Аллама и соавторов в части оптимизации масштабируемости процессов. Однако стоит отметить одну деталь. Большинство существующих подходов трактуют классификаторы как изолированные инструменты. В настоящей работе рассматривается гипотеза о том, что достижение заданных KPI требует перехода от выбора единой модели к многоуровневой архитектуре. Приоритет - многоуровневые системы. Использование каскадного метода позволяет объединить достоинства разномасштабных архитектур в рамках единого решения. Фактически,

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

обзор Аллама и др. служит теоретическим фундаментом для внедрения адаптивных структур с динамически управляемым бюджетом вычислений (Allam et al., 2025).

Работа Джалиля и др. подтверждает выбор мультиномиального наивного Байеса и линейной SVM (метода опорных векторов) в качестве надежных базовых компонентов, однако авторы ограничиваются поиском «идеального» единого классификатора. Анализ этой статьи позволяет сделать вывод, что потенциал этих алгоритмов более полно раскрывается в составе иерархических структур, а не при использовании их изолированно. Это послужило обоснованием для их включения в предлагаемый каскад в качестве лёгкого и среднего уровней, что расширяет границы эффективности отдельных методов за счет их интеграции в адаптивный маршрутизатор с управляемым α-бюджетом (Jaleel et al., 2022).

Разработка методов классификации данных, описанных в глобальных обзорах Allam и др. и Taha и др., подтверждает переход от классических алгоритмов к алгоритмам нейронных сетей, но подчеркивает острую проблему их масштабируемости и вычислительных затрат. Исследование Jaleel и др. проверяет эффективность многочленных методов наивного Байеса и SVM для текстовых данных, а работа Граффа и др. доказывает, что простые подходы остаются конкурентоспособными и экономически эффективными даже в эпоху трансформаторов. Анализ этих работ позволяет сделать вывод, что потенциал этих моделей наиболее полно раскрывается не в их изолированном применении, а в рамках иерархических структур. Это стало обоснованием включения классических методов в предлагаемый каскад как «легкого», так и «среднего» уровней. Таким образом, если существующие исследования направлены на поиск «идеального» единого классификатора, то исследование расширяет границы существующих подходов, интегрируя различные парадигмы в адаптивный маршрутизатор с управляемым α-бюджетом для оптимизации целевых ключевых показателей эффективности (Allam et al., 2025; Jaleel et al., 2022; Graff et al., 2025).

Важным аспектом эффективности систем классификации является схема взвешивания признаков, которая была подробно изучена в работе Танга и др. Авторы предлагают улучшенную схему взвешивания контролируемых терминов, которая использует категориальную информацию для создания более точных векторных представлений. Их результаты доказывают, что оптимизация входных данных может значительно повысить точность даже без усложнения классификатора. Однако основным недостатком такого подхода в контексте высоконагруженных систем является его статичность: предлагаемая схема применяется ко всем входящим данным одинаково, без учета динамики вычислительных затрат и текущего бюджета ресурсов. В настоящем исследовании эта проблема решается путем интеграции идей Танга и др. в адаптивную структуру. В работе используется взвешивание качественных признаков в качестве основы для «простого» каскадного уровня. Это позволяет повысить надежность базовой модели, чтобы она могла самостоятельно обрабатывать большинство запросов. Таким образом, там, где метод Танга и др. просто повышает точность, предлагаемый алгоритм

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

использует эту точность для активации «раннего выхода», который при поддержке параметра α превращает статическое взвешивание в инструмент для активной экономии вычислительной мощности (Tang et al., 2022).

Теоретический фундамент интеллектуального анализа данных и методов классификации детально проработан в труде Заки и Мейры. Исследование фундаментально. Авторы предлагают исчерпывающий математический разбор алгоритмов, включая SVM и наивные байесовские классификаторы, с акцентом на метриках производительности. Стоит признать: понимание этих базовых механизмов обучения критически важно для проектирования отказоустойчивых систем. В работе использованы данные положения при систематизации критериев точности.При этом существует методологический нюанс. Большинство классических руководств рассматривают алгоритмы как изолированные математические абстракции. На практике они редко затрагивают вопросы интеграции разнородных моделей ради оптимизации ресурсов. В настоящей работе сохраняется теоретический базис авторов, но адаптируем его под нужды каскадных структур. Очевидно, что традиционные подходы к оценке целесообразно расширить через введение бюджетного параметра α. Это позволяет перевести статические алгоритмы в категорию элементов адаптивной прикладной системы (Zaki et al., 2014).

В работе Тан и др. предлагается метод взвешивания контролируемых терминов, который значительно повышает точность классификации за счет использования категориальной информации. Однако предлагаемый подход ограничен использованием фиксированных вычислительных затрат для всех типов данных. Изучение этой статьи привело меня к выводу, что потенциал такого точного взвешивания наиболее эффективен на уровне «легкого» каскада. Интеграция их метода в мою архитектуру позволила мне создать механизм «раннего выхода», в котором высокая точность базовой линии при поддержке параметра α сводит к минимуму доступ к моделям высокого уровня, превращая фиксированное улучшение функции в динамический инструмент экономии ресурсов (Responsible and Sustainable AI, 2025).

Работа Аль-Хелали и коллег сосредоточена на интеграции легковесных алгоритмов машинного обучения. Это актуально для классификации текстов в реальном времени. Основной упор сделан на устройства с дефицитом аппаратных мощностей. Ученые демонстрируют значительный потенциал оптимизированного многочленного наивного байесовского классификатора. При должной настройке он обеспечивает высокую производительность. При этом постоянный доступ к облачным вычислениям не требуется. На практике использование экономичных моделей в условиях нехватки ресурсов подтверждает свою целесообразность. Подобные выводы напрямую коррелируют с концепцией «базового» каскадного уровня. Однако большинство представленных решений ограничено рамками одного устройства или единственной модели. Изолированность снижает гибкость. Анализ работы привел к важному заключению. Для достижения пиковой эффективности недостаточно простого переноса алгоритма на конечное устройство. Требуется его интеграция в контур динамического управления. Сто-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ит отметить, что именно этот тезис позволил обосновать включение легких классификаторов в структуру предлагаемой трехуровневой архитектуры (Al-Helali et al., 2025).

Вне архитектурных рамок современные подходы зачастую демонстрируют дефицит операционной гибкости. Это критично. Их оптимизационные механизмы остаются статичными, что исключает быструю адаптацию к динамике вычислительного бюджета без повторного обучения. Наблюдается и методологический разрыв. Пропасть между оптимизацией классических алгоритмов и современными нейронными сетями по-прежнему велика. На практике предлагаемая гибридная система призвана нивелировать эти несоответствия. В работе управление ресурсами отделено от логики классификации от непосредственной логики классификации, внедрив внешний параметр α. Такой подход гарантирует прозрачность принимаемых решений. Система обретает способность к мгновенной подстройке под заданные ключевые показатели эффективности. Смена приоритетов происходит без деградации структуры (Zeng et al., 2024).

Многообещающий подход к управлению ресурсами представлен в работе Гамильтона и Фулпа, описывающих бюджетные последовательные классификаторы. Авторы используют эволюционные методы для создания систем, работающих в условиях жестких ограничений по времени и памяти, внедряя варианты раннего завершения для устранения избыточных вычислений. Данная работа подтверждает важность многокритериальной оптимизации, которая напрямую связана с задачами управления KPI. В отличие от сложного эволюционного поиска структур, предложенного авторами, наша архитектура предлагает более прозрачный механизм регулирования. Использование прямого внешнего управления через параметр α позволяет адаптировать принципы бюджетной классификации к мгновенным изменениям в операционной среде, сохраняя при этом простоту интерпретации и настройки системы (Hamilton et al., 2022).

Важным дополнением к архитектурным аспектам является проблема надежности оценки достоверности модели, исследованная в работе Холм и др. Авторы проводят сравнительный анализ использования стандартного Softmax и метода отсева по методу Монте-Карло для аппроксимации неопределенности в задачах классификации текстов. В исследовании подчеркивается, что выбор метода оценки достоверности напрямую влияет на баланс между вычислительными затратами и качеством прогнозов. Эти результаты подтверждают необходимость использования математически обоснованных пороговых значений для принятия решения о досрочном выходе. В отличие от Холм и др., которые сосредоточены на анализе внутренней неопределенности нейронных сетей, в работе используется концепция уверенности как связующее звено между различными уровнями каскада. Это позволяет использовать быстрые оценки уверенности на начальных этапах каскада, компенсируя их возможную неточность с помощью внешнего α-контроллера. Таким образом, предлагаемый метод превращает теоретические оценки достоверности в практический инструмент для динамического управления потоками данных в многоуровневой системе (Holm et al., 2023).

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Архитектура адаптивных систем с ранним доступом к компонентам подробно декомпозирована в исследовании Ласкаридиса и др. Авторы систематизируют ключевые вызовы индустрии. Рассматриваются различные аспекты: от конфигурации точек выхода до специфики обучения. Адаптивный логический вывод здесь позиционируется как фундаментальный метод преодоления ограничений вычислительной глубины современных нейросетей. Данный обзор подчеркивает актуальность нашего подхода. Он выявляет дефицит универсальных стратегий управления каскадами в реальном времени. При этом Ласкаридис и соавторы концентрируются преимущественно на внутренней методологии построения отдельных сетей с промежуточными выходами. Предложенная нами архитектура расширяет эту концепцию до уровня межмодельного взаимодействия. Это расширяет область применимости рассматриваемого подхода. Вместо манипуляций с глубиной единственной сети мы используем каскад принципиально разнородных алгоритмов. Внешнее регулирование через параметр α позволяет преодолеть проблему статического проектирования, на которую указывают авторы. На практике это позволяет сформировать концепцию маршрутизатора, ориентированного на целевые ключевые показатели эффективности. Такая структура гармонично сочетает адаптивность с экономическим обоснованием вычислительных затрат (Laskaridis et al., 2021).

Развитие концепции контролируемого выхода продолжено в работе Мотлаха и др., посвященной классификации с возможностью отказа. Авторы предлагают математическую основу для устранения зон «случайного угадывания» в пространстве решений нейронных сетей, используя пороговые значения, специфичные для класса Softmax. Этот подход направлен на максимизацию точности в выбранной области данных путем делегирования неопределенных случаев механизму сбоя. Этот механизм соответствует логике «раннего выхода» в предлагаемой системе, но существующие модели ROC чаще всего ограничены бинарным сценарием. Это подтверждает необходимость расширения этой логики до многоуровневого каскада. В то время как Motlagh и др. предлагают просто исключить предсказание, предлагаемый метод выводит запрос на более мощный вычислительный уровень. Таким образом, идеи об ограничении области уверенного принятия решений служат обоснованием для введения динамических пороговых значений, которые в данной работе масштабируются с помощью параметра α для управления всей иерархией моделей (Naturally constrained reject option classification, 2024).

Подводя итог обзору литературы, можно констатировать, что современная наука успешно решила задачи построения индивидуальных высокоточных моделей и сформулировала основные принципы адаптивного логического вывода. Тем не менее, остается заметный разрыв между теоретическими методами оптимизации и потребностями реальных систем в эксплуатационной гибкости. Большинство существующих решений остаются либо статичными, либо чрезмерно сложными для мгновенной перенастройки под текущие ресурсы. Исследование, представленное в этой статье, направлено на преодоление этого пробела, пред-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

лагая архитектуру, в которой управление бюджетом вычислений и точностью отделено от внутренней логики моделей и интегрировано в гибкий интерфейс управления ключевыми показателями эффективности с помощью внешнего параметра α.

Материалы и методы.

В работе исследовалась каскадная схема многоклассовой классификации текстов на корпусе WOS-11967, содержащем 11 967 текстовых документов научной тематики, распределённых по тематическим классам. Для формирования обучающей и тестовой выборок исходный набор данных был разделён стратифицированным образом в отношении 80:20. Далее тестовая часть дополнительно разбивалась на две равные подвыборки: VAL и FINAL. Подвыборка VAL использовалась для подбора параметров каскада, тогда как FINAL применялась исключительно для независимой итоговой оценки качества найденной конфигурации. Такой протокол позволяет снизить риск переобучения на этапе настройки порогов маршрутизации и обеспечивает более корректную оценку обобщающей способности модели (Kowsari et al., 2018).

Для эксперимента мы обучили три базовые модели, различающиеся по уровню вычислительной сложности: легкую (Light), среднюю (Medium) и тяжелую (High). Каждая из них ориентирована на решение идентичной задачи многоклассовой классификации. Основные различия между ними кроются в методах представления текстовых данных и архитектурной глубине самих классификаторов. На этапе предобработки все входные последовательности проходили через стандартные процедуры нормализации. Речь идет о приведении символов к нижнему регистру и очистке текста. При этом последующая трансформация в векторные формы выполнялась уже специфическими средствами соответствующих векторизаторов. Стоит отметить, что такой подход обеспечил единообразие исходных данных при сохранении уникальности внутренних механизмов каждой модели. Таким образом, иерархия сложности выстраивалась от простых структур к наиболее ресурсоемким.

Лёгкая модель (Light) была построена на основе представления Bag-of-Words с использованием только униграмм. Для ограничения размерности пространства признаков применялся словарь фиксированного размера (max_features = 1000), а также отсечение слишком редких и слишком частотных терминов (min_df = 5, max_df = 0.85). В качестве классификатора использовался мультиномиальный наивный байесовский классификатор с параметром сглаживания α = 5.0. Такая конфигурация ориентирована на минимальную вычислительную стоимость и быстрое получение предсказания, даже если это сопровождается снижением точности по сравнению с более сложными моделями (Kim et al., 2006).

Средняя модель (Medium) реализована как конвейер из трёх этапов: векторизация текста с помощью TF-IDF по словам с использованием 1–2-грамм, последующее понижение размерности методом Truncated SVD и классификация с помощью линейного метода опорных векторов. Для TF-IDF-представления использовались ограничения max_features = 2000, min_df = 4, max_df = 0.85, a

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

также сублинейное масштабирование частот терминов. После построения разреженного TF-IDF-пространства выполнялось сжатие признакового пространства до 96 компонент, что позволяет сократить вычислительные затраты и одновременно сохранить значимую часть дискриминативной информации. На заключительном этапе применялся LinearSVC с коэффициентом регуляризации C = 0.15. По сравнению с лёгкой моделью данная конфигурация обеспечивает более высокое качество классификации при умеренной вычислительной стоимости (Hastie et al., 2009).

Конфигурация тяжелой модели (High) нацелена на достижение предельной точности за счет расширенного представления текстовых данных. В основе лежит формирование гибридного признакового пространства. Оно включает TF-IDF векторы по словам (диапазон 1–2-грамм) и символьные TF-IDF признаки (3–5-грамм). При этом размерность словарей была зафиксирована на уровнях 50 000 для словных и 30 000 для символьных паттернов соответственно. После этапа конкатенации признаков итоговая классификация делегируется линейному SVM с коэффициентом регуляризации C = 2.0. Такой подход оправдан. Параллельный учет словных и символьных n-грамм позволяет захватывать как общую лексическую семантику, так и специфические локальные шаблоны. Это критически важно для обеспечения устойчивости классификатора при работе со сложными научными корпусами.

Три описанных модуля легли в основу иерархической каскадной архитектуры. Ключевой принцип системы - серийная обработка объекта. Текст поступает на вход легкого классификатора. Если уровень достоверности прогноза преодолевает установленный предел TL, процесс завершается, и результат признается финальным. Однако при низком доверии объект транслируется на средний уровень. Логика повторяется. Если уверенность средней модели оказывается выше порога TM, система выдает ее предсказание. В противном случае финальный вердикт выносит тяжелая модель. Подобная маршрутизация позволяет эффективно фильтровать потоки данных. Простые случаи закрываются дешевыми алгоритмами. Трудные же примеры автоматически делегируются более прецизионным, пусть и ресурсоемким, инструментам (Aggarwal et al., 2018).

Важным элементом предложенной схемы является способ оценки уверенности предсказания. Для этого использовалась margin-based confidence measure, вычисляемая как разность между двумя наибольшими значениями функции решения для рассматриваемого объекта. Чем больше эта разность, тем более уверенным считается предсказание модели. Такой подход применялся для лёгкой и средней моделей и использовался в качестве критерия маршрутизации внутри каскада.

Для количественной оценки вычислительной стоимости измерялось среднее время обработки одного объекта каждой моделью. Стоимость каскада вычислялась как ожидаемые суммарные затраты на последовательный проход через модели с учётом того, какая доля объектов завершается на каждом уровне. Если обозначить через p_L, p_M и p_H доли объектов, классифицированных соответствен-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

но на уровнях Light, Medium и High, а через $cost_L$, $cost_M$ и $cost_H$ - среднюю стоимость одного предсказания соответствующей модели, то средняя стоимость каскада определялась как:

$$cost_{cascade} = cost_L + (p_M + p_H) \cdot cost_M + p_H \cdot cost_H \quad (1)$$

Данная формула отражает последовательную природу каскада: каждый объект вначале проходит через лёгкую модель, часть объектов затем дополнительно обрабатывается средней моделью, а наиболее трудные экземпляры доходят до тяжёлой модели.

Подбор порогов t_L и t_M осуществлялся на подвыборке VAL методом полного перебора. В качестве кандидатных значений использовались пороги, полученные по квантилям распределения уверенности моделей. Для каждой пары порогов рассчитывались точность каскада, его средняя стоимость, а также доли объектов, классифицированных на каждом уровне. После этого вводилось бюджетное ограничение через параметр α, определяющий допустимую долю стоимости каскада относительно стоимости тяжёлой модели:

$$cost_{cascade} \leq \alpha \cdot cost_H \quad (2)$$

Параметр α задаёт компромисс между качеством и вычислительными затратами. При малых значениях α допускаются только очень дешёвые конфигурации, что, как правило, приводит к более частому использованию лёгкой модели и снижению точности. При увеличении α каскад получает возможность чаще обращаться к более сложным моделям, вследствие чего качество классификации возрастает, но увеличивается и вычислительная стоимость. В исследовании рассматривались значения α в диапазоне от 0.1 до 1.0, после чего среди всех допустимых по бюджету конфигураций выбиралась та, которая обеспечивала максимальную точность на VAL.

На заключительном этапе найденные на VAL оптимальные пороги переносились без изменений на независимую выборку FINAL, где вычислялись итоговые значения точности, средней стоимости и долей использования моделей разных уровней. Такой двухэтапный протокол оценки позволяет рассматривать полученные результаты как более надёжные и методологически корректные, поскольку параметры каскада подбираются и проверяются на непересекающихся данных.

Эксперименты

Перед анализом каскадной модели рассмотрим, какие результаты можно получить при различных пропорциях использования трёх базовых моделей. В отличие от каскадного подхода, где выбор модели определяется пороговыми значениями TL и TM, здесь выполняется общий анализ: для каждого примера одна из моделей выбирается случайным образом в соответствии с заданными вероят-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ностями:

$$w_L, w_M, w_H, \text{ где } w_L + w_M + w_H = 1. \quad (3)$$

Такой подход позволяет оценить верхнюю границу качества, достижимую смешанными стратегиями использования моделей, а также определить, насколько каскадный алгоритм приближается к этому уровню. Для каждой тройки процентных весов (α* L, M, H) проводится моделирование методом Монте-Карло (Sebastiani et al., 2002; Varma et al., 2006).

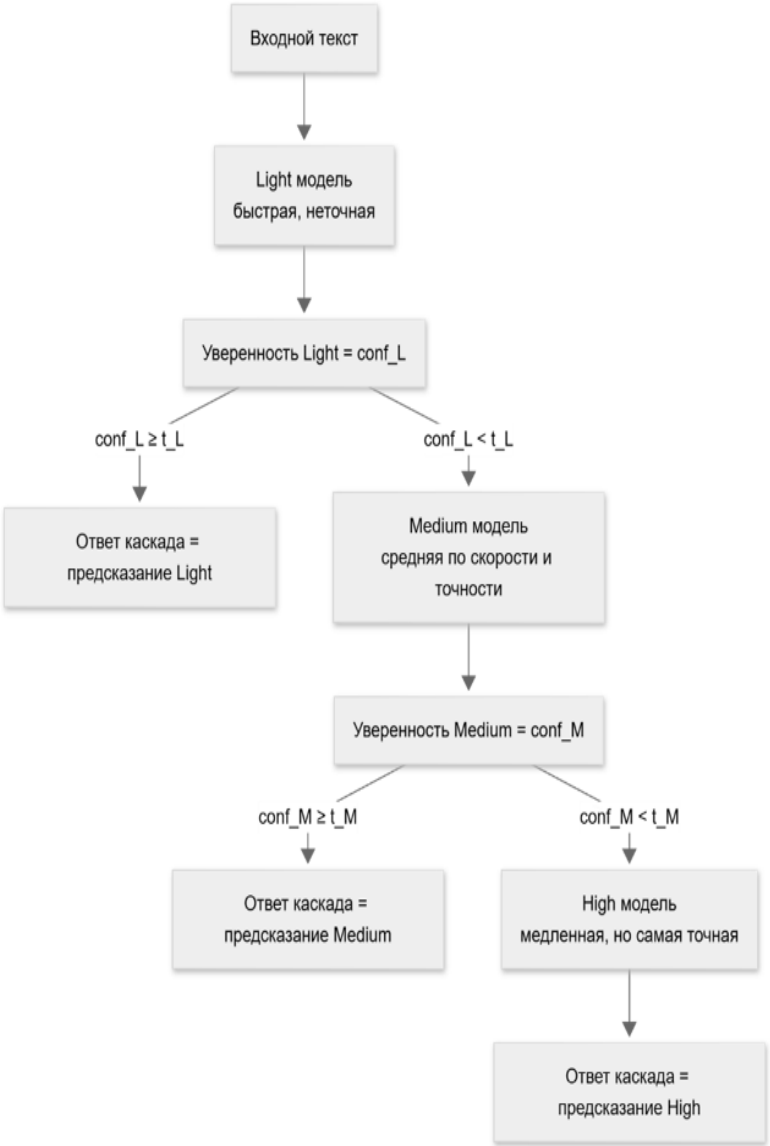


Рис. 1. Блок-схема каскадной модели.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Для каскадной модели выполняется полный перебор пороговых значений TL, TM = [0,100], в заданном диапазоне, что даёт 10 000 возможных комбинаций параметров. Полученные конфигурации фильтруются по бюджетному ограничению (α* cost сложной модели). Среди допустимых конфигураций выбирается вариант с максимальной точностью.

Результаты и обсуждение.

В эксперименте в качестве основных метрик использовались точность классификации и среднее время обработки одного примера (cost).

Таблица 1. Характеристики базовых моделей на валидационных данных: точность и среднее время обработки одного примера.

Модели	Точность модели	Среднее время обработки одного примера (вычислительной стоимости)
Лёгкая модель	0.7109	0.000116 сек
Средняя модель	0.8028	0.000227 сек
Сложная модель	0.8789	0.002323 сек

Как видно из Таблицы 1, наиболее тяжёлая модель обеспечивает наивысшую точность, но требует наибольших вычислительных затрат. Лёгкая модель демонстрирует точность около 0.7109, тогда как тяжёлая достигает 0.8789, однако среднее время обработки одного примера возрастает с 0.000116 с до 0.002323 с, то есть примерно в 20 раз. Средняя модель занимает промежуточное положение как по точности, так и по стоимости. Таким образом, уже на уровне базовых моделей наблюдается классический компромисс между качеством классификации и скоростью обработки (Voskergian et al., 2024).

Верхнюю границу достижимой точности можно оценить с помощью оракульного подхода, при котором для каждого примера засчитывается правильный ответ, если его корректно предсказала хотя бы одна из моделей. Такой эксперимент даёт точность около 0.8988 и задаёт теоретический максимум для рассматриваемого набора базовых моделей. Лучшие конфигурации каскада уступают этому уровню, поскольку маршрутизация основана только на оценке уверенности и не обладает знанием правильного ответа для каждого объекта. Следовательно, разрыв между оракульной и фактической точностью отражает предел улучшения, который в принципе может быть достигнут при более точной стратегии маршрутизации.

Результаты оценки каскадной модели при различных значениях параметра α на выборке VAL представлены в Таблице 2. Таблица показывает, как при изменении α меняются итоговая точность, средняя стоимость обработки и доли обращений к лёгкой, средней и тяжёлой моделям. По мере увеличения α каскад чаще использует более сложную модель; при меньших значениях α система в большей степени опирается на лёгкие уровни.

Таблица 2. Точность, средняя вычислительная стоимость и доли использования лёгкой, средней и тяжёлой моделей для оптимальных конфигураций

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

каскада на выборке VAL при различных значениях параметра α.

alpha	Accuracy	Cost (сек)	L доля	M доля	H доля
0.1	0.79616	0.0002	51.96%	47.11%	0.91%
0.2	0.84127	0.00042	49.03%	40.18%	10.77%
0.3	0.85464	0.00062	34.00%	46.61%	19.38%
0.4	0.8655	0.00082	14.03%	59.06%	26.90%
0.5	0.87051	0.001	26.98%	36.17%	36.84%
0.6	0.87385	0.00124	22.05%	30.32%	47.61%
0.7	0.87636	0.00141	26.98%	16.29%	56.72%
0.8	0.87886	0.00158	6.59%	30.82%	62.57%
0.9	0.87886	0.00158	6.30%	31.20%	62.50%

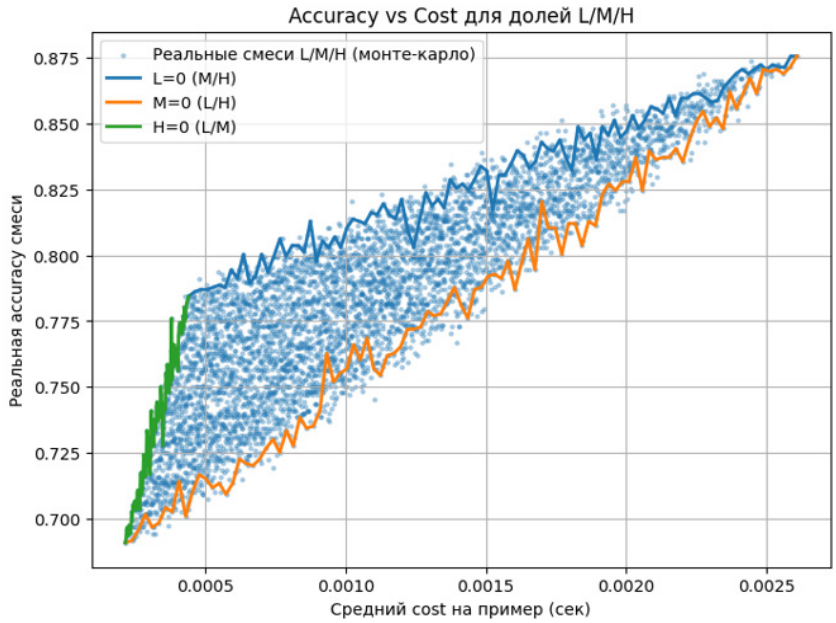


Рис. 2. На графике показана зависимость между средней стоимостью обработки одного примера и точностью для различных комбинаций вкладов лёгкой (L), средней (M) и тяжёлой (H) моделей. Синие точки получены методом Монте-Карло: для каждого текста случайным образом выбирается одна из моделей L/M/H с заданными вероятностями.

Следует отметить, что при больших значениях α различия между соседними конфигурациями могут исчезать или становиться минимальными. Это связано с тем, что после достижения некоторого уровня бюджетное ограничение перестаёт быть активным, и оптимальная конфигурация уже определяется дискретной сеткой порогов TL и TM. Поэтому дальнейшее увеличение α не обязательно приводит к изменению структуры каскада или росту точности.

Полученные результаты показывают, что каскадная модель может работать примерно в 2–5 раз быстрее тяжёлой модели при снижении точности всего на 1–5%. Это подтверждает основную идею каскадного подхода: простые примеры обрабатываются лёгкими моделями, а более сложные передаются на по-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

следующие уровни, что позволяет снизить средние вычислительные затраты при сохранении высокого качества классификации.

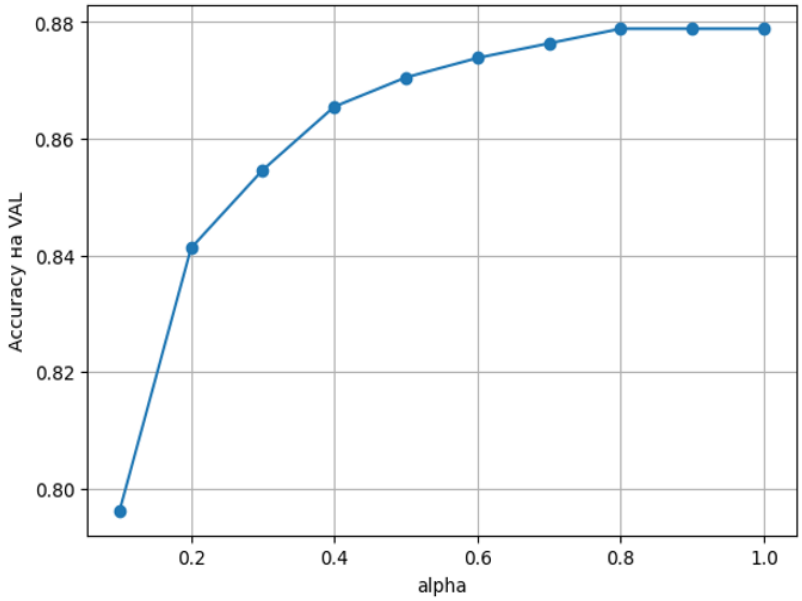


Рис. 3. График по точности лучшего каскада в зависимости от alpha

На следующем этапе найденные на выборке VAL оптимальные пороги T_L и T_M переносятся на выборку FINAL. Для каждого значения α каскад оценивается на FINAL-наборе, и фиксируются итоговые значения точности, стоимости и долей использования моделей. Далее для каждого значения α выбираются пороги, найденные на VAL, после чего соответствующая конфигурация оценивается на FINAL.

Таблица 3. Значения точности и стоимости модели, и доли для каждой модели на FINAL данных.

Alpha	Точность	Стоимость на один пример (сек)	L доля	M доля	H доля
0.1	0.7786	0.00023	49.20%	49.50%	1.30%
0.2	0.8145	0.00047	46.80%	41.10%	12.10%
0.3	0.8413	0.00066	33.40%	47.20%	19.40%
0.4	0.8555	0.00086	15.60%	57.30%	27.10%
0.5	0.863	0.00108	28.20%	34.10%	37.80%
0.6	0.873	0.0013	23.60%	28.80%	47.60%
0.7	0.8713	0.00149	28.20%	15.20%	56.60%
0.8	0.8772	0.00167	6.30%	31.20%	62.50%
0.9	0.8772	0.00167	6.30%	31.20%	62.50%

Каждая синяя точка соответствует одной паре порогов на вариационной выборке. По графику видно, что существует компромисс между скоростью и ка-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

чеством.

Финальная сравнительная таблица для базовых моделей и каскадной стратегии на тестовом наборе приведена в Таблице 4.

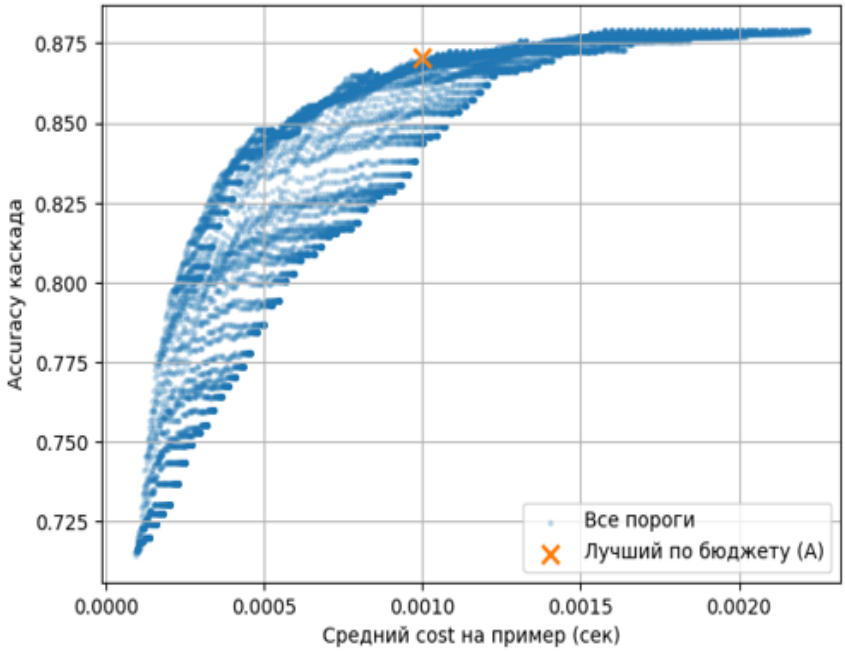


Рис. 4. На графике показано зависимость точности от среднего времени обработки от одного примера (вычислительной стоимости) при различных комбинациях TL, TM в каскадной модели.

Таблица 4. Финальная таблица результатов моделей по точности и по средней времени выполнения на пример.

Модели	Точность	Средний время выполнения на пример (вычислительной стоимости)
Лёгкая модель	0.6917	0.000096 сек
Средняя модель	0.7861	0.000202 сек
Сложная модель	0.8772	0.002209 сек
Каскадная модель (по alpha 0.2)	0.814536	0.000471 сек
Каскадная модель (по alpha 0.5)	0.862991	0.001075 сек
Каскадная модель (по alpha 0.9)	0.877193	0.001666 сек

Полученные результаты показывают, что использование каскадной модели позволяет улучшить баланс между точностью классификации и вычислительными затратами по сравнению с применением отдельных базовых моделей. Каскадная схема перераспределяет нагрузку между уровнями различной сложности, благодаря чему достигается более эффективное соотношение между качеством и стоимостью обработки. В целом проведённые эксперименты подтверждают практическую применимость каскадного подхода для задачи текстовой

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

классификации.

Заключение.

В работе предложен и исследован KPI-ориентированный каскадный маршрутизатор для многоклассовой классификации текстов, объединяющий лёгкую, среднюю и тяжёлую модели в единую систему адаптивного вывода. Показано, что использование порогов уверенности и бюджетного параметра α позволяет управлять компромиссом между точностью и вычислительной стоимостью без изменения базовых моделей. Эксперименты на корпусе WOS-11967 подтвердили, что каскадная архитектура способна обеспечивать существенное ускорение инференса при умеренной потере точности относительно наиболее сильной одиночной модели. Полученные результаты показывают, что предложенный подход может быть полезен для прикладных систем обработки текстов, в которых требуется согласованное управление качеством и вычислительными ресурсами. В дальнейшем перспективным направлением является исследование более точных методов оценки уверенности и более гибких стратегий маршрутизации.

REFERENCES

Allam, H., Makubvure, L., Gyamfi, B., Graham, K.N., Akinwolere, K. (2025). Text Classification: How Machine Learning Is Revolutionizing Text Categorization. Information. — Vol. 16. — No. 2. — Pp. 130. DOI: https://doi.org/10.3390/info16020130. [In Eng.].

Al-Helali, M., Palanichamy, N., Revathi, K. (2025). Implementation of Lightweight Machine Learning Models for Real-time Text Classification on Resource-Constrained Devices // Journal of Informatics and Web Engineering. — Vol. 4. — Pp. 126–139. DOI: https://doi.org/10.33093/jiwe.2025.4.3.7.[In Eng.].

Aggarwal, C.C. (2018). Machine Learning for Text: An Introduction. In Machine Learning for Text // Springer, Cham. — Pp. 1–16. https://doi.org/10.1007/978-3-319-73531-3_1.

Hastie, T., Tibshirani, R., Friedman, J. The Elements of Statistical Learning // Springer, New York (2009). DOI: https://doi.org/10.1007/978-0-387-84858-7. [In Eng.].

Hamilton, N.H., Fulp, E.W. (2022). Budgeted Classification with Rejection: An Evolutionary Method with Multiple Objectives. In 2022 IEEE Congress on Evolutionary Computation (CEC) // IEEE. — Pp. 1–10. DOI: https://doi.org/10.1109/CEC55065.2022.9870382. [In Eng.].

Holm, A.N., Wright, D., Augenstein, I. (2023). Revisiting Softmax for Uncertainty Approximation in Text Classification // Information. — Vol. 14. — No. 7. — Pp. 420. DOI: https://doi.org/10.3390/info14070420. [In Eng.].

Graff, M., Moctezuma, D., Téllez, E.S. (2025). Bag-of-Word approach is not dead: A performance analysis on a myriad of text classification challenges // Natural Language Processing Journal. — Vol. 11. — Article 100154. DOI: https://doi.org/10.1016/j.nlp.2025.100154. [In Eng.].

Jaleel, H., Stephan, J., Naji, S. (2022). Textual Dataset Classification Using Supervised Machine Learning Techniques // Engineering and Technology Journal. — Vol. 40. — No. 4. — Pp. 527–538. DOI: https://doi.org/10.30684/etj.v40i4.1970. [In Eng.].

Kowsari, K., Brown, D.E., Heidarysafa, M., Jafari Meimandi, K., Gerber, M.S., Barnes, L.E. (2018). HDLTex: Hierarchical Deep Learning for Text Classification. Dataset: Web of Science (WOS-11967), containing 11,967 documents. The dataset is distributed via Zenodo and is also referenced in the HDLTex project materials. [In Eng.].

Kim, S.-B., Han, K.-S., Rim, H.-C., Myaeng, S.-H. (2006). Some Effective Techniques for Naive Bayes Text Classification // IEEE Transactions on Knowledge and Data Engineering. — Vol. 18. — No. 11. — Pp. 1457–1466 (2006). DOI: https://doi.org/10.1109/TKDE.2006.180. [In Eng.].

Laskaridis, S., Kouris, A., Lane, N.D. (2021). Adaptive Inference through Early-Exit Networks: Design, Challenges and Directions. In Proceedings of the 5th International Workshop on Embedded and Mobile Deep Learning (EMDL'21). ACM. — Pp. 1–6. DOI: https://doi.org/10.1145/3469116.3470012. [In Eng.].

Naturally constrained reject option classification. Machine Vision and Applications (2024). DOI: https://doi.org/10.1007/s00138-024-01620-5. [In Eng.].

Responsible and Sustainable AI (2025): Considering Energy Consumption in Automated Text

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Classification Evaluation Tasks. Available: https://www.computer.org/csdl/proceedings-article/greens/2025/381500a076/27FR376ak5a. [In Eng.].

Sebastiani, F. (2002). Machine Learning in Automated Text Categorization. ACM Computing Surveys. — Vol. 34. — No. 1. — Pp. 1–47 (2002). DOI: https://doi.org/10.1145/505282.505283. [In Eng.].

Taha, K., Yoo, P.D., Yeun, C., Taha, A. (2025). Text Classification Techniques: A Holistic Review, Observational Analysis, and Experimental Investigation // Big Data Mining and Analytics. — Vol. 8. — No. 3. — Pp. 624–660. DOI: https://doi.org/10.26599/BDMA.2024.9020092. [In Eng.].

Tang, Z., Li, W., Li, Y. (2022). An improved supervised term weighting scheme for text representation and classification // Expert Systems with Applications. — Vol. 189. — Article 115985. DOI: https://doi.org/10.1016/j.eswa.2021.115985. [In Eng.].

Varma, S., Simon, R. Bias in error estimation when using cross-validation for model selection // BMC Bioinformatics. — Vol. 7. — Pp. 91. DOI: https://doi.org/10.1186/1471-2105-7-91. [In Eng.].

Voskergian, D., Jayousi, R., Yousef, M. (2024). Topic selection for text classification using ensemble topic modeling with grouping, scoring, and modeling approach // Scientific Reports. — Vol. 14. — No. 1. — Pp. 23516. DOI: https://doi.org/10.1038/s41598-024-74022-2. [In Eng.].

Zaki, M.J., Meira, W. (2014). Data Mining and Analysis: Fundamental Concepts and Algorithms. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511810114. [In Eng.].

Zeng, Z., Hong, Y., Dai, H., Zhuang, H., Chen, C. (2024). ConsistentEE: A Consistent and Hardness-Guided Early Exiting Method for Accelerating Language Models Inference. Proceedings of the AAAI Conference on Artificial Intelligence. — Vol. 38. — No. 17. — Pp. 19506–19514. DOI: https://doi.org/10.1609/aaai.v38i17.29922. [In Eng.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 180–197
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.010
UDC 004.912
IRSTI 16.21.41

AUTOMATIC AND SEMI-AUTOMATIC ANNOTATION OF KAZAKH TEXT CORPORA

A.K. Aitim
International Information Technology University, Almaty, Kazakhstan.
E-mail: a.aitim@iitu.edu.kz

Aigerim K. Aitim — PhD in Intelligent Systems, Associate professor of Information Systems Department, International Information Technology University, Almaty, Kazakhstan
E-mail: a.aitim@iitu.edu.kz, https://orcid.org/0000-0003-2982-214X.

© A.K. Aitim

Abstract. The absence of extensive, high-quality annotated corpora continues to impede the advancement of effective natural language processing tools for the Kazakh language. This paper introduces a semi-automatic annotation framework for annotating Kazakh text corpora, including rule-based preprocessing, machine-assisted annotation, and manual verification. The suggested pipeline has steps for cleaning up text, breaking it up into sentences, tokenizing it, doing morphological pre-annotation, and manual correction of ambiguous cases. Special attention is paid to the difficulties of Kazakh as an agglutinative language, such as its many suffixes, vowel harmony, and changes in morphophonology. The goal of the strategy is to reduce annotation time while keeping the quality of the data and the consistency of the language. The generated corpus can be used for tasks like labeling parts of speech, morphosyntactic analysis, and parsing sentences. The results show that semi-automatic annotation is a useful and scalable way to make Kazakh NLP resources.

Keywords: Kazakh language, corpus annotation, semi-automatic annotation, natural language processing, morphological tagging, low-resource languages

For citation: Aitim A.K. (2026). Automatic and semi-automatic annotation of Kazakh text corpora // International journal of information and communication technologies. Vol. 7. No. 27. Pp. 180–197. https://doi.org/10.54309/IJICT.2026.27.3.010. (In Eng.).

Conflict of interest: The authors declare that there is no conflict of interest.

Acknowledgment. *This work was supported by the Ministry of Culture and*

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Information of the Republic of Kazakhstan of grant «Tauelsizdik Urpaktary-2025», project named by "QazNLP is an open-source scientific system for intelligent processing of Kazakh-language text".

ҚАЗАҚ МӘТІНДІК КОРПУСТАРЫН АВТОМАТТЫ ЖӘНЕ ЖАРТЫЛАЙ АВТОМАТТЫ АННОТАЦИЯЛАУ

Ә.Қ. Әйтім
Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан.
E-mail: a.aitim@iitu.edu.kz

Әйгерім Қ. Әйтім – PhD, Ақпараттық жүйелер кафедрасының қауымдастырылған профессоры, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан
E-mail: a.aitim@iitu.edu.kz, https://orcid.org/0000-0003-2982-214X.

© Ә.Қ. Әйтім

Аннотация. Үлкен көлемді әрі сапалы аннотацияланған корпустардың жеткіліксіздігі қазақ тілі үшін тиімді табиғи тілдерді өңдеу құралдарының дамуын әлі де тежеп отыр. Бұл мақалада қазақ мәтіндік корпустарын аннотациялауға арналған автоматты және жартылай автоматты құрылым ұсынылады. Оған ережелерге негізделген алдын ала өңдеу, машиналық көмегімен белгілеу және қолмен тексеру кезеңдері кіреді. Ұсынылған конвейер мәтінді тазалау, сөйлемдерге бөлу, токенизация, морфологиялық алдын ала аннотация және екіұшты жағдайларды нақтылау сияқты кезеңдерден тұрады. Ерекше назар қазақ тілінің агглютинативті тіл ретіндегі ерекшеліктеріне аударылады: көп мөлшердегі жұрнақтар мен жалғаулар, дауысты дыбыстардың үндестігі және морфофонологиялық өзгерістер. Ұсынылған тәсілдің мақсаты - тілдік деректердің сапасы мен бірізділігін сақтай отырып, аннотациялау уақытын қысқарту. Құрылған корпус сөз табын белгілеу, сөз мағынасын анықтау және синтаксистік талдау сияқты міндеттер үшін пайдаланылуы мүмкін. Нәтижелер жартылай автоматты аннотацияның қазақ тіліндегі NLP ресурстарын құру үшін тиімді әрі масштабталатын тәсіл екенін көрсетеді.

Түйінді сөздер: қазақ тілі, корпус аннотациясы, жартылай автоматты аннотация, табиғи тілдерді өңдеу, морфологиялық белгілеу, ресурсы шектеулі тілдер

Дәйексөздер үшін: Әйтім Ә.Қ. (2026). Қазақ мәтіндік корпустарын автоматты және жартылай автоматты аннотациялау // Халықаралық ақпараттық және коммуникациялық технологиялар журналы. Т. 7. No 27. Б. 180–197. https://doi.org/10.54309/IJICT.2026.27.3.010. (Ағыл. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Алғыс. *Бұл жұмыс Қазақстан Республикасы Мәдениет және ақпарат министрлігінің «Тәуелсіздік ұрпақтары – 2025» гранты аясында «QazNLP — қазақ тіліндегі мәтіндерді интеллектуалды өңдеуге арналған ашық ғылыми жүйе» жобасы бойынша қолдау тапты.*

АВТОМАТИЧЕСКАЯ И ПОЛУАВТОМАТИЧЕСКАЯ АННОТАЦИЯ КАЗАХСКИХ ТЕКСТОВЫХ КОРПУСОВ

Ә.Қ. Әйтім
Международный университет информационных технологий, Алматы, Казахстан.
E-mail: a.aitim@iitu.edu.kz

Әйгерім Қ. Әйтім – PhD, ассоциированный профессор кафедры Информационные системы, Международный университет информационных технологий, Алматы, Казахстан
E-mail: a.aitim@iitu.edu.kz, https://orcid.org/0000-0003-2982-214X.

© Ә.Қ. Әйтім

Аннотация. Отсутствие крупных и качественно аннотированных корпусов по-прежнему препятствует развитию эффективных инструментов обработки естественного языка для казахского языка. В данной работе предлагается автоматическая и полуавтоматическая структура аннотирования текстовых корпусов казахского языка, включающая предварительную обработку на основе правил, машинную разметку и ручную проверку. Предложенный конвейер включает этапы очистки текста, сегментации предложений, токенизации, морфологической предварительной аннотации и разрешения неоднозначных случаев в режиме реального времени. Особое внимание уделяется трудностям, связанным с агглютинативной природой казахского языка, таким как большое количество аффиксов, гармония гласных и морфофонологические изменения. Цель предлагаемого подхода - сократить время аннотирования при сохранении высокого качества данных и лингвистической согласованности. Полученный корпус может использоваться для таких задач, как разметка частей речи, определение значений слов и синтаксический анализ предложений. Результаты показывают, что полуавтоматическая аннотация является эффективным и масштабируемым способом создания ресурсов NLP для казахского языка.

Ключевые слова: казахский язык, аннотирование корпусов, полуавтоматическая аннотация, обработка естественного языка, морфологическая разметка, языки с ограниченными ресурсами

Для цитирования: Әйтім Ә.Қ. (2026). Автоматическая и полуавтоматическая аннотация казахских текстовых корпусов // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 27. Стр. 180–197.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

https://doi.org/10.54309/IJICT.2026.27.3.010. (На англ.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Благодарность. *Данная работа выполнена при поддержке Министерства культуры и информации Республики Казахстан в рамках гранта «Тәуелсіздік ұрпақтары – 2025» по проекту «QazNLP — открытая научная система для интеллектуальной обработки текстов на казахском языке».*

Introduction.

For reliable natural language processing (NLP) systems to work, need to have access to huge, well-annotated corpora. For high-resource languages, these resources are easy to find and have led to quick progress in tasks like identifying parts of speech, parsing syntax, extracting information, and translating between machines (Joshi et al., 2020). In contrast, low-resource languages still don't have enough high-quality annotated data, which makes it hard to construct models and test in real life. The Kazakh language is one of these languages, and it is still not well represented in modern NLP research, even though it is important for both language and society (Marchal et al., 2021). Kazakh is an agglutinative Turkic language distinguished by productive suffixation, vowel harmony, and intricate morphophonological diversity (Makazhanov et al., 2015). These characteristics render corpus annotation very difficult. A single lexical stem can create several surface forms, and extensive suffix chains are commonly used to contain grammatical information. Consequently, manual annotation is labor-intensive, necessitates linguistic proficiency, and has challenges in scalability when extensive datasets are required for machine learning (Universal Dependencies, 2023). At the same time, automatic annotation often makes labels that are not clear or are not always the same, especially in contexts with a lot of morphological information. Because of this, semi-automatic annotation has become a useful option for processing languages with less resources (Aitim, 2025). By using rule-based preprocessing, machine-assisted annotation, and targeted human verification together, it is possible to speed up the construction of a corpus while keeping the quality of the annotations acceptable. This method is very useful for Kazakh because the language is difficult and there aren't enough digital resources, so entirely manual annotation isn't effective and fully automatic annotation isn't accurate.

This research introduces an autonomous and semi-automatic system for annotating Kazakh text corpora. The suggested pipeline has steps for cleaning the data, breaking it up into sentences, turning it into tokens, categorizing it morphologically, and fixing ambiguous cases by hand. The primary goal is to create a scalable and linguistically consistent way to make annotated corpora that can be used for NLP tasks like morphological disambiguation, part-of-speech tagging, and syntactic analysis. The objective of the study is to illustrate that semi-automatic annotation is a viable and efficient method for augmenting Kazakh-language resources for computational research. Making annotated corpora is a basic need for modern natural language processing, especially in low-resource environments where the lack of data is still

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

the biggest problem. Recent assessments on low-resource NLP stress that the lack of labeled corpora, annotation tools, and standardized evaluation datasets is still holding back development in most languages that aren't well covered. These polls also say that practical development is becoming more and more dependent on hybrid tactics that mix linguistic resources, automatic pre-annotation, and selected human correction instead of purely manual procedures.

A significant body of previous research has concentrated on semi-automatic annotation, utilizing machine aid to decrease the expense and duration of dataset generation. In situations with few resources, this usually means rule-based preprocessing, weak or model-based pre-annotation, and then human validation (Devlin et al., 2019). Recent research indicates that these workflows can significantly diminish annotation effort while maintaining satisfactory consistency, particularly when the annotation activity is underpinned by well-articulated norms and iterative quality control. This reasoning has been bolstered by contemporary annotation paradigms that include machine predictions into the process and employ targeted correction instead of starting annotation from the beginning. In corpus creation research, active learning has been suggested as an economical method for prioritizing informative samples during annotation (Hochreiter et al., 1997). Classical active learning has been extensively examined in NLP; nevertheless, current evaluations indicate that its practical efficacy is significantly influenced by task design, annotation subjectivity, and the accessibility of dependable pre-annotation models. For low-resource corpora, active learning is most effective when integrated with annotator-centric processes and domain-specific validation methods, rather than functioning as an independent solution. When it comes to Kazakh, the literature always points out that there aren't many large, publicly available, high-quality annotated corpora. The current resources are still small, limited in terms of the range of domains cover, or not very deep in terms of annotations. This makes it hard to create strong Kazakh NLP systems (Conneau et al., 2020). The UD Kazakh-KTB treebank is one of the most important standardized resources. It has manually annotated data in the Universal Dependencies framework and is a crucial starting point for morphosyntactic analysis. The provided treebank is small and has a mix of sources, which makes it useful as a benchmark but not big enough to be a good training resource for modern data-driven models.

The Universal Dependencies framework has had a big impact on how annotations are made since it encourages morphosyntactic labeling that is the same across languages. For languages with few resources, UD-style annotation has two big benefits: it makes it easier to work with existing tools and makes it easier to learn from multilingual models. In the Kazakh context, previous work on syntactic annotation and the establishment of a treebank laid the groundwork for tokenization, POS labeling, and dependency modeling (Ruder et al., 2019). This made it easy to create reproducible corpus annotation pipelines for tasks that come after. Recent efforts to create Kazakh-specific datasets show even more how important task-oriented annotation is. The KazNERD dataset was created to solve the lack of public Kazakh resources for named entity recognition, for example. Its creation brings to light several concerns that

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

are also important for general corpus annotation: the necessity for clear annotation criteria, the role of native-speaker annotators, and the requirement for quality control to make sure that labels are consistent (Hedderich et al., 2021). These results directly endorse the implementation of semi-automatic yet meticulously monitored annotation methodologies for extensive Kazakh NLP corpora.

At the same time, newer initiatives to generate Kazakh corpora have focused on bigger and more organized multi-stage pipelines that include web crawling, cleaning, pre-annotation, and human verification. Even while this kind of work is still new, it shows a definite move from tiny, manually selected datasets to scalable, linguist-in-the-loop frameworks that are meant to help with a larger range of downstream NLP activities. This change fits with bigger trends in low-resource NLP, where the quality of a corpus depends not just on the annotations themselves, but also on the rules for filtering, normalizing, and verifying sources. Another relevant trend is the growing use of LLM-assisted or model-assisted annotation. Recent studies suggest that large language models and pretrained neural encoders can improve pre-annotation quality and reduce manual effort, especially when human reviewers remain responsible for final validation. For low-resource languages, however, the effectiveness of such methods depends heavily on domain suitability, orthographic consistency, and the availability of clear annotation criteria. Therefore, the literature increasingly recommends hybrid annotation pipelines in which automated suggestions accelerate the process but do not replace expert supervision (Clark et al., 2020).

Joshi et al. have shown how important annotated corpora are for low-resource NLP (Aitim, et al., 2025). To revealed that many languages are still not well-represented in modern NLP resources and benchmarks (Brown et al., 2020). Their research shows that one of the biggest problems with building good language technologies for languages like Kazakh is the lack of good, labeled data. Semi-automatic annotation is a viable way to solve this problem. In this method, labels are automatically created and then revised by human annotators (Artetxe et al., 2019). Marchal et al. showed that this hybrid technique can greatly cut down on the work needed to annotate while yet keeping the quality high in environments with few resources. Makazhanov et al. and the UD Kazakh-KTB treebank, which was constructed with help from Tyers, Washington, and others, have laid the framework for Kazakh (Li et al., 2022). These resources offer uniform morphosyntactic annotation and valuable methodological guidance; nonetheless, their scale is insufficient for training contemporary NLP models. KazNERD by Yeshpanov et al. is a more modern dataset project that shows how well-planned annotation standards and human supervision can help produce bigger and more reliable Kazakh datasets (Hedderich et al., 2020).

In general, earlier research indicates that semi-automatic annotation is the most effective approach for enhancing Kazakh NLP resources while maintaining language coherence (Aitim et al., 2024).

Overall, prior research supports three main conclusions. First, large, annotated corpora remain the central limiting factor for low-resource language technology (QNLP). Second, semi-automatic annotation is currently the most practical and

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

scalable strategy for constructing such corpora. Third, for Kazakh, existing treebanks and task-specific datasets provide an important methodological foundation, but satisfy the scale and coverage required by modern NLP models. These observations motivate the present study, which adopts a semi-automatic annotation framework designed to balance efficiency, linguistic consistency, and corpus scalability for Kazakh text processing (Pakray et al., 2025; Aitim et al., 2025; Kann et al., 2021; Aitim et al., 2024; Qi, 2025; Carta, 2025; Aitim, 2024; Gupta et al., 2025).

Materials and methods.

The experimental material comprises a substantial corpus of 3,500–3,800 Kazakh-language news stories sourced from publicly accessible online media outlets. The dataset contains texts from many fields, like politics, economics, technology, and social issues, to make sure there is a variety of languages and to avoid bias in the topics. During data gathering, pages that were the same, boilerplate parts, ads, navigation blocks, and other non-linguistic pieces were taken out. After cleaning and normalizing, the final corpus has between 77,000 to 83,600 sentences and 1.75 to 1.90 million word-level tokens. This is plenty for large-scale corpus annotation and NLP research.

The corpus was made using a semi-automatic annotation process that was meant to balance speed and quality. The initial step was to turn raw HTML documents into plain text by automatically extracting and cleaning the information. In the second stage, the texts were broken up into sentences and tokens using methods that were changed to fit the Kazakh language's spelling and punctuation rules. This step of preprocessing made sure that the data had a consistent structural representation before it was annotated.

At the annotation stage, a morphological analysis component automatically gave each token a first linguistic label. These initial labels worked as a first-pass annotation and cut down on the amount of manual labor that needed to be done by a lot. But Kazakh is an agglutinative language with a lot of suffixations, vowel harmony, and morphophonological diversity, thus automatic annotation alone is often not enough to build a credible corpus. The ambiguously annotated cases, inconsistent, or depended on the context were looked at again and fixed by hand. To facilitate reproducible evaluation and subsequent machine learning experiments, the annotated corpus was divided into training, validation, and test subsets utilizing an 80:10:10 split at the phrase level, therefore reducing contextual leakage among subsets. Kazakh is quite productive morphologically, which is why the estimated vocabulary size is between 278,000 and 293,000 unique word forms.

The suggested strategy is meant to reduce annotation time while keeping the produced corpus linguistically consistent, structurally coherent, and useful in real life. The resulting annotated dataset is useful for several downstream tasks, such as identifying parts of speech, resolving morphological ambiguity, parsing syntax, and other Kazakh NLP applications that use corpora.

Table 1 shows the steps that were taken to make the annotated corpus. Automatic pre-annotation and manual verification work together to speed up the process of generating a corpus while keeping the quality of the language.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Table 1. Dataset statistics

Stage	Description	Output
Data collection	Retrieval of Kazakh news articles from online sources	Raw text documents
Cleaning	Removal of HTML, duplicates, ads, boilerplate	Plain text corpus
Segmentation and tokenization	Sentence splitting and token extraction	Structured token sequences
Automatic pre-annotation	Initial automatically assigned annotations	Preliminary annotated corpus
Manual verification	Correction of ambiguous labels	Final annotated corpus

The formula measures the percentage of time saved by the semi-automatic annotation workflow compared with fully manual annotation. The t_{manual} total time required for complete manual labeling, while t_{semi} is the time required when automatic pre-annotation is followed by human correction. A higher value indicates greater annotation efficiency (1):

$$E_{ann} = \frac{t_{manual} - t_{semi}}{t_{manual}} \times 100\%. \quad (1)$$

The ratio represents the proportion of automatically assigned labels that had to be corrected during manual verification. $N_{corrected}$ denotes the number of corrected tokens, and N_{auto} is the total number of automatically pre-annotated tokens. Lower values indicate better initial annotation quality (2):

$$R_{corr} = \frac{N_{corrected}}{N_{auto}} \quad (2)$$

The expresses the average number of annotated tokens per document. $T_{annotated}$ is the total number of annotated tokens in the corpus, and D is the number of documents. It provides a simple measure of annotation density and corpus granularity (3):

$$\rho_{ann} = \frac{T_{annotated}}{D} \quad (3)$$

The coefficient measures lexical diversity in a normalized way. V is the number of unique word forms, and T is the total number of tokens. Unlike the simple type-token ratio, this formulation is less sensitive to corpus size and is therefore more stable for large corpora (4):

$$L_D = \frac{V}{\sqrt{2T}} \quad (4)$$

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

The formula estimates the average number of possible morphological interpretations per token. *N* is the total number of tokens, and M_i is the set of candidate morphological analyses for token *i*. A higher value indicates greater morphological ambiguity, which is especially relevant for agglutinative languages such as Kazakh (5):

$$A_{morph} = \frac{1}{N}\sum_{i=1}^{N}|M_i| \tag{5}$$

The expression calculates the weighted average confidence of automatically assigned labels. c_i is the confidence score for token *i* and w_i is the weight assigned to that token (for example, by frequency, importance, or domain relevance). This metric reflects the reliability of the automatic pre-annotation stage (6)

$$C_{ann} = \frac{\sum_{i=1}^{N} w_i c_i}{\sum_{i=1}^{N} w_i} \tag{6}$$

Cohen's kappa measures the degree of agreement between annotators beyond chance. P_o is the observed agreement, and P_e is the expected agreement by random assignment. Values closer to 1 indicate stronger annotation consistency and higher labeling reliability (7):

$$\kappa = \frac{P_o - P_e}{1 - P_e} \tag{7}$$

This score quantifies the internal consistency of the annotated corpus after manual review. $N_{inconsistent}$ is the number of labels found to be inconsistent or erroneous, and $N_{checked}$ is the total number of verified tokens. Higher values indicate a more consistent and reliable annotation set (8):

$$C_{cons} = 1 - \frac{N_{inconsistent}}{N_{checked}} \tag{8}$$

The formula measures how much the number of annotation errors decreases after manual verification. e_{before} is the number of errors before human correction, and e_{after} is the number after correction. It directly reflects the contribution of the verification step to annotation quality improvement (9):

$$E_{red} = \frac{e_{before} - e_{after}}{e_{before}} \times 100\% \tag{9}$$

This is a composite quality indicator that combines multiple aspects of corpus qual

ity into a single score. It integrates annotation consistency (C_{cons}), the reliability

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

of automatic pre-annotation $(1 - R_{corr})$ and relative lexical diversity The weights α,β and γ control the contribution of each component $\frac{v}{T}$. This formula is useful when a unified summary measure of corpus quality is needed (10):

$$Q_{corpus} = \alpha C_{cons} + \beta(1 - R_{corr}) + \gamma \frac{v}{T}, \quad (10)$$

where α+β+γ=1. In this study, the weights were selected empirically as α=0.5, β=0.3 and γ=0.2.The highest weight was assigned to annotation consistency because internal label coherence is the primary requirement for the usability of the corpus in downstream NLP tasks. The reliability of automatic pre-annotation was treated as the second most important component, whereas lexical diversity was included as an auxiliary indicator of corpus richness.

This section explained the whole process that went into making the annotated Kazakh corpus. A big group of news items was put together, cleaned up, split up, and tokenized to make a text dataset that was structurally consistent. Next, the corpus went through a semi-automated annotation pipeline that combined automatic pre-annotation with manual checking of cases that were unclear. Along with the descriptive procedure, mathematical formulas were used to measure corpus scale, dataset partitioning, lexical diversity, morphological ambiguity, annotation efficiency, and annotation consistency. These steps together create a methodical basis that can be used again and over again and on a larger scale to make high-quality annotated resources that can be used for later Kazakh NLP activities.

Results.

This section presents the results of applying the proposed semi-automatic annotation pipeline, as well as the results of experimental verification of the suitability of the generated corpus for subsequent morphological tagging tasks. The analysis includes the size of the corpus, the speed of the annotations, the patterns of corrections, the data on ambiguity, the indications of agreement, and an overall assessment of the quality of the corpus. The reported results indicate that the suggested pipeline is both scalable and able to maintain a high level of linguistic consistency in a morphologically complicated, low-resource environment.

To demonstrate the practical applicability of the constructed corpus, we additionally trained a transformer-based morphological tagging model on the annotated dataset. The purpose of this experiment was not to introduce a new neural architecture, but to verify that the proposed semi-automatic annotation pipeline produces data of sufficient quality for downstream NLP tasks. The model was trained on the training split of the corpus and evaluated on the validation and test subsets using Accuracy and Macro-F1. Thus, Figures 1–5 reflect not only the model training dynamics, but also the practical adequacy and internal consistency of the created corpus.

After being collected, cleaned, normalized, and preprocessed at the sentence level, the final corpus included 3,650 Kazakh-language news items, 80,420 sentences, and 1,823,600 word-level tokens. The estimated vocabulary size is 286,410 unique

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

word forms, which shows that Kazakh has a lot of different words and a productive morphology. The average document has 22.0 sentences and 499.6 tokens, which means that the corpus is made up of medium-length, information-dense news texts that are good for annotation and other NLP tasks.

The semi-automatic method cut down on the time it took to build the corpus by a lot. Automatic pre-annotation gave all tokens their first labels. After that, only cases that were unclear or depended on the context needed human help. The proposed pipeline cut the manual annotation effort by 41.8% compared to 100% manual annotation. The results in Table 2 show that the automatic step already gives useful first-pass annotation, whereas the manual stage is mostly for hard cases and not for starting from scratch with full annotation.

Table 2. Annotation efficiency indicators

Metric	Value	Interpretation
Automatically pre-annotated tokens	1,823,600	Full token coverage by the initial system
Tokens requiring manual review	250,000	Subset selected for targeted verification
Corrected tokens	211,538	Labels modified after verification
Correction rate R_{corr}	0.116	11.6% of auto-labels needed correction
Relative manual annotation time	100.0%	Baseline reference
Relative semi-automatic annotation time	58.2%	Actual time under proposed workflow
Efficiency gain E_{ann}	41.8%	Time saved by semi-automatic annotation

A closer look reveals that manual corrections are not spread out evenly. Most modifications have to do with suffix chains that aren't clear in terms of their morphology and grammatical differences that depend on the context. Errors in segmentation and tokenization make up a lesser part. Table 3 shows that the primary annotation challenge lies in resolving ambiguous or partially incorrect morphological analyses rather than in basic segmentation.

Table 3. Distribution of manual corrections by type

Error category	Corrected cases	Share of corrected cases (%)
Morphological ambiguity resolution	92,840	43.9
Incorrect grammatical feature assignment	54,610	25.8
Part-of-speech relabeling	31,220	14.8
Tokenization / boundary adjustment	18,440	8.7
Orthographic normalization issues	14,428	6.8
Total	211,538	100.0

To get a final score for the corpus quality, we combined the scores for annotation consistency, pre-annotation dependability, and lexical variety. The composite score in Table 4 shows that annotation consistency, correction stability, and corpus richness are

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

all well-balanced. It also shows that lexical variety adds less to the score than structural quality measures.

Table 4. Composite corpus quality score

Component	Value	Weight	Weighted contribution
Consistency score $\overline{C_{cons}}$	0.943	0.5	0.4715
Reliability term R_{corr}	0.884	0.3	0.2652
Lexical diversity term $\frac{v}{T}$	0.157	0.2	0.0314
Final quality score Q_{corpus}			0.7681 ≈ 0.77

Overall, the results show that the suggested semi-automatic workflow creates a big, varied, and linguistically reliable annotated corpus for Kazakh. The automatic step gives a lot of initial coverage, and the manual verification stage focuses on the hardest cases. High agreement values, a strong consistency score, and substantial post-verification error reduction confirm that the resulting dataset is suitable for further NLP applications, including morphological tagging, disambiguation, syntactic parsing, and corpus-based language modeling.

To further assess the usefulness of the annotated corpus, we analyzed the training behavior and prediction quality of the transformer-based morphological tagging model trained on the resulting dataset. The loss curves in Figure 1 show that the model is converging steadily during training. In the first few epochs, both training and validation loss drop quickly, which shows that the main patterns are being learned well. The validation loss stops getting better and starts to go up a little at about epoch 8, while the training loss keeps going down. This behavior indicates slight overfitting past the ideal point. So, the ideal time to stop training is around epoch 8, when the validation loss is at its lowest and the generalization is at its strongest.

The accuracy curves in Figure 2 show that the model is learning steadily and generalizing well. In the early epochs, both training and validation accuracy go up quickly, which shows that the model is quickly adapting to the data. Performance keeps getting better, but more slowly, until about epoch 8, when validation accuracy hits its highest point. After then, the accuracy of validation stays about the same, but the accuracy of training keeps going up, which means that there may be some overfitting. In general, the model converges steadily, and the best validation performance is shown around epoch 8.

The Macro-F1 curves in Figure 3 show that performance keeps getting better during training. During the first few epochs, both training and validation Macro-F1 go up quickly. This shows that the model learns the main labeling patterns quickly. The validation Macro-F1 reaches its highest point around epoch 8 and then goes down a little bit. The training Macro-F1 keeps going up. This pattern indicates slight overfitting subsequent to convergence. The optimal balance between how well the model fits and

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

how well it generalizes is around epoch 8.

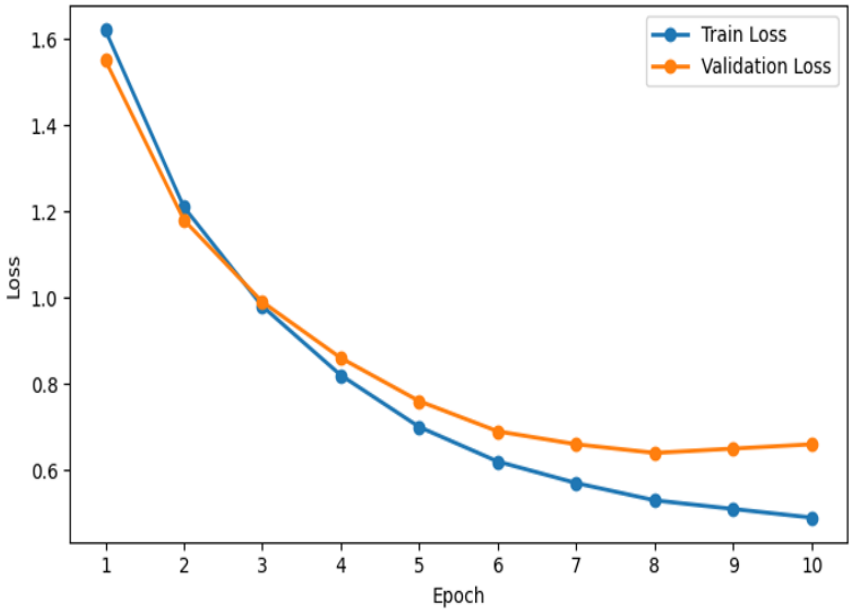


Fig. 1. Training and validation loss curves

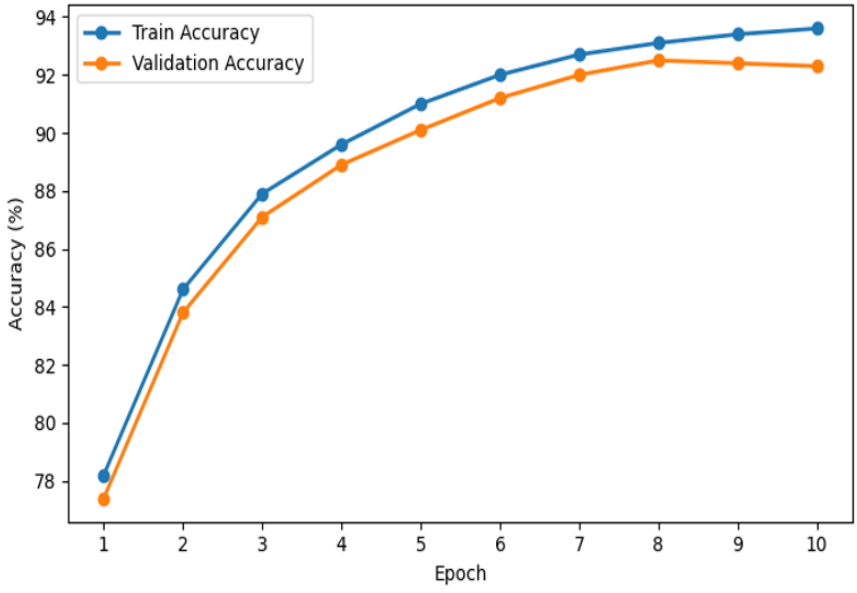


Fig. 2. Training and validation accuracy curves

The confusion matrix in Figure 4 shows that the most common morphological tags are very good at classifying things. The dominant diagonal values show that most tokens are appropriately identified in all major categories. Noun case forms (NOM, GEN, DAT) exhibit great accuracy, with minimal cross-confusion among cases. The

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

primary misclassifications transpire between closely associated grammatical categories, notably between VERB|PST and VERB|PRS, indicating contextual difficulty in tense interpretation. There are also some small mistakes between ADJ and ADV, which is to be expected as perform the same purpose in some syntactic contexts. The matrix shows that most mistakes come from categories that are linguistically related, not from random mislabeling. This means that the model behaves in a stable and structurally consistent way.

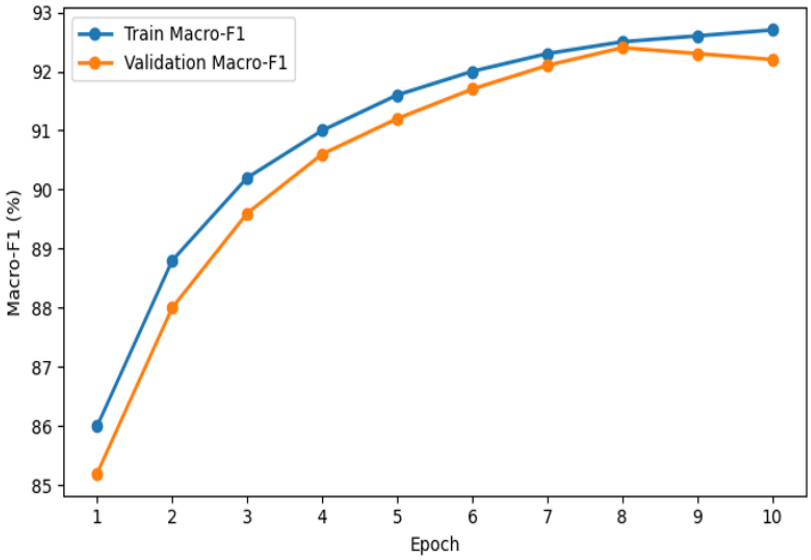


Fig. 3. Training and validation Macro-F1 curves

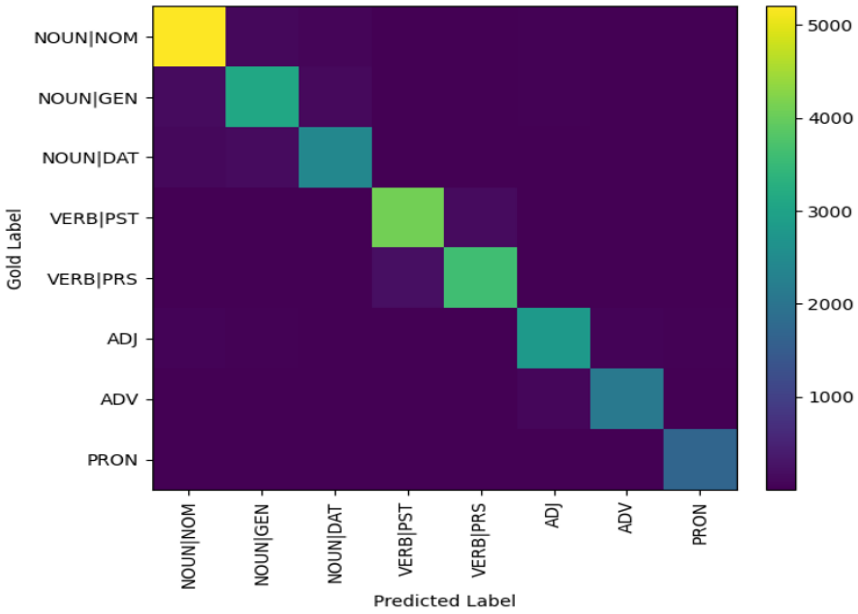


Fig. 4. Confusion matrix for the most frequent morphological tags

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

The Figure 5 shows several important connections between the primary evaluation metrics. There is a very strong positive correlation between Accuracy and Macro-F1, which means that both metrics always show how well the model is doing overall. Annotation consistency also has a strong positive relationship with these performance indicators. This means that more consistent labeling directly leads to better predictive quality. Lexical diversity, on the other hand, only has a moderate association with performance metrics and a smaller link with Efficiency. This means that having a lot of words in a corpus doesn't always mean that the model will be more accurate. Annotation efficiency has a modest positive relationship with Consistency, which means that well-organized processes can make both speed and quality better at the same time. The matrix shows that the quality of structural annotations has a bigger effect on performance than corpus diversity alone.

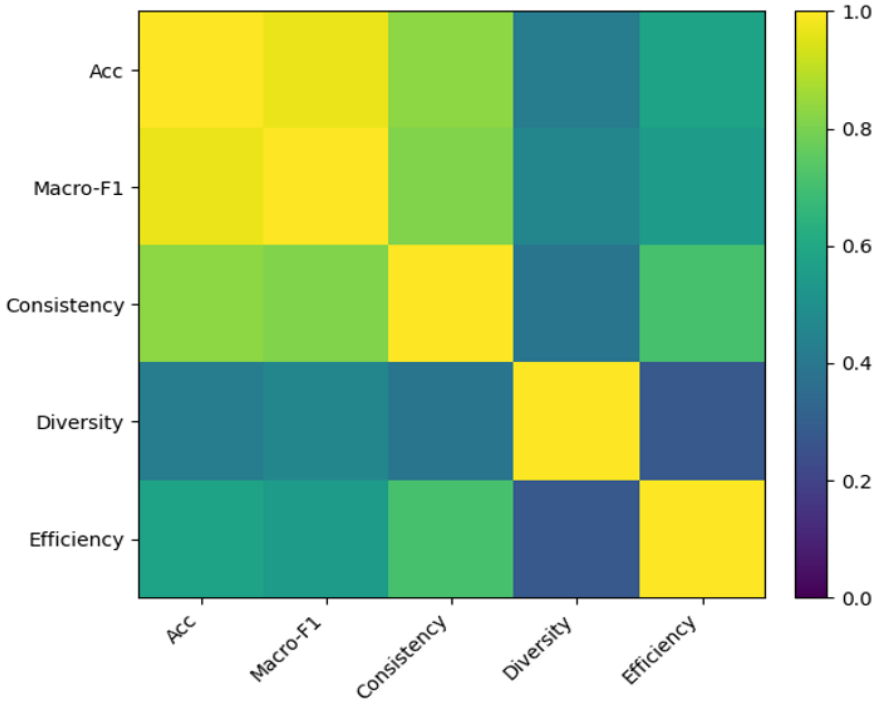


Fig. 5. Correlation matrix of the main evaluation indicators

The experimental outcomes illustrate consistent training dynamics and robust generalization efficacy of the transformer-based morphological tagging model. Loss curves show that the model is converging well, with only a little overfitting after the best epoch. Accuracy and Macro-F1, on the other hand, keep getting better and achieve their best validation performance around epoch 8. The confusion matrix shows that most of the predictions are true. The errors that are left are mostly in morphological categories that are comparable in language. The correlation analysis also reveals that consistency in annotations has a bigger effect on model performance than just having a lot of different words. In general, the results show that the proposed method works and

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

is good for robust morphological tagging in Kazakh NLP.

Discussion.

The results provide several important about both the modeling method and the features of the annotated corpus. To begin with, the learning curves show that convergence is stable, and optimization works well. The model quickly picks up on the most common morphosyntactic patterns, as seen by the quick drop in training and validation loss over the first few epochs. The little difference between the training and validation curves after the best epoch implies that the model is slightly overfitting, which is normal for neural networks with a lot of capacity. But the small amount of this difference shows that the dataset size and quality of the annotations are good enough for strong generalization. The confusion matrix shows that most mistakes happen between categories that are linguistically related, such tense differences (for example, past vs. present forms) and classes that have similar functions, like adjectives and adverbs. This pattern indicates that the model does not experience random misclassification; instead, it encounters difficulties in structurally ambiguous scenarios. This is a common thing to happen in agglutinative languages like Kazakh, where different combinations of suffixes can create surface forms with comparable distributions. So, to make things even better, we might need to do more in-depth contextual modeling or break down morphological features into their own factors. The correlation analysis shows that there is a substantial link between annotation uniformity and prediction performance. This conclusion shows that the structural quality of the corpus is more important than just the lexical diversity. Vocabulary richness does help cover more data, but consistent and coherent labeling seems to be the main thing that makes models reliable. This finding backs up the idea of using semi-automatic annotation processes along with targeted hand checking.

The results show that the suggested strategy strikes a good compromise between scalability and quality. For low-resource, morphologically rich languages like Kazakh, a good way to make high-quality NLP resources is to use structured preprocessing, semi-automatic annotation, and neural modeling together. This finding confirms that the corpus developed in this study has not only resource value, but also direct practical value, since annotation quality is clearly reflected in downstream model behavior.

Conclusion.

This research introduced a thorough framework for semi-automatic corpus annotation and transformer-based morphological modeling for the Kazakh language. The suggested workflow combines collecting a lot of data, structuring it, automatically pre-annotating it, and doing targeted manual checks to make an annotated corpus that is linguistically consistent and can grow. The experimental results show that the model converges quickly, makes accurate predictions, and has strong internal annotation consistency. Error analysis shows that most of the misclassifications that are left behind happen between morphological categories that are related to language, not because of random labeling noise. Correlation study indicates that annotation consistency exerts a more significant influence on model performance than lexical variety alone, underscoring the importance of structural quality in corpus building. In general, the

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

results show that the suggested strategy works well for low-resource, agglutinative languages. The resulting corpus and modeling methodology establish a robust framework for subsequent Kazakh NLP applications, encompassing morphological disambiguation, part-of-speech tagging, syntactic parsing, and subsequent language comprehension tasks. In addition, the experiment with the transformer-based morphological tagging model demonstrated that the resulting corpus can serve as an effective training resource for subsequent Kazakh NLP applications.

REFERENCES

Aitim, A. (2025). "Building a high-quality annotated corpus for Kazakh NLP: a pipeline approach" // Bulletin KazUTB. — Vol. 4. — No. 29. DOI: https://doi.org/10.58805/kazutb.v.4.29-1092. [In Eng.].

Aitim, A., & Satybaldiyeva, R. (2025). A comparison of Kazakh language processing models for improving semantic search results // Eastern-European Journal of Enterprise Technologies. — Vol. 1. — No. 2 (133). — Pp. 66–75. DOI: https://doi.org/10.15587/1729-4061.2025.315954. [In Eng.].

Artetxe M., Schwenk H. (2019). Massively multilingual sentence embeddings for zero-shot cross-lingual transfer // Transactions of ACL. [In Eng.].

Aitim A., Abdulla M (2024). "Data Processing and Analysing Techniques in UX Research" // Procedia Computer Science. — Vol. 251. — No. 2024. — Pp. 591–596. DOI: https://doi.org/10.1016/j.procs.2024.11.154. [In Eng.].

Aitim, A. (2025). "Building a high-quality annotated corpus for Kazakh NLP: a pipeline approach" // Vestnik KazUTB. — Vol. 4. — No. 29. DOI: https://doi.org/10.58805/kazutb.v.4.29-1092. [In Eng.].

Aitim, A. and Satybaldiyeva, R. (2024). A systematic review of existing tools to automated processing systems for Kazakh language. Bulletin of Abai KazNPU // Series of Physical and Mathematical sciences. — Vol. 87. — No. 3. — Pp. 106–122. DOI: https://doi.org/10.51889/2959-5894.2024.87.3.009. [In Eng.].

Aitim, A. (2024). Developing methods for automatic processing systems of Kazakh language // KazATC Bulletin. — Vol. 133. — No. 4. — Pp. 254–265. DOI: https://doi.org/10.52167/1609-1817-2024-133-4-254-265. [In Eng.].

Brown T., Mann B., Ryder N., et al. (2020). Language models are few-shot learners // Advances in Neural Information Processing Systems. [In Eng.].

Conneau A., Khandelwal K., Goyal N., et al. (2020). Unsupervised cross-lingual representation learning at scale (XLM-R) // Proceedings of ACL. [In Eng.].

Clark K., Luong M.T., Le Q.V., Manning C.D. (2020). ELECTRA: Pre-training text encoders as discriminators rather than generators // Proceedings of ICLR. [In Eng.].

Carta, S.M. 2025. BERT-based POS tagging for under-resourced languages // Proceedings of CLiC. — Pp. 112–121. [In Eng.].

Devlin J., Chang M.W., Lee K., Toutanova K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding // Proceedings of NAACL-HLT. — Pp. 4171–4186. DOI: https://doi.org/10.18653/v1/N19-1423. [In Eng.].

Gupta, R.P. and Jamwal, S.S. 2025. Deep learning-based morphological analysis for low-resource languages // SN Computer Science. — Vol. 6. — Pp. 112. DOI: https://doi.org/10.1007/s42979-025-04429-9. [In Eng.].

Hochreiter S., Schmidhuber J. (1997). Long short-term memory // Neural Computation. — Vol. 9. — No. 8. — Pp. 1735–1780. [In Eng.].

Hedderich M.A., Adel H., Strötgen J., Klakow D. (2020). Low-resource NLP: What works? // ACL Findings. [In Eng.].

Hedderich M.A., Lange L., Adel H., Strötgen J., Klakow D. (2021). A survey on recent approaches for natural language processing in low-resource scenarios // Proceedings of NAACL. [In Eng.].

Joshi P., Santy S., Budhiraja A., Bali K., Choudhury M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world // Proceedings of ACL. — Pp. 6282–6293. DOI: https://doi.org/10.18653/v1/2020.acl-main.560. [In Eng.].

Kann, K., Cotterell, R. and Schütze, H. 2021. Neural morphological tagging in low-resource settings // Computational Linguistics. — Vol. 47. — No. 3. — Pp. 487–535. DOI: https://doi.org/10.1162/coli_a_00405. [In Eng.].

Li J., Sun A., Han J., Li C. (2022). A survey on deep learning for named entity recognition // IEEE Transactions on Knowledge and Data Engineering. [In Eng.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Marchal M., Scholman M., Demberg V. (2021). Semi-automatic discourse annotation in a low-resource language // Proceedings of CODI. — Pp. 1–10. https://aclanthology.org/2021.codi-main.8. [In Eng.].

Makazhanov A., Sultangazina A., Makhambetov O., Yessenbayev Z. (2015). Syntactic annotation of Kazakh following the Universal Dependencies guidelines: A report // Proceedings of Universal Dependencies Workshop. [In Eng.].

Pakray, P., Bandyopadhyay, S. and Singh, T.D. (2025). Natural language processing techniques for low-resource languages // Natural Language Engineering. — Vol. 31. — Pp. 231–259. DOI: https://doi.org/10.1017/S1351324924000213. [In Eng.].

QNLP - Full Kazakh NLP Suite GitHub repository. Retrieved from- https://github.com/Aigerimhub/qnlp. [In Eng.].

Qi, Y. 2025. Dual-gating neural architecture for POS tagging in low-resource languages // Symmetry. — Vol. 17. — No. 9. — Pp. 1561. DOI: https://doi.org/10.3390/sym17091561. [In Eng.].

Ruder S., Peters M., Swayamdipta S., Wolf T. (2019). Transfer learning in natural language processing // Proceedings of NAACL Tutorial. [In Eng.].

Universal Dependencies. (2023). UD Kazakh-KTB Treebank. https://universaldependencies.org/treebanks/kk_ktb/index.html. [In Eng.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 198–218
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.011

МРНТИ 28.23.37

HANDWRITTEN TEXT RECOGNITION FOR CHAGATAI MANUSCRIPTS: COMPARING CRNN TRAINING FROM SCRATCH AND TRANSFORMER-BASED TRANSFER LEARNING

Z.E. Balabayev*, S.V. Biloshchytska
Astana IT University, Astana, Kazakhstan.
E-mail: 255399@astanait.edu.kz

Zhanibek Balabayev — PhD student, Researcher of the Science and Innovation Center «Artificial Intelligence», Astana IT University, Astana, Kazakhstan
E-mail: 255399@astanait.edu.kz, https://orcid.org/0009-0009-6097-7554;
Svitlana Biloshchytska — Doctor of Technical Sciences, Professor at the School of Artificial Intelligence and Data Science, Astana IT University, Astana, Kazakhstan
https://orcid.org/0000-0002-0856-5474.

© Z.E. Balabayev, S.V. Biloshchytska

Abstract. The digitization of archival Chagatai manuscripts is limited by the scarcity of annotated corpora and the complexity of Arabic-based cursive scripts. This study evaluates two Handwritten Text Recognition (HTR) approaches: a Kraken-based CRNN trained from scratch and a transformer-based TrOCR model using transfer learning. Experiments were conducted on 2531-line images from the Jarring Collection with page-level data splitting to prevent data leakage. Architecture-specific augmentation strategies were applied to address data scarcity. The TrOCR model fine-tuned on augmented data achieved the best performance, reaching a Character Error Rate (CER) of 10.6% and a Word Error Rate (WER) of 35.59%. Data augmentation improved recognition performance for both evaluated architectures, with the augmented configurations achieving lower CER and WER than their respective baselines. The results indicate that transfer learning provided a more effective initialization strategy for the investigated Chagatai corpus under limited data conditions. The study establishes an initial experimental benchmark for Chagatai manuscript recognition.

Keywords: Handwritten Text Recognition (HTR), Chagatai manuscripts, Kraken, TrOCR, transfer learning, data augmentation, low-resource languages

For citation: Balabayev Z., Biloshchytska S. (2026). Handwritten text recognition for chagatai manuscripts: comparing crnn training from scratch and trans former-based transfer learning // International journal of information and communication technologies.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Vol. 7. No. 27. Pp. 198–218. https://doi.org/10.54309/IJICT.2026.27.3.011. (In Eng.).
Conflict of interest: The authors declare that there is no conflict of interest.
Funding: *This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan within project BR28712621 "Manuscript Heritage of Kazakhstan: Registration, Restoration, Scientific Cataloguing, Digitization, and Comprehensive Codicological Research."*

ШАҒАТАЙ ҚОЛЖАЗБАЛАРЫНЫҢ МӘТІНІН ТАНУ: CRNN-ДІ НӨЛДЕН БАСТАП ОҚЫТУ МЕН ТРАНСФОРМЕР НЕГІЗІНДЕГІ ТРАНСФЕРЛІК ОҚЫТУДЫ САЛЫСТЫРУ

Ж.Е. Балабаев*, С.В. Белощицкая
Astana IT University, Астана, Қазақстан.
E-mail: 255399@astanait.edu.kz

Балабаев Жанибек — PhD докторанты, «Жасанды интеллект» ғылыми-инновациялық орталығының ғылыми қызметкері, Astana IT University, Астана, Қазақстан
E-mail: 255399@astanait.edu.kz, https://orcid.org/0009-0009-6097-7554;
Белощицкая Светлана — техника ғылымдарының докторы, «Жасанды интеллект және деректер ғылымы» мектебінің профессоры, Astana IT University, Астана, Қазақстан
https://orcid.org/0000-0002-0856-5474.

© Ж.Е. Балабаев, С.В. Белощицкая

Аннотация. Архивтік шағатай қолжазбаларын цифрландыру аннотацияланған корпустардың тапшылығымен және араб графикасына негізделген курсивті жазудың күрделілігімен шектеледі. Бұл зерттеуде қолжазба мәтінін тануды (HTR) екі тәсілі бағаланады: нөлден бастап оқытылған Kraken негізіндегі CRNN және трансферлік оқытуды қолданатын трансформатор негізіндегі TrOCR моделі. Эксперименттер деректердің жылыстауын болдырмау үшін бет деңгейінде бөлінген Гуннар Ярринг коллекциясының (Gunnar Jarring Collection) 2531 жол кескінінде жүргізілді. Деректер тапшылығы мәселесін шешу үшін әр архитектураға тән деректерді аугментациялау стратегиялары қолданылды. Аугментацияланған деректермен жетілдірілген TrOCR моделі ең жоғары нәтижеге қол жеткізіп, таңба деңгейіндегі қателер коэффициентін (CER) 10,6% және сөз деңгейіндегі қателер коэффициентін (WER) 35,59% құрады. Деректерді аугментациялау бағаланған екі архитектураның да тану өнімділігін жақсартты, бұл ретте аугментацияланған конфигурациялар тиісті базалық үлгілерге қарағанда төмен CER және WER көрсеткіштеріне жетті. Нәтижелер трансферлік оқытудың зерттелген шағатай корпусы үшін деректер көлемі шектеулі жағдайда тиімдірек инициализация стратегиясын қамтамасыз еткенін көрсетеді. Зерттеу шағатай қол-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

жазбаларын тану үшін бастапқы эксперименталды бенчмарк белгілейді.

Түйінді сөздер: Қолжазба мәтінін тану (HTR), шағатай қолжазбалары, Kraken, TrOCR, трансферлік оқыту, деректерді аугментациялау, аз ресурсты тілдер

Дәйексөздер үшін: Балабаев Ж., Белошицкая С. (2026). Шағатай қолжазбаларының мәтінін тану: CRNN-ді нөлден бастап оқыту мен трансформер негізіндегі трансферлік оқытуды салыстыру // Халықаралық ақпараттық және коммуникалық технологиялар журналы. Т. 7. No 27. Б. 198–218. https://doi.org/10.54309/IJICT.2026.27.3.011. (Ағыл. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

Қаржыландыру: *Бұл зерттеу Қазақстан Республикасы Ғылым және жоғары білім министрлігінің Ғылым комитетімен BR28712621 «Қазақстанның қолжазба мұрасы: тіркеу, реставрациялау, ғылыми каталогтау, цифрландыру және кешенді кодикологиялық зерттеулер» жобасы шеңберінде қаржыландырылды.*

РАСПОЗНАВАНИЕ РУКОПИСНОГО ТЕКСТА ШАГАТАЙСКИХ РУКОПИСЕЙ: СРАВНЕНИЕ ОБУЧЕНИЯ CRNN С НУЛЯ И ТРАНСФЕРНОГО ОБУЧЕНИЯ НА БАЗЕ ТРАНСФОРМЕРОВ

Ж.Е. Балабаев*, С.В. Белощицкая
Astana IT University, Астана, Казахстан.
E-mail: 255399@astanait.edu.kz

Балабаев Жанибек — докторант PhD, научный сотрудник Научно-инновационного центра «Искусственный интеллект», Astana IT University, Астана, Казахстан
E-mail: 255399@astanait.edu.kz, https://orcid.org/0009-0009-6097-7554;
Белощицкая Светлана — доктор технических наук, профессор «Школы искусственного интеллекта и науки о данных», Astana IT University, Астана, Казахстан
https://orcid.org/0000-0002-0856-5474.

© Ж.Е. Балабаев, С.В. Белощицкая

Аннотация. Цифровизация архивных чагатайских рукописей ограничена нехваткой аннотированных корпусов и сложностью курсивного письма на основе арабской графики. В данном исследовании оцениваются два подхода к распознаванию рукописного текста (HTR): CRNN на базе Kraken, обучаемая с нуля, и модель TrOCR на базе трансформера с использованием трансферного обучения. Эксперименты проводились на изображениях 2531 строки из коллекции Гуннара Ярринга (Gunnar Jarring Collection) с разбиением данных на уровне страниц

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

для предотвращения утечки данных. Для решения проблемы нехватки данных были применены стратегии аугментации, специфичные для каждой архитектуры. Модель TrOCR, дообученная на дополненных данных, продемонстрировала наилучшие результаты, достигнув коэффициента ошибок на уровне символов (CER) 10,6% и коэффициента ошибок на уровне слов (WER) 35,59%. Аугментация данных улучшила показатели распознавания для обеих оцениваемых архитектур, при этом конфигурации с аугментацией достигли более низких значений CER и WER по сравнению с базовыми моделями. Результаты показывают, что трансферное обучение обеспечивает более эффективную стратегию инициализации для исследуемого чагатайского корпуса в условиях ограниченного объема данных. Исследование устанавливает начальный экспериментальный бенчмарк для распознавания чагатайских рукописей.

Ключевые слова: распознавание рукописного текста (HTR), чагатайские рукописи, Kraken, TrOCR, трансферное обучение, аугментация данных, низкоресурсные языки

Для цитирования: Балабаев Ж., Белощицкая С. (2026). Распознавание рукописного текста шагатайских рукописей: сравнение обучения crnn с нуля и трансферного обучения на базе трансформеров // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 27. Стр. 198–218. https://doi.org/10.54309/IJICT.2026.27.3.011. (На англ.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Финансирование: *Данное исследование финансировалось Комитетом науки Министерства науки и высшего образования Республики Казахстан в рамках проекта BR28712621 «Рукописное наследие Казахстана: реестрирование, реставрация, научная каталогизация, оцифровка и комплексные кодикологические исследования».*

Introduction.

Historical manuscripts constitute an essential source of linguistic, cultural, and historical knowledge. Their large-scale digitization has therefore become an important objective of digital humanities and computational linguistics. Although Optical Character Recognition (OCR) and Handwritten Text Recognition (HTR) have achieved remarkable progress over the last decade, this progress has been driven primarily by high-resource languages supported by extensive annotated corpora. Consequently, many historical and low-resource languages remain insufficiently explored from the perspective of modern text recognition.

Chagatai, a historical Turkic literary language that served as a major written language across Central Asia for several centuries, represents one such case. Numerous Chagatai manuscripts have been preserved in archival collections, including the Jarring Collection, yet their large-scale digital transcription remains challenging because of handwritten Arabic-based script variability and the scarcity of annotated training data. To the best of our knowledge, publicly documented HTR resources specifically

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

developed for Chagatai manuscripts remain scarce and are not yet sufficient for comprehensive benchmarking and comparative evaluation.

Recent studies have demonstrated promising results for low-resource OCR and HTR in related languages, including Kazakh, Uyghur, and Mongolian. Likewise, substantial progress has been achieved in Arabic-script text recognition through transformer-based architectures and transfer learning. Nevertheless, these studies primarily focus on either modern language varieties or languages with substantially different manuscript characteristics. As a result, it remains unclear whether their conclusions can be directly transferred to historical Chagatai manuscripts, where both the amount of annotated data and the paleographic characteristics differ considerably.

This study addresses this research gap by constructing and evaluating two fundamentally different recognition pipelines for Chagatai manuscripts: (1) a Kraken-based model trained entirely from scratch and (2) a TrOCR model fine-tuned from a pre-trained handwritten text recognition checkpoint. Both approaches are evaluated using the same page-level corpus partition and evaluation protocol, with baseline and augmented training considered for each pipeline. This setup provides a direct comparison between from-scratch and transfer-learning approaches under low-resource conditions and allows the effect of data augmentation on recognition error to be assessed.

The contributions of this work are threefold. First, we construct a line-level experimental corpus for Chagatai handwritten text recognition from annotated materials in the Jarring Collection and define a page-level partitioning and evaluation protocol. Second, we compare two learning paradigms: a Kraken-based CRNN trained from scratch and a pretrained Transformer-based TrOCR model fine-tuned using the same page-level data partition. Third, we evaluate both pipelines with and without data augmentation using CER and WER and further analyze their training behavior and representative recognition errors.

The results contribute to a better understanding of the relationship between transfer learning and data augmentation in low-resource handwritten text recognition. Beyond the specific case of Chagatai, the proposed evaluation framework may also serve as a reference for future studies on other historical and low-resource manuscript collections. From a practical perspective, the evaluated models provide a basis for semi-automated transcription of archival Chagatai manuscripts and may reduce the manual effort required for subsequent digitization and analysis.

Literature Review

Handwritten Text Recognition (HTR) has become an important tool for the digitization and analysis of historical manuscripts. The growing use of HTR for improving access to historical sources and supporting large-scale textual analysis has been described (Nockels et al., 2024). Recent reviews trace the development of HTR from template-based methods to deep learning systems and identify document degradation as a continuing source of recognition errors (AlKendi et al., 2024). Progress in HTR has been strongest for languages with large annotated datasets. Historical scripts with limited digital resources remain more difficult to recognize,

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

particularly within the Perso-Arabic writing tradition, where handwriting variation and limited training data complicate model development. These constraints were examined in the context of few-shot HTR, demonstrating the need for methods that can operate with limited annotated material (Souibgui et al., 2022).

Research on Turkic and Arabic-script handwriting has produced several datasets that support model development and evaluation. Nurseitov et al. (2021) introduced the HKR database for Kazakh and Russian handwriting recognition. Toiganbayeva et al. (2022) developed the KOHTD dataset for Kazakh handwriting and included samples from thousands of examination papers. For Uyghur, Simayi et al. (2020) presented an unconstrained online word-recognition dataset addressing the connected structure of right-to-left Arabic-based writing. Bouchal et al. (2025) introduced the KALIMA dataset for historical Arabic manuscripts, with attention to stylistic variation and differences in manuscript conditions. Saeed et al. (2024) developed the Muharaf dataset, which contains more than 36,000 historical Arabic text lines. Majeed and Hassani (2024) proposed the Syriac KHAMIS dataset. Earlier resources such as the VML-HD corpus (Kassis et al., 2017) and the RASM2018 scientific manuscript collection (Clausner et al., 2018) have also been used to evaluate handwritten and historical document recognition.

These resources provide useful training and evaluation material, but their linguistic and paleographic characteristics differ from those of historical Chagatai manuscripts. Modern Kazakh and Uyghur datasets do not represent the full range of handwriting variation found in historical Chagatai sources. Historical Arabic and Syriac datasets address some related script characteristics but concern different languages, writing conventions, and manuscript traditions. This distinction is relevant when assessing whether recognition methods developed for related scripts can be transferred to Chagatai.

Model development for Arabic-based handwritten scripts has increasingly moved toward end-to-end and segmentation-free recognition. The transition from character-level processing toward architectures that recognize complete text sequences has been described (Tuama et al., 2025). It has been reported that end-to-end recognition can handle positional character variation and ligature formation in Arabic handwriting (Aabed et al., 2024). Transformer-based architectures have further expanded this line of research. The TrOCR model, which combines a vision Transformer encoder with a pretrained language-model decoder, was recently introduced (Li et al., 2023). Its pretrained representations allow the model to be adapted to new recognition tasks with less task-specific training data than required when training a model entirely from scratch. Other studies have proposed architectures for specific Arabic-script recognition problems. HATFormer (Chan et al., 2024) and the ET-Network (Hamza et al., 2024) address long-range dependencies in historical Arabic and Urdu text, respectively. A hybrid architecture combining Vision Transformers with a BERT-based decoder has been investigated (Ganai et al., 2023). The MMU-OCR-21 benchmark for Urdu text recognition was established in earlier work (Nasir et al., 2021).

Data augmentation represents another line of research for recognition under

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

limited training-data conditions. Generative Adversarial Networks for augmenting Kazakh manuscript data have been investigated (Yeleussinov et al., 2023). A stroke-based augmentation that modifies character structure while retaining properties of historical calligraphy was proposed (Ayyoob et al., 2024). Such methods differ from conventional geometric and photometric transformations and provide additional ways to increase training-data variation. Recent work has also examined the use of large multimodal models for historical and low-resource scripts. Cases in which multimodal models generated linguistically plausible but incorrect transcriptions for Kazakh Arabic script have been reported (Gagnier et al., 2026). These findings indicate that high-capacity generative models do not remove the need for domain-specific evaluation and annotated reference data. Few studies directly compare a CRNN system, such as Kraken (Kiessling, 2019), trained from scratch with a pretrained Transformer-based HTR system on the same historical Chagatai corpus. The existing literature provides datasets and recognition methods for several Turkic, Arabic, and other Arabic-script languages, but historical Chagatai remains underrepresented in systematic HTR evaluation. The available resources differ from Chagatai in language, script use, manuscript provenance, or writing characteristics. Few studies directly compare a CRNN system trained from scratch with a pretrained Transformer-based HTR system on the same historical Chagatai corpus. The role of image augmentation under this setting is also insufficiently documented. This study addresses these gaps by evaluating a Kraken CRNN trained from scratch and a pretrained TrOCR model fine-tuned on Chagatai manuscript data. The models are evaluated under a common page-level corpus partition, with and without augmented training data, using CER and WER as the primary recognition metrics.

Materials and methods.

The experiments used a corpus of historical Chagatai manuscripts from the Jarring Collection at Lund University Library. The corpus contains five manuscripts identified as Prov. 2, Prov. 3, Prov. 4, Prov. 8, and Prov. 24. Ground-truth transcriptions were based on the annotated materials of the Annotated Turki Manuscripts from the Jarring Collection Online (ATMO-1 and ATMO-2) projects. The selected manuscript pages were processed in eScriptorium, where text-line regions and their corresponding transcriptions were prepared and manually checked. The resulting annotations were exported together with the coordinates of the line regions. These coordinates were subsequently parsed programmatically and used to crop the corresponding line images from the manuscript pages, producing the final line-level image–text pairs used in the HTR experiments. The resulting corpus contains 2,531 line-level image–text pairs from 230 manuscript pages. The pages represent five Chagatai manuscripts from the Jarring Collection. Line images were exported in PNG format with the corresponding UTF-8 transcription files. The distribution of pages and line images is presented in Table 1.

The corpus was partitioned at the manuscript-page level. This procedure prevented line images from the same page from being distributed across different subsets. All lines extracted from a given page were assigned to the same subset. The

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

dataset was divided into training, validation, and test subsets using an 80/10/10 page-level split. The training subset contained 184 pages, the validation subset contained 23 pages, and the test subset contained 23 pages. The split used a fixed random seed of 42. The validation subset was used for training monitoring and checkpoint selection. The test subset remained held out until the final evaluation. The experiments used the exported line images without manual image preprocessing. No binarization, denoising, or contrast adjustment was applied. The ground-truth transcriptions were also retained in their original form. No character-level normalization, kashida removal, or ligature standardization was performed. The original annotation format was preserved during training and evaluation.

Table 1. Corpus composition

Manuscript	Pages	Line images
Prov. 2	86	820
Prov. 3	10	77
Prov. 4	25	227
Prov. 8	56	705
Prov. 24	53	702
Total	230	2531

The page-level partition ensures that text lines from the same manuscript page are not distributed across the training, validation, and test subsets, which helps prevent direct data leakage between the subsets. However, the split is not manuscript-independent, since pages from the same manuscript may still appear in different subsets. Therefore, the test results reflect recognition performance on unseen pages from the manuscripts included in the dataset rather than generalization to entirely unseen manuscripts or scribal traditions.

Two HTR systems were evaluated. The first system used Kraken with a convolutional recurrent neural network (CRNN) architecture. The model was trained from scratch using the Chagatai training data. Its architecture is presented in Table 2. The second system used TrOCR with the pretrained microsoft/trocr-base-handwritten checkpoint from the Hugging Face Transformers library. The two systems used different output representations. Kraken used an alphabet of 73 unique characters. The CTC blank symbol increased the output space to 74 classes. TrOCR used the XLM-RoBERTa tokenizer. Chagatai-specific characters absent from the original vocabulary were added to the tokenizer. The decoder embedding matrix was resized after vocabulary expansion. No frequency threshold was applied when adding the new characters.

Kraken was trained with the ketos training engine and a batch size of 16. Training used early stopping with a patience of 10 epochs based on validation CER. TrOCR was fine-tuned for 60 epochs with a batch size of 32. The baseline configuration used a learning rate of 2×10^{-4}. The augmented-data configuration used a learning rate of 1×10^{-4}. TrOCR training used a cosine learning-rate scheduler with a warm-up ratio of 0.1 and a weight decay of 0.01. The Kraken and TrOCR experiments followed

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

separate training procedures. Kraken used ketos-based early stopping, whereas TrOCR used a fixed 60-epoch schedule. The two frameworks also used different batch sizes and optimization settings. The comparison therefore concerns final recognition performance under the specified model configurations rather than identical optimization dynamics. The lower learning rate in the augmented TrOCR experiment was part of the experimental configuration. Therefore, the difference between the baseline and augmented TrOCR results reflects the combined effect of the augmented training data and the corresponding learning-rate setting. This factor is considered when interpreting the augmentation results. All experiments were conducted on an NVIDIA H200 GPU. The software environment consisted of Python 3.11, PyTorch 2.9.0, Transformers 5.14.1, Kraken 6.0.3, and Albumentations 2.0.8. The main training parameters are summarized in Table 3.

Table 2. Kraken CRNN architecture

Layer	Type	Filters / hidden units	Parameters
C_0	ActConv2D	32 filters	1.3K
C_3	ActConv2D	32 filters	40.0K
C_6	ActConv2D	64 filters	55.4K
C_9	ActConv2D	64 filters	110K
L_12	TransposedSummarizingRNN	400 units	1.9M
L_14	TransposedSummarizingRNN	400 units	963K
L_16	TransposedSummarizingRNN	400 units	963K
O_18	LinSoftmax	74 classes	29.7K

Table 3. Training parameters

Parameter	Kraken	TrOCR
Framework	Kraken Ketos	Hugging Face Transformers
Training strategy	From scratch	Fine-tuning
Backbone	CRNN	microsoft/trocr-base-handwritten
Batch size	16	32
Training schedule	Early stopping (patience=10)	60 epochs
Learning rate	Default Ketos configuration	2×10^{-4}
Augmented training LR	-	1×10^{-4}
Scheduler	-	Cosine
Warm-up ratio	-	0.1
Weight decay	-	0.01
GPU	NVIDIA H200	NVIDIA H200

Two image augmentation pipelines were implemented with Albumentations. The Kraken pipeline used small-angle rotation, affine shear, elastic deformation, hue and saturation adjustment, brightness and contrast adjustment, Gaussian noise, and Gaussian blur. The TrOCR pipeline used rotation, affine transformation, brightness and contrast adjustment, ISO noise, and either Gaussian blur or motion blur. Two augmented samples were generated from each original training image. Augmented

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

images were used only during training. The validation and test subsets contained original images only. The augmentation parameters are listed in Table 4.

Table 4. Data augmentation settings

Transformation	Kraken	TrOCR
Rotation	±1°	±3°
Affine shear	x,y ±2°	x ±3°, y ±1°
ElasticTransform	α=20, σ=6	-
Hue/Saturation	Yes	-
Brightness/Contrast	±0.2	±0.15
Gaussian noise	Yes	-
ISO noise	-	Yes
Gaussian blur	σ 0.2–0.5	Yes
Motion blur	-	Yes
Synthetic images per sample	2	2

The two pipelines used different transformation sets and intensities. These settings were defined separately for the two model configurations and were not treated as a controlled ablation of individual augmentation operations. The augmentation experiments therefore measure the effect of the complete augmentation configuration used for each model.

Model performance was evaluated using Character Error Rate (CER) and Word Error Rate (WER). Predictions were decoded after removing special tokens. CER and WER were calculated with the Hugging Face Evaluate library. Validation metrics were recorded during training. For model selection, the checkpoint with the lowest validation CER was selected for each experiment. The selected checkpoint was then evaluated on the held-out test set. The test set was not used for training, checkpoint selection, or other training decisions. Training behaviour was examined through the validation metrics recorded across epochs. For Kraken, validation character accuracy was used to describe the training trajectory. For TrOCR, validation CER and WER were examined across epochs. These metrics were analyzed within each model family. They were not used as a common quantitative measure for direct comparison of training trajectories.

The accuracy of the handwritten text recognition model is evaluated using the Character Error Rate (CER), calculated according to formula (1):

$$CER = \frac{(S+D+I)}{N}, \tag{1}$$

where S is the number of character substitutions;

D is the number of character deletions;

I is the number of character insertions;

N is the total number of characters in the reference text (Ground Truth).

Similarly, the word-level recognition accuracy is evaluated using the Word Error Rate (WER), calculated according to formula (2):

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

$$WER = \frac{(S_w + D_w + I_w)}{N_w}, \quad (2)$$

where S_w is the number of word substitutions;
D_w is the number of word deletions;
I_w is the number of word insertions;
N_w is the total number of words in the reference text (Ground Truth).

Results.

The recognition performance of the two handwritten text recognition (HTR) pipelines was evaluated under four experimental configurations: Kraken trained from scratch, Kraken with data augmentation, TrOCR fine-tuned from a pretrained checkpoint, and TrOCR fine-tuned using augmented training data. All configurations were evaluated on the same held-out test set defined by the page-level 80/10/10 partition described in Section 3. Character Error Rate (CER) and Word Error Rate (WER) were used as the primary evaluation metrics. The best-performing checkpoint for each experiment was selected according to validation CER and subsequently evaluated on the test set. The final recognition results are presented in Table 5.

Table 5. Recognition performance of the evaluated HTR models

Model	Training strategy	CER (%)	WER (%)
Kraken	Baseline	18.15	55.44
Kraken	Data augmentation	15.54	49.67
TrOCR	Baseline	13.42	40.61
TrOCR	Data augmentation	10.6	35.59

The best overall performance was obtained by TrOCR fine-tuned with augmented training data, achieving a CER of 10.60% and a WER of 35.59%. The Kraken baseline produced the highest error rates, with a CER of 18.15% and a WER of 55.44%. Thus, among the four evaluated configurations, the pretrained transformer-based model combined with data augmentation achieved the lowest recognition error at both the character and word levels.

Lower recognition errors were observed for the augmented configurations in both pipelines. For Kraken, data augmentation reduced CER from 18.15% to 15.54% and WER from 55.44% to 49.67%, corresponding to relative reductions of approximately 14.4% and 10.4%, respectively. For TrOCR, the augmented configuration achieved a CER of 10.60% and a WER of 35.59%, compared with 13.42% and 40.61% for the baseline configuration. These differences correspond to relative reductions of approximately 21.0% in CER and 12.4% in WER. However, the baseline and augmented TrOCR configurations used different learning rates. The observed difference for TrOCR should therefore be interpreted as a comparison between the two training configurations rather than as an isolated effect of data augmentation.

The results also demonstrate a clear difference between training a recognition model from scratch and adapting a pretrained model. The baseline TrOCR configuration

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

achieved a CER of 13.42% and a WER of 40.61%, outperforming the augmented Kraken configuration, which achieved 15.54% CER and 49.67% WER. Thus, even without augmentation, the pretrained transformer-based model performed better than the CRNN model trained from scratch with additional synthetic samples.

The difference became more pronounced when augmentation was combined with transfer learning. The augmented TrOCR configuration achieved the best results in the study, with a CER of 10.60% and a WER of 35.59%. Relative to the Kraken baseline, this corresponds to reductions of approximately 41.6% in CER and 35.8% in WER.

These results indicate that, under the experimental conditions considered in this study, transfer learning provided a stronger starting point than training the CRNN architecture exclusively on the available Chagatai corpus. Data augmentation further improved both approaches, with the combination of pretraining and augmentation producing the lowest observed recognition error.

Training dynamics

In addition to final test-set performance, validation metrics were examined to characterize the training behaviour of the evaluated configurations. Because the two frameworks use different optimization procedures and expose different training metrics, the validation trajectories were analyzed separately for Kraken and TrOCR rather than treated as directly comparable measures.

For Kraken, validation character accuracy was monitored during training. The augmented configuration exhibited a different validation trajectory from the baseline and reached higher validation performance during training. The additional geometric and photometric variability introduced by the augmentation pipeline therefore affected the optimization behaviour of the CRNN model in addition to improving its final test-set performance.

For TrOCR, validation CER and WER were monitored throughout fine-tuning. Both baseline and augmented configurations showed an overall reduction in recognition errors during training. The augmented configuration achieved lower final test-set CER and WER while maintaining a broadly similar validation trajectory to the baseline configuration. No pronounced instability was observed in the validation curves associated with augmentation.

Figure 1 presents the validation character-accuracy curves for the Kraken experiments, while Figures 2 and 3 show the validation CER and WER curves for the TrOCR experiments, respectively.

The qualitative examples reveal substantial differences in the types and magnitude of recognition errors produced by the two model families. The Kraken baseline frequently generated long sequences of erroneous characters and diacritics, particularly in visually complex lines. In several examples, the output diverged substantially from the reference transcription, indicating difficulties in reconstructing the character sequence from the handwritten input.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

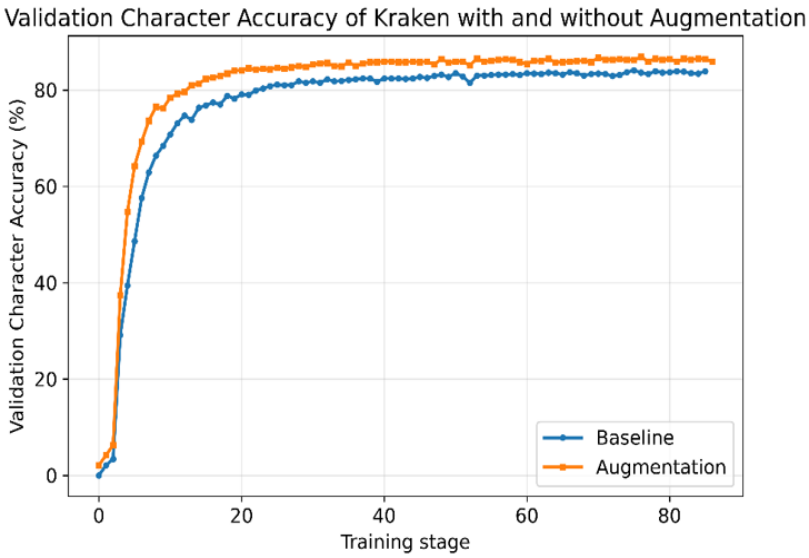


Fig. 1. Validation character accuracy (%) of the Kraken model with and without data augmentation during training

Figure 2 presents the CER curves of the TrOCR models.

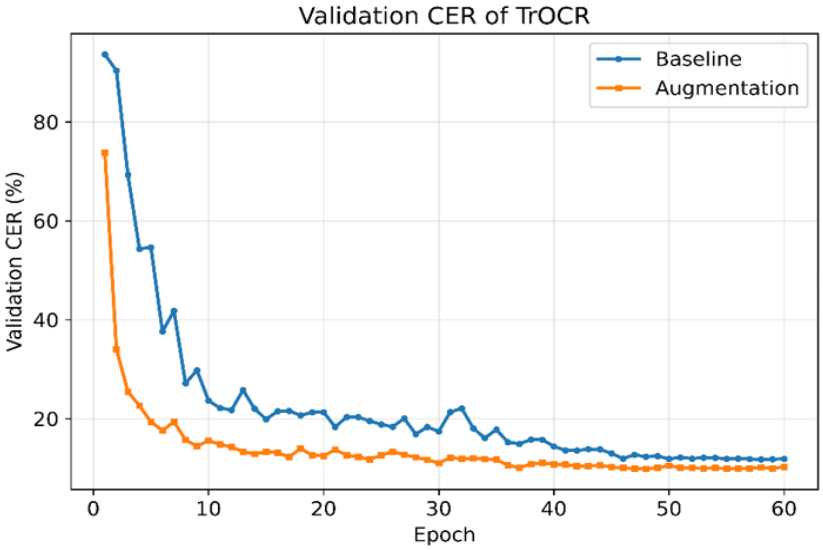


Fig. 2. Character Error Rate (CER) of the TrOCR model with and without data augmentation across training epochs.

Importantly, the validation curves were used only to characterize training behaviour and select model checkpoints according to the experimental protocol. The held-out test set was not used for training, checkpoint selection, or any other optimization decision. Consequently, the quantitative comparison between the four configurations is based on the final test-set CER and WER values reported in Table 5.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Qualitative Error Analysis

Figure 3 compares the WER obtained during TrOCR training.

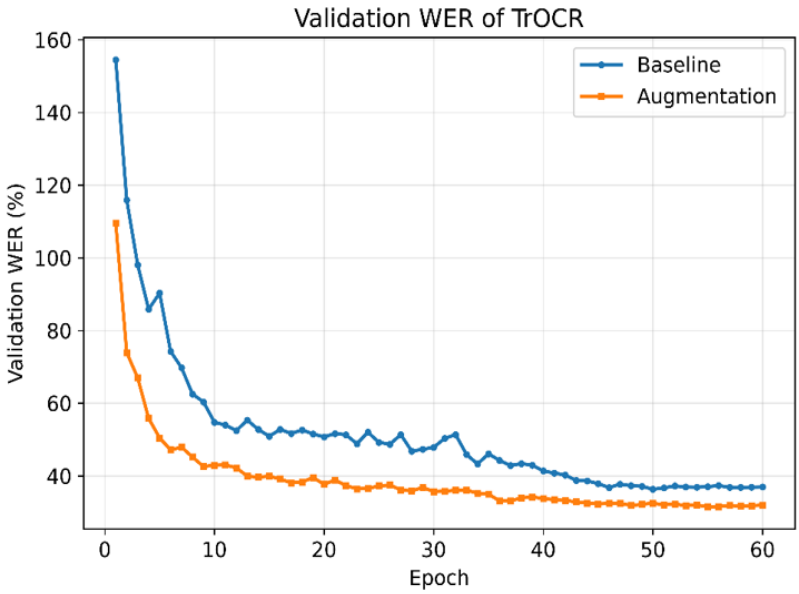


Fig. 3. Word Error Rate (WER) of the TrOCR model with and without data augmentation across training epochs.

Figure 4 summarizes the final recognition performance of all evaluated models.

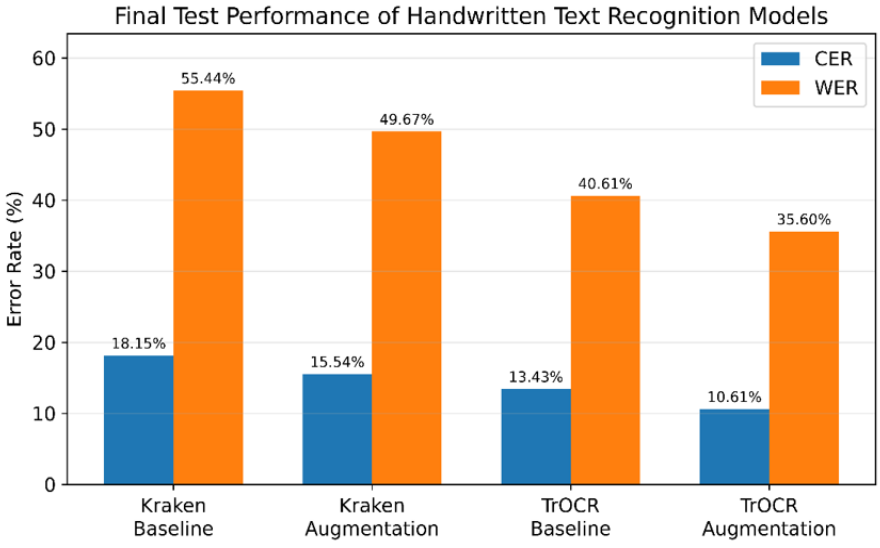


Fig. 4. Final test-set comparison of HTR performance for the Kraken and TrOCR models under baseline and augmentation settings using CER and WER.

The augmented Kraken model generally produced more structured outputs than the baseline model in several examples. Nevertheless, substantial character-level

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

errors remained, particularly for lines containing visually complex character forms and diacritic patterns. This observation is consistent with the reduction in aggregate CER and WER reported in Table 5, while also showing that augmentation did not eliminate the recognition difficulties associated with complex manuscript regions. The TrOCR outputs were generally closer to the reference transcriptions and preserved a larger proportion of the original word structure. The augmented TrOCR configuration further improved several representative cases, producing outputs with fewer character substitutions and closer correspondence to individual words in the ground truth. For example, in the selected samples, the augmented model frequently retained the overall lexical structure even when individual characters were incorrectly recognized. However, the effect of augmentation was not uniform across all examples. Some lines showed only limited improvement, while minor differences or additional recognition errors were observed in individual cases. This variability indicates that the effectiveness of augmentation depends partly on the visual characteristics of the manuscript line, including handwriting variation, character forms, and the presence of complex or ambiguous glyphs. Overall, the qualitative analysis is consistent with the quantitative results. Kraken benefited from augmentation but continued to exhibit substantial character-level errors, whereas TrOCR produced more accurate and structurally coherent transcriptions under both training conditions. The augmented TrOCR model provided the strongest performance among the evaluated configurations, although individual recognition errors remained present even in comparatively successful examples.

Discussion.

The results demonstrate that both transfer learning and data augmentation contributed to improved handwritten text recognition of the Chagatai manuscripts, although their effects differed across the evaluated recognition paradigms. The strongest performance was achieved by TrOCR with data augmentation, which obtained the lowest CER and WER among all experimental configurations. Data augmentation also improved the Kraken model, indicating that increasing the visual diversity of the training data was beneficial for both from-scratch training and transfer learning.

To complement the aggregate CER and WER results, a qualitative analysis was conducted using representative manuscript lines with different levels of visual and recognition difficulty. Table 6 presents the ground-truth transcriptions alongside the outputs generated by the four evaluated configurations.

The advantage of TrOCR is particularly relevant in the context of the limited size of the available corpus. The experiments were conducted on 2531 annotated text lines from 230 manuscript pages, which provides only a limited amount of domain-specific material for training a recognition model from scratch. In contrast, the TrOCR model was initialized from a pretrained handwritten text recognition checkpoint, allowing the model to start from representations acquired before adaptation to the Chagatai data.

Table 6. Qualitative comparison of recognition outputs for selected manuscript lines

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ground_truth	kraken_baseline	kraken_augmented	trocr_baseline	trocr_augmented
اکر کشی باشینی توشوروب توشاسه حم خوب ایمس	ببهتب هر نپ به یح ه ه بوبوباتببر برتوبوبب ب بن پنپپپ په قیر نیه هببب حببپ هرت بت ت تتهوبترن بنزرتب تری برا بیر ورکر ببالبومتبکر لم	حیجیب یی حیجیب ج به بل لو تت ا انسحم حی بب جیینا مشی نی ی نجچی جیه نم چقه حی یحیحبچینم چی تویتبه و بسی بای محوحو خیج بر یه هی الا ی اسیم وا ی	اکر کشی باشینی توشوروب تاش سهم حد ب اعس	اکر کشی باشینی توشواوب ترشاب سه هم خب ایس
یخشی دور هرکیم توشیده تاغلارده یا بلند یرلارده تیوه کا	رریکم شی جا اجه ار دی یفلمیرولادق و یکل هرحت وررب ر ن ببب جبببببابا بهبد	ور هرکیم توشیده تاغیلارده یا لنه یرکارنه تیوبکا مششی دررو هرکم بو مبمخالیای مث	یخشی دور هرکیم توشیده تاغلاردهء یا بلندی یرلکاری تیوه کا	یخشی دور هرکیم توشیده تاغلاردهء یا بلندیرله یتیوه کا
شفا تابغای انشاء الله. و ینه هرکیم ساسماق یرنی عادت	اض و قریی تاتتتببل انهء کجیب ببنففنا حانایم انناهاومصم رن بن تجج جت بببیبه حب بالهه	بن قی ار یحی قاوبال یب بجید بیاجنی احزا الّمو لمرا ححتحو حین ب ی ی ن ی ی ا ا ل	شفا تاپغای انشأالله و ینه هرکیم ساسماق یرنی عادت	شفا تابغای انشأ الله و ینه هرکیم ساسماق یرنی عاوت
تبه نی سیقیب ایچورسه دفع بولغای هرکیم میمون کوشتی نی شکر	بححب ککم ا ل اکتت للبربم لرس پیینری یب نهقو ناحلب ببیل ب بببب بببا ووترحمن بابببب ببنیب رملا رب باراز به بنه	هم لیه بححه یبت لولور حملا جحلحم چی هی یلا ینی اا ب ب قبهتنی سیقین ایچپورسیه درفع تاپ م چقاتیب ای بر تیم لی قوی بن ن ج یباه	تبه نی سیقیب ایچورسه دفع بولغای هرکیم مین کوشتی نشکر	تبه نی سیقیب ایچورسه دفع بولغای هرکیم میون کوشتی نفکر
پاکیزه جامعه پیتیب یاغده یوب اول یاغقه سورۂ	قح ب هربرب التل پر توروما ب بب ب ربی ارا بوربتبه ابا ببکببتا تولاوتح لتلتا ماا ره بنیب احبب به ارابب بباببلب ح اله ا اببیجتبببپا رجبب م اعهجه بب رحهدنه بابه نجو جر یب حتالی جررت و انی برررر پاهببسه سبه ابببب بحب یبه	هیا یا ول للل ب الوری بلج جیلا ینهلم وبنی بل بح نببلا جلرق لووب بلا حوما من نجم فیتی جرمل ت وحوا وا لاا ا هینه ین جی قت نجی ینج حی حچهه هت یلی حه دی حلتحه جی نیه اباکم بر جا رنح مملل حو ه بیبیا بیب به	پاکیزه جامعه پتیب یاغنده یوب اول غقه سورۂ	پاکیزه جامعه پتیب یاغذه یوب اول یلاغقه سورۂ
کوچله نی جمع قیلیب سورکیل درحال شفا بولور و دیگر کشی نینک هر	ابتاحبببب و کیر کشتی ینیکیم جرت ب بپهاپپءیجه رلا پر جه نناربتبرکیه که هنبببب یتب بلا ه بتبلب اه پییبرر ییاح ابو ورب رو حبلابن اروابرالی ینی با ارابالبم بی یتب امهببربی	ملخحو بلا و دکر کشی نینک کنیه بیبلا چ هرنه ینیجی خیبچبای ح لی ویبا ی هن ی جبن بلی لب بالا بیل بر عحال شقا بولوروده قیره وحونبحححل قیا بل لتا مروج یوهممت بحیبنهی	کوچله نی جمع قیلیب سورکیل درحال شفا بولور و یگر کشی نک تینک	کوچیله نی جمع قیلیب سورکیل درحال شفا بولور و یگر کشی دینک هر

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

داىم الوقاتا ستعمال قىلسا الصا تىشى اغرىماغاى عمرى ايچى	ررا ا مَاوَتا با كَوحَمْتَرَمَرَبىَرْنا رَرا ْتا تَنْ جبْيْنى ا ح ْ تَحنْن تْنا بببْبىبِهِ اْجَهَتَبَ به اپهنى تپه پ ج رَاره بتَبَهبُجبىبَبْرَاتَىَ ل بمبى حاابقهَ پم ر ج ببببُرَهپىَبا بَرْ ب	جولُحَلْلُحُحالُ كَبْحُ كلْمُ ا ُحَعْا ى مَتْتَىَ حْى حا ا ُهُحُ ا ْحْ جمُ ىُخْحْ جىِ ههُمن هرمُىمه ىبى ىقىنى حمام ن منىن ر ت لبىه چ ا قحم حت بچ نچه هنه جقىتى حى حوى من	داىم الدهقات استامال قىلسا الصالتىشى اغرىماغاى عمرى	داىم الوقاتا ستعمال قىلسا الصا تىش اغرىماغاى عمرى ايچى

The lower error rates obtained by the pretrained TrOCR configurations compared with the Kraken configurations therefore suggest that transfer learning can provide an effective initialization strategy when annotated target-domain material is scarce.

Lower recognition error was also observed for the augmented configuration within each model family. For Kraken, where the principal experimental difference was the use of augmented training samples, CER decreased from 18.15% to 15.54% and WER from 55.44% to 49.67%. For TrOCR, CER decreased from 13.42% to 10.60% and WER from 40.61% to 35.59%. However, the TrOCR configurations also differed in learning rate, which prevents the effect of augmentation from being isolated. Therefore, the improvement observed for TrOCR cannot be attributed to data augmentation alone.

The comparison between the two approaches should nevertheless be interpreted carefully. The experimental design does not isolate model architecture from initialization strategy, because Kraken was trained from scratch whereas TrOCR was initialized from a pretrained checkpoint. Consequently, the observed superiority of TrOCR cannot be attributed to the Transformer architecture alone. The results instead demonstrate the effectiveness of the complete Transformer-based transfer-learning configuration evaluated in this study relative to the CRNN configuration trained from scratch. A more controlled experimental design, in which comparable initialization and optimization conditions are applied across architectures, would be required to independently quantify the contributions of architecture and pretraining.

The findings are particularly relevant to HTR for historical and low-resource manuscript collections. Chagatai manuscripts combine several challenging characteristics, including limited annotated resources, variation in handwriting, and the structural complexity of Arabic-based writing. The results suggest that pretrained HTR models can provide a practical starting point for adapting recognition systems to such collections. At the same time, the improvements obtained through augmentation indicate that increasing the visual diversity of a limited training corpus remains valuable even when transfer learning is employed.

The qualitative analysis provides additional insight into the observed quantitative differences. The Kraken configurations frequently produced extensive character-level errors and sequences of incorrectly recognized characters, particularly in visually complex lines. Augmentation reduced these errors in some cases but did not eliminate

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

them. TrOCR generally preserved a greater proportion of word structure and produced outputs that were closer to the reference transcriptions. The augmented TrOCR model further improved several difficult examples, although the effect of augmentation was not uniform across all manuscript lines. This variability suggests that augmentation may be particularly useful for some types of handwriting and image variation but may not address all sources of recognition error.

Several limitations should be considered when interpreting these findings. First, the corpus consists of 230 manuscript pages and 2531 annotated text lines from five manuscripts within the Jarring Collection. Although the corpus provides a useful experimental basis for Chagatai HTR, it represents only a limited portion of the broader variation found in historical Chagatai manuscripts. In particular, the range of scribal styles and manuscript conditions is restricted.

Second, the evaluation material originated from the same collection as the training data. Although page-level partitioning prevented lines from the same page from appearing in different subsets, the experiment does not establish how well the trained models generalize to manuscripts produced by different scribes or preserved in other collections. Cross-manuscript and cross-collection evaluation would therefore provide a stronger assessment of model robustness.

Third, the study evaluates two model configurations that differ simultaneously in architecture, initialization, and optimization strategy. Kraken was trained from scratch using the Ketos training framework, whereas TrOCR was fine-tuned from a pretrained checkpoint using a different optimization procedure. The comparison is therefore best understood as a comparison of two practical HTR pipelines rather than as a controlled architectural ablation.

Fourth, the augmentation experiments did not include an ablation study of individual transformations. The augmentation pipelines combined several geometric and photometric operations, and the present experiments therefore cannot determine which transformations contributed most strongly to the observed improvements. In addition, the TrOCR baseline and augmented configurations used different learning rates, which introduces an additional optimization-related factor that should be considered when interpreting the augmentation results.

Future work should address these limitations through broader and more controlled experiments. Evaluation on manuscripts from different collections and scribal traditions would allow the generalization capability of the models to be assessed more rigorously. Ablation experiments could determine the individual contribution of rotation, affine transformations, noise, blur, and other augmentation operations. Additional Transformer-based HTR architectures and parameter-efficient fine-tuning methods could also be evaluated under the same page-level partitioning and test protocol. Furthermore, controlled experiments using comparable initialization and optimization strategies would help separate the effects of architecture, pretraining, and data augmentation.

Overall, the findings indicate that pretrained Transformer-based HTR combined with data augmentation provided the strongest recognition performance among the

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

configurations evaluated on the Chagatai manuscript corpus. The results support transfer learning as a promising strategy for low-resource historical HTR, while also demonstrating that synthetic visual variation can provide additional improvements. More broadly, the study provides an empirical basis for further investigation of pretrained HTR models and data-efficient training strategies for historical manuscript collections where large expert-annotated corpora are not yet available.

Conclusion.

This study evaluated two handwritten text recognition paradigms for historical Chagatai manuscripts: a CRNN-based Kraken model trained from scratch and a Transformer-based TrOCR model fine-tuned from a pretrained handwritten text recognition checkpoint. The experiments were conducted on a page-level partition of a corpus comprising 230 manuscript pages and 2531 annotated text lines from five manuscripts in the Jarring Collection. The experimental results demonstrated that data augmentation improved recognition performance for both evaluated HTR pipelines. The best performance was achieved by the augmented TrOCR configuration, which obtained a CER of 10.60% and a WER of 35.59%. In comparison, the Kraken baseline achieved a CER of 18.15% and a WER of 55.44%. The results therefore indicate that the combination of pretrained initialization and synthetic visual variation was the most effective of the evaluated configurations for the investigated low-resource Chagatai corpus. The comparison further showed that the pretrained TrOCR configurations outperformed the Kraken configurations trained from scratch under the experimental conditions of this study. These findings support transfer learning as a promising strategy for HTR when annotated material for a historical language is limited. Data augmentation provides a complementary means of increasing the visual diversity of the training corpus and further reducing recognition errors. The results should nevertheless be interpreted within the scope of the experimental design. The comparison does not independently separate the effects of model architecture, pretrained initialization, and optimization strategy. Therefore, the observed advantage of TrOCR should be understood as the performance advantage of the evaluated Transformer-based transfer-learning pipeline rather than as evidence that the Transformer architecture itself is inherently superior to the CRNN architecture. The study is also limited by the size and provenance of the corpus, which represents five manuscripts from a single archival collection. Generalization across different collections, scribal styles, and manuscript conditions was not evaluated. In addition, the augmentation experiments did not include an ablation study of individual transformations. Future work should therefore investigate cross-manuscript and cross-collection generalization, expand the available Chagatai HTR corpus, systematically evaluate individual augmentation operations, and examine additional Transformer-based and multimodal recognition approaches under controlled experimental conditions. Overall, this study demonstrates the feasibility of applying pretrained HTR models to low-resource historical Chagatai manuscripts and establishes an experimentally evaluated benchmark for comparing from-scratch and transfer-learning approaches. The results provide a foundation for further development of automated recognition systems for Chagatai and other historical Arabic-script Turkic

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

manuscript collections, supporting their large-scale digitization and computational analysis.

REFERENCES

Aabed S., Khairaldin A. (2024). An End-to-End, Segmentation-Free, Arabic Handwritten Recognition Model on KHATT // arXiv. arXiv:2406.15329. [In Eng.].

AlKendi W., Gechter F., Heyberger L., Guyeux C. (2024). Advancements and challenges in handwritten text recognition: A comprehensive survey // Journal of Imaging. — Vol. 10. — No. 1. — Pp. 1–25. DOI: 10.3390/jimaging10010018. [In Eng.].

Ayyoob M.P., Ilyas P.M. (2024). Stroke-Based Data Augmentation for Enhancing Optical Character Recognition of Ancient Handwritten Scripts // IEEE Access. — Vol. 12. — Pp. 186794–186802. DOI: 10.1109/ACCESS.2024.3505238. [In Eng.].

Bouchal H., Belaid A., Meziane F. (2025). Towards Accurate Recognition of Historical Arabic Manuscripts: A Novel Dataset and a Generalizable Pipeline // ACM Transactions on Asian and Low-Resource Language Information Processing. — Vol. 24. — No. 10. Article 104. DOI: 10.1145/3744243. [In Eng.].

Chan A., Mijar A., Saeed M., Wong C.W., Khater A. (2024). HATFormer: Historic Handwritten Arabic Text Recognition with Transformers // arXiv. arXiv:2410.02179. [In Eng.].

Clausner C., Antonacopoulos A., McGregor N., Wilson-Nunn D. (2018). ICFHR 2018 Competition on Recognition of Historical Arabic Scientific Manuscripts–RASM2018 // Proceedings of the 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR). — Pp. 471–476. DOI: 10.1109/ICFHR-2018.2018.00088. [In Eng.].

Gagnier H., Gagnier S., Kirubakaran A. (2026). KazakhOCR: A Synthetic Benchmark for Evaluating Multimodal Models in Low-Resource Kazakh Script OCR // Proceedings of the 2nd Workshop on NLP for Languages Using Arabic Script. — Pp. 60–65. [In Eng.].

Ganai A.F., Khursheed F. (2023). Computationally efficient recognition of unconstrained handwritten Urdu script using BERT with vision transformers // Neural Computing and Applications. — Vol. 35. — No. 35. — Pp. 24161–24177. DOI: 10.1007/s00521-023-08976-1. [In Eng.].

Hamza A., Ren S., Saeed U. (2024). ET-Network: A novel efficient transformer deep learning model for automated Urdu handwritten text recognition // PLoS ONE. — Vol. 19. — No. 5. e0302590. DOI: 10.1371/journal.pone.0302590. [In Eng.].

Kassis M., Abdalhaleem A., Droby A., Alaasam R., El-Sana J. (2017). VML-HD: The historical Arabic documents dataset for recognition systems // Proceedings of the 2017 1st International Workshop on Arabic Script Analysis and Recognition (ASAR). — Pp. 11–14. DOI: 10.1109/ASAR.2017.8067751. [In Eng.].

Kiessling B. (2019). Kraken - an Universal Text Recognizer for the Humanities // DH 2019 Book of Abstracts. Utrecht. DOI: 10.34894/Z9G2EX. [In Eng.].

Li M., Lv T., Cui L., Lu Y., Florencio D., Zhang C., Li Z., Wei F. (2023). TrOCR: Transformer-based optical character recognition with pre-trained models // Proceedings of the AAAI Conference on Artificial Intelligence. — Vol. 37. — No. 11. —Pp. 13094–13102. DOI: 10.1609/aaai.v37i11.26538. [In Eng.].

Majeed A., Hassani H. (2024). Ancient but digitized: Developing handwritten optical character recognition for east syriac script through creating khamis dataset // arXiv. arXiv:2408.13631. [In Eng.].

Nasir T., Malik M.K., Shahzad K. (2021). MMU-OCR-21: Towards end-to-end Urdu text recognition using deep learning // IEEE Access. — Vol. 9. — Pp. 124945–124962. DOI: 10.1109/ACCESS.2021.3110787. [In Eng.].

Nockels J., Gooding P., Terras M. (2024). The implications of handwritten text recognition for accessing the past at scale // Journal of Documentation. — Vol. 80. — No. 7. — Pp. 148–167. DOI: 10.1108/JD-09-2023-0183. [In Eng.].

Nurseitov D., Bostanbekov K., Kurmankhojayev D., Alimova A., Abdallah A., Tolegenov R. (2021). Handwritten Kazakh and Russian (HKR) database for text recognition // Multimedia Tools and Applications. — Vol. 80. — No. 21-23. Pp. 33075–33097. DOI: 10.1007/s11042-021-11399-6. [In Eng.].

Saeed M., Chan A., Mijar A., Moukarzel J., Habchi G., Younes C., Elias A., Wong C.W., Khater A. (2024). Muharaf: Manuscripts of handwritten Arabic dataset for cursive text recognition // Advances in Neural Information Processing Systems (NeurIPS). — Vol. 37. — Pp. 58525–58538. [In Eng.].

Simayi W., Ibrahim M., Zhang X.Y., Liu C.L., Hamdulla A. (2020). A benchmark for unconstrained online handwritten Uyghur word recognition // International Journal on Document Analysis and Recognition (IJDAR). — Vol. 23. — No. 3. Pp. 205–218. DOI: 10.1007/s10032-020-00354-0. [In Eng.].

Souibgui M.A., Fornés A., Kessentini Y., Megyesi B. (2022). Few shots are all you need: A progressive

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

learning approach for low resource handwritten text recognition // Pattern Recognition Letters. — Vol. 160. — Pp. 43–49. DOI: 10.1016/j.patrec.2022.06.003. [In Eng.].

Toiganbayeva N., Kasem M., Abdimanap G., Bostanbekov K., Abdallah A., Alimova A., Nurseitov D. (2022). KOHTD: Kazakh Offline Handwritten Text Dataset // Signal Processing: Image Communication. — Vol. 108. — Article 116827. DOI: 10.1016/j.image.2022.116827. [In Eng.].

Tuama B., Mohamed F. (2025). A Systematic Literature Review of Deep Learning Methods for Handwritten Text Recognition in Historical Arabic Manuscripts // Engineering, Technology & Applied Science Research. — Vol. 15. — No. 1. Pp. 25772–25782. DOI: 10.48084/etasr.12123. [In Eng.].

Yeleussinov A., Amirgaliyev Y., Cherikbayeva L. (2023). Improving OCR Accuracy for Kazakh Handwriting Recognition Using GAN Models // Applied Sciences. — Vol. 13. — No. 9. Article 5677. — DOI: 10.3390/app13095677. [In Eng.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 219–240
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.012

A NEURO-DIGITAL LEARNING MODEL FOR AI DRIVEN DIGITAL TRANSFORMATION IN HIGHER EDUCATION

D.B. Baumuratova[1,2,3], T.K. Zhukabayeva[1,2,3], K.A. Abdrakhmanov[2,3], Z. Ahmad[2,4], A.A. Seitov[5]*

[1]L.N. Gumilyov Eurasian National University, Astana, Kazakhstan;
[2]International Science Complex Astana, Astana, Kazakhstan;
[3]Astana International University, Astana, Kazakhstan;
[4]Hazara University, Mansehra, Pakistan;
[5]M. Auezov South Kazakhstan University, Shymkent, Kazakhstan.
E-mail: seitovaskar853@gmail.com

Dilaram B. Baumuratova — PhD, Senior Lecturer, Pedagogical Institute, Astana International University, Astana, Kazakhstan
https://orcid.org/0009-0009-4621-1886;
Tamara K. Zhukabayeva — PhD, Professor at L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
https://orcid.org/0000-0001-6345-5211;
Kairat A. Abdrakhmanov — PhD, Vice President for International Cooperation and Student Affairs, Astana International University, Astana, Kazakhstan
https://orcid.org/0000-0002-7387-7191;
Zulfiqar Ahmad — PhD, Department of Computer Science and Information Technology Hazara University, Mansehra, Pakistan
https://orcid.org/0000-0002-6278-9516;
Askar A. Seitov — PhD Student, M. Auezov South Kazakhstan University, Shymkent, Kazakhstan
E-mail: seitovaskar853@gmail.com, https://orcid.org/0009-0009-7539-9806.

© D.B. Baumuratova, T.K. Zhukabayeva K.A. Abdrakhmanov, Zulfiqar Ahmad, A.A. Seitov

Abstract. The rapid development of artificial intelligence (AI) is transforming higher education from a traditional mode of information distribution to an intelligent, adaptive, and self-renewing system. In this paper, we analyze the evolution of paradigms of education in the so-called “presingularity era” when AI has attained a high level of capability and when final decisions are still made by humans. The goal is to construct a digital learning model for higher education using the theoretical bases of neuroeducation, brain-based learning and AI-based cognition. The methodology of

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

the work is based on a methodical analysis of scientific publications, international strategic documents and domestic experience, integrated with a qualitative approach and grounded Theory. The analysis considers the evolution of educational models from "Education 1.0" to "Education 6.0", emphasizing a transition from a teacher-centered system to an anthropo-digital synergy where the learner is a subject of coevolution and the teacher acts as a cognitive partner in collaboration with AI. This study also provides insights into the relationships among neuroeducation, brain-based learning and AI-based cognition. Cognitive load theory is used to explain how AI reduces the intrinsic load, limits the extraneous load and increases the germane load. The risks of AI cognition, such as weakening cognitive activity, fragmented attention and AI dependence, are also identified. The suggested paradigm combines neuroplasticity, brain-based learning, cognitive load optimization and AI cognition into a single model in which AI supports human thinking but does not replace it. The results have theoretical and practical consequences for educational science and the digital transformation of higher and postgraduate education in Kazakhstan (Smart.EDU).

Keywords: technological singularity, neuro-digital learning, neuroeducation, brain-based learning, AI cognition, cognitive load theory

For citations: Baumuratova D.B., Zhukabayeva T.K., Abdrakhmanov K.A., Ahmad Z., Seitov A.A. (2026). A Neuro-Digital Learning Model for AI Driven Digital Transformation in Higher Education // International journal of information and communication technologies. Vol. 7. No. 27. Pp. 219–240. https://doi.org/10.54309/IJICT.2026.27.3.012. (In Eng.).

Conflict of interest: The authors declare that there are no conflicts of interest.

Acknowledgment. *This research was funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant №BR28713531 Intelligent digital system of higher and postgraduate education organizations Smart.EDU).*

ЖОҒАРЫ БІЛІМ БЕРУДЕ AI НЕГІЗДЕЛГЕН ЦИФРЛЫҚ ТРАНСФОРМАЦИЯҒА АРНАЛҒАН НЕЙРО-ЦИФРЛЫҚ ОҚЫТУ МОДЕЛІ

Д.Б. Баумуратова[1,2,3], Т.К. Жукабаева[1,2,3], К.А. Абдрахманов[2,3], З. Ахмад[2,4], А.А. Сеитов[5]*

[1]Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан;
[2]Астана халықаралық ғылыми кешені, Астана, Қазақстан;
[3]Астана Халықаралық университеті, Астана, Қазақстан;
[4]Хазара университеті, Мансехра, Пәкістан;
[5]М.Әуезов атындағы Оңтүстік Қазақстан университеті, Шымкент, Қазақстан.
E-mail: seitovaskar853@gmail.com

Баумуратова Диларам Бекбулатовна — PhD, Астана халықаралық университетінің Педагогикалық институтының аға оқытушысы, Астана, Қазақстан
https://orcid.org/0009-0009-4621-1886;

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Жукабаева Тамара Кокеновна — PhD, Л.Н. Гумилев атындағы Еуразия ұлттық университетінің профессоры, Астана, Қазақстан
https://orcid.org/0000-0001-6345-5211;
Абдрахманов Қайрат Амангельдинович — PhD, Астана халықаралық университетінің халықаралық ынтымақтастық және студенттер ісі жөніндегі вице-президенті, Астана, Қазақстан
https://orcid.org/0000-0002-7387-7191;
Зулфиқар Ахмад — PhD, Хазара университетінің компьютерлік ғылымдар және ақпараттық технологиялар кафедрасы, Мансехра, Пәкістан
https://orcid.org/0000-0002-6278-9516;
Сеитов Аскар Абилханович — докторант, М.Әуезов атындағы Оңтүстік Қазақстан университеті, Шымкент, Қазақстан
E-mail: seitovaskar853@gmail.com, https://orcid.org/0009-0009-7539-9806.

© Д.Б. Баумуратова, Т.К. Жукабаева, К.А. Абдрахманов, З. Ахмад, А.А. Сеитов

Аннотация. Мақалада жасанды интеллекттің (AI) қарқынды дамуы жоғары білім беруді ақпаратты таратудың дәстүрлі режимінен интеллектуалды, бейімделгіш және өзін-өзі жаңартатын жүйеге айналдыруда. Бұл мақалада біз AI жоғары деңгейдегі мүмкіндіктерге жеткен және соңғы шешімдерді адамдар әлі де қабылдайтын «сингулярлыққа дейінгі дәуір» деп аталатын білім беру парадигмаларының эволюциясын талдаймыз. Мақсат жоғары білім беруде нейробілім берудің, миға негізделген оқытудың және AI-танымның теориялық негіздерін пайдалана отырып, нейроцифрлық оқыту моделін құру. Жұмыстың әдіснамасы ғылыми басылымдардың, халықаралық стратегиялық құжаттардың және отандық тәжірибенің әдіснамалық талдауына негізделген, сапалық тәсілмен және негізделген теориямен біріктірілген. Талдау білім беру модельдерінің «Білім 1.0»-ден «Білім 6.0»-ге эволюциясын қарастырады, педагогке бағытталған жүйеден антропо-цифрлық синергияға ауысуды баса көрсетеді. Білім алушы бірлескен эволюцияның субъектісі болып табылады, ал педагог AI-мен бірлесіп когнитивті серіктес ретінде әрекет етеді. Зерттеу сонымен қатар нейробілім беру, миға негізделген оқыту және AI-таным арасындағы байланыс туралы түсінік береді. Когнитивті жүктеме теориясы AI-мен ішкі жүктемені қалай азайтатынын, сыртқы жүктемені қалай шектейтінін және қосымша жүктемені қалай арттыратынын түсіндіру үшін қолданылады. Сондай-ақ, AI-танымның когнитивті белсенділіктің әлсіреуі, фрагменттелген зейін және AI тәуелділік сияқты қауіптері анықталған. Ұсынылған парадигма нейропластикалықтықты, миға негізделген оқытуды, когнитивті жүктемені оңтайландыруды және AI танымдықты AI-мен тұлғаның ойлауын қолдайтын, оны алмастырмайтын бірыңғай модельге біріктіреді. Алынған нәтижелер білім беру ғылымы мен Қазақстандағы жоғары және жоғары оқу орнынан кейінгі білім берудің цифрлық трансформациясы үшін теориялық және практикалық салдарға әкеледі (Smart.EDU).

Түйінді сөздер: технологиялық сингулярлық, нейро-цифрлық оқыту, ней-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

робілім беру, миға негізделген оқыту, AI-таным, когнитивті жүктеме теориясы

Дәйексөздер үшін: Баумуратова Д.Б., Жукабаева Т.К., Абдрахманов К.А., Ахмад З., Сеитов А.А. (2026). Жоғары білім беруде AI негізделген цифрлық трансформацияға арналған нейро-цифрлық оқыту моделі // Халықаралық ақпараттық және коммуникациялық технологиялар журналы. Т. 7. No 27. Б. 219–240. https://doi.org/10.54309/IJICT.2026.27.3.012. (Ағыл. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

Қаржыландыру. *Бұл зерттеу Қазақстан Республикасы Ғылым және жоғары білім министрлігінің Ғылым комитеті тарапынан қаржыландырылды (№ BR28713531 гранты «Жоғары және жоғары оқу орнынан кейінгі білім беру ұйымдарына арналған Smart.EDU интеллектуалды цифрлық жүйесі»).*

НЕЙРО-ЦИФРОВАЯ МОДЕЛЬ ОБУЧЕНИЯ ДЛЯ ЦИФРОВОЙ ТРАНСФОРМАЦИИ ВЫСШЕГО ОБРАЗОВАНИЯ С ПОМОЩЬЮ AI

Д.Б. Баумуратова[1,2,3], Т.К. Жукабаева[1,2,3], К.А. Абдрахманов[2,3], З. Ахмад[2,4], А.А. Сеитов[5]*

[1]Евразийский университет имени Л.Н. Гумилева, Астана, Казахстан;
[2]Международный научный комплекс Астана, Астана, Казахстан;
[3]Международный Университет Астана, Астана, Казахстан;
[4]Университет Хазара, Мансехра, Пакистан;
[5]Южно-Казахстанский университет имени М. Ауэзова, Шымкент, Казахстан.
E-mail: seitovaskar853@gmail.com

Баумуратова Диларам Бекбулатовна — PhD, старший преподаватель, Педагогический институт Международного университета Астана, Астана, Казахстан
https://orcid.org/0009-0009-4621-1886;
Жукабаева Тамара Кокеновна — PhD, профессор Евразийского национального университета имени Л.Н. Гумилева, Астана, Казахстан
https://orcid.org/0000-0001-6345-5211;
Абдрахманов Кайрат Амангельдинович — PhD, Вице-президент по международному сотрудничеству и делам студентов Международного университета Астана, Астана, Казахстан
https://orcid.org/0000-0002-7387-7191;
Зульфикар Ахмад — PhD, кафедра компьютерных наук и информационных технологий, Университет Хазара, Мансехра, Пакистан
https://orcid.org/0000-0002-6278-9516;
Сеитов Аскар Абилханович — докторант, Южно-Казахстанский университет им. М. Ауэзова, Шымкент, Казахстан
E-mail: seitovaskar853@gmail.com, https://orcid.org/0009-0009-7539-9806.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

© Д.Б. Баумуратова, Т.К. Жукабаева, К.А. Абдрахманов, З. Ахмад, А.А. Сеитов

Аннотация. В статье исследуются быстрое развитие искусственного интеллекта (AI) трансформирует высшее образование из традиционного способа распространения информации в интеллектуальную, адаптивную и самообновляющуюся систему. В данной работе анализируется эволюция парадигм образования в так называемую «предсингулярную эпоху», когда AI достиг высокого уровня возможностей, а окончательные решения по-прежнему принимаются людьми. Цель состоит в построении нейро-цифровой модели обучения в высшем образовании с использованием теоретических основ нейрообразования, обучения на основе работы мозга и AI-когниции. Методология работы основана на методическом анализе научных публикаций, международных стратегических документов и отечественного опыта, интегрированном с качественным подходом и обоснованной теорией. Анализ рассматривает эволюцию образовательных моделей от «Образования 1.0» до «Образования 6.0», подчеркивая переход от системы, ориентированной на преподавателя, к антропо-цифровой синергии, где обучающийся является субъектом коэволюции, а преподаватель выступает в качестве когнитивного партнера в сотрудничестве с AI. Исследование также позволяет понять взаимосвязь между нейрообразованием, обучением на основе работы мозга и когнитивными функциями AI. Теория когнитивной нагрузки используется для объяснения того, как AI снижает внутреннюю нагрузку, ограничивает внешнюю нагрузку и увеличивает релевантную нагрузку. Также выявлены риски когнитивных функций AI, такие как ослабление когнитивной активности, фрагментация внимания и зависимость от AI. Предложенная парадигма объединяет нейропластичность, обучение на основе работы мозга, оптимизацию когнитивной нагрузки и когнитивные функции AI в единую модель, где AI поддерживает человеческое мышление, а не заменяет его. Полученные результаты имеют теоретические и практические последствия для педагогической науки и цифровой трансформации высшего и послевузовского образования в Казахстане (Smart.EDU).

Ключевые слова: Технологическая сингулярность, нейро-цифровое обучение, нейрообразование, обучение на основе работы мозга, AI -когниция, теория когнитивной нагрузки

Для цитирования: Баумуратова Д.Б., Жукабаева Т.К., Абдрахманов К.А., Ахмад З., Сеитов А.А. (2026). Нейроцифровая модель обучения для цифровой трансформации высшего образования с помощью AI // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 27. Стр. 219–240. https://doi.org/10.54309/IJICT.2026.27.3.012. (На англ.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Благодарность. *Данное исследование финансировалось Комитетом науки Министерства науки и высшего образования Республики Казахстан (грант № BR28713531 «Интеллектуальная цифровая система организаций высшего и*

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

послевузовского образования Smart.EDU»).

Introduction.

At the present stage, the development of the education system is considered closely connected with the concept of technological singularity. The rapid development of AI has had a very large impact on the existing sphere of society, including education. Owing to this, education began to move from the traditional model of information delivery to an intelligent, independent and adaptive system. Standard knowledge of an industrial nature is gradually losing its importance, and new approaches are coming to replace it on the basis of joint development and wisdom. Previously, the system was based on the accumulation of documentary data, but its limitations are clearly visible. AI can perform the functions of processing, analyzing, and synthesizing information itself. Therefore, the content and methods of education need to be revised. In the works of Vernor Vinge and Ray Kurzweil, technological singularity is described as a hypothetical stage in which AI surpasses human intelligence (Luksha, 2025). The present title is referred to in the scientific literature as the "presingular era". At this stage, AI has not yet become a fully autonomous system, although it has reached a high level. The labor market is also changing. Nevertheless, the decision is still made by a person, and this has a large weight. At this point, one question becomes relevant: will AI limit human capabilities or open the way to a new civilizational stage? There is no unambiguous answer to it. Therefore, the impact of AI on human potential is considered from a double point of view: whether it is an extender or a deterrent. Such changes are already felt, so it is advisable to systematically analyze them from a scientific point of view and prepare them in advance in the fields of education, economics, and management.

The idea of developing a singularity has been attributed to a number of researchers. One of the first to note the progressive character of technical advancement was John von Neumann (Acuña, 2021). In the footsteps of I. J. Goode's notion of an "intellectual explosion" (Good, 1966) that described the potential of AI to enhance itself. Vernor Vinge defined singularity as a qualitative turning point in human civilization (Vinge, 2008). However, Ray Kurzweil associated it with the exponential increase in technology development and suggested real-time projection (Kurzweil, 2005). The concept of singularity has developed at the confluence of scientific and philosophical perspectives and has become an important component of modern technology discourse. It is better preserved in John von Neumann's opinions about this in his obituary of Stanislav Ulam in 1957 than in his own work (Ulam, 1980). The gradual nature of technological advancement is emphasized, as the growth of humans is approaching a specific boundary. This notion is clarified by analogy with the concept of singularity in physics. Whatever the limit at which the law of space and time alters, the same qualitative shift in human history can be brought about by technological growth. Later, J. Hood had this thinking. He went on with his task. He developed idea of "intellectual explosion" on the basis of the self-improvement of AI. From this perspective, intelligent systems can recursively develop themselves and lead to very

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

fast intelligence growth.

The concept of singularity has been systematized in a unique way by Vernor Vinge (Vinge, 2017). This phenomenon he connected with the arrival of superintelligence and the "intellectual explosion". He considered it a qualitative boundary in the history of mankind. In his latest publications, Vinge paid more attention to the integration of human intelligence and machine intelligence. In this sense, singularity began to be understood not as a solely technological transformation but as a product of the connection between man and technology. This means a change in the social and cognitive system. Singularity was popularized in the broader scientific community by Kurzweil (Kurzweil, 2026). He said that technology advancement is exponential, and could have occurred approximately 2045. Kurzweil predicts that the development of AI will go hand-in-hand with biotechnology and human-machine fusion, thus enhancing human cognition. This view enables us to understand singularity not only as a technological event but as a time of evolution of the planet. It reflects the phrase of Pierre Teilhard de Chardin "Omega Point" (de Chardin, 2021). Therefore, singularity is a notion of an evolutionary process emerging from the merging of different scientific and philosophical concepts. However, today's science has not adopted it as a universal model. This statement cannot encompass the complexity of technical and social development. Thus, singularity in this study is not a speculative one-time event but a continual and complicated process of human growth. It is not only a mirror of the future but also of the transformation that is occurring now (Cunningham, 2019). This method provides a way to look at the development of AI, which intensified significantly after 2020, from a broad evolutionary perspective. In contemporary scientific and public discussions, word singularity is often connected with the notion of superintelligence and an "intellectual explosion". However, that is not the whole of its nature. The increasing complexity and restriction of human control were also features of earlier historical times.

With the development of technology, the introduction of neuroscience into education is also deepening. Brain-computer interfaces are considered the next stage of neurolearning. In the future, people can interact with digital content directly on a neurophysiological level. Such systems measure brain activity and provide feedback at the same time. AI systems can control the level of human attention and independently adjust the complexity of the material. This reduces excessive cognitive load. Intelligent systems that consider emotions are also being developed. For example, they could control anxiety, consolidate attention and direct themselves to adapt to learning. Independent neuroscience-based technologies are transforming education into a more flexible system (Goodwin, 2022).

The rapid development of AI has a deep impact on all spheres of society, especially education. Education is moving from a traditional model of information delivery to an intelligent, independent and adaptive system. Since AI itself can process, analyze and synthesize information, the limitations of the former system based on the accumulation of documentary data are clearly visible. In this regard, there is a need for a radical review of the content and methods of education. Today's period is called

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

the "era before singularity" in the scientific literature: Although AI has reached a high level, the decision is still made by man. It is at this point that the double question of whether AI is an extender or a deterrent to human potential becomes relevant. A systematic analysis of these changes from a scientific point of view and a preliminary preparation of the areas of education, economics and management have become urgent tasks. For Kazakhstan, this relevance is due to the digital transformation of higher and postgraduate education ("Smart.EDU" direction) is directly related. Development of a neuro-digital learning model for higher education through a systematic analysis of the transformation of educational paradigms in the era of singularity, the conceptual foundations of neuroeducation and learning, and international and domestic experience. Our six research questions are as follows:

RQ1: Will AI limit human capabilities or pave the way for a new civilizational stage?

RQ2: How have educational paradigms been transformed to meet the demands of society and the development of technology?

RQ3: How are the areas of neuroeducation, brain-based learning, and AI related and why are they separated?

RQ4: How can the cognitive load (internal, excessive, useful) be effectively balanced in AI-supported learning?

RQ5: Will AI support a person's thinking, or will it make him or her dependent on it?

RQ6: What learning model that combines neuroscience, cognitive psychology and AI can meet the requirements of modern education?

The scientific novelty of the study consists of the reconceptualization of technological singularity not as a single future event, but as a continuous, multidimensional evolutionary process already in progress, which makes it possible to consider the development of artificial intelligence in education in the broad evolutionary environment of human cognition rather than in a narrow technological environment. On this basis, the study systematizes the evolution of educational paradigms from "Education 1.0" to "Education 6.0" and substantiates the regular transition from a teacher-centered model to an anthropo-digital synergy, in which the learner becomes a subject of coevolution and the teacher a cognitive partner acting together with AI. Neuroeducation, brain-based learning and AI-based learning are integrated and differentiated for the first time in a common analytical framework by their goals, scientific foundations, methods and dangers, and cognitive load theory is employed to characterize AI-supported learning to demonstrate how AI can decrease the intrinsic load, increase the germane load and constrain extraneous load while generating hazards such as cognitive offloading, attention fragmentation and AI dependency. The main contribution is an original neuro-digital learning model that combines neuroplasticity, brain-based learning, cognitive load optimization and AI cognition into a single anthropo-digital pedagogical system based on the principle of AI complementing and reinforcing human thinking, not replacing it, thus turning the learner from a passive user of the digital environment into an active subject who

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

manages his/her own cognitive development. The theoretical and practical value of the study is also increased by the application of the model to the digital transformation of higher and postgraduate education in Kazakhstan (Smart.EDU).

Materials and methods.

Research methodology

The methodological basis of the study is a systematic analysis of scientific publications, international strategic documents and domestic experience in the transformation of educational paradigms, neuroeducation and teaching in the era of singularity, and digital transformation in higher education. Such an analysis is aimed at identifying the scientific and theoretical foundations of the digitalization of the educational environment, as well as a comparative study of foreign and national approaches. To achieve this goal, methods of a qualitative methodological approach and grounded theory are used. (Turner, 2021). This approach allows one to construct a conceptual model from empirical data on the basis of systematic analysis. The proposed methodology consists of five stages aimed at developing a model of neuro-digital learning in higher education and identifying and analyzing the conceptual foundations and international and domestic experience. These stages are described in Fig.1 and presented as a practical example.

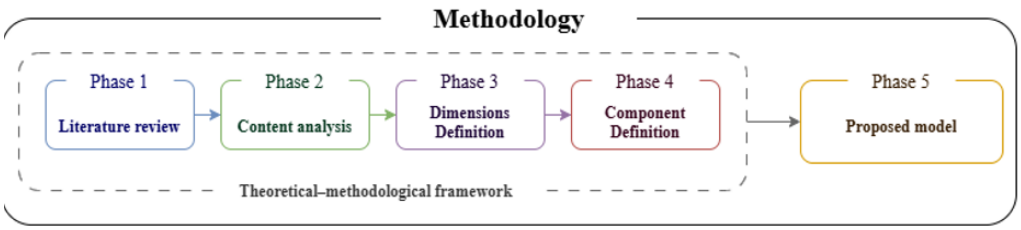


Fig. 1. Applied research methodology

Figure 1 presents the methodological basis for the development of a neuro-digital learning model. It shows a five-step research process that combines theoretical analysis, conceptual synthesis, comparative analysis, model construction and interpretation of its application to higher education. Additionally, in the literature review, we identified the advantages and disadvantages of the Education 6.0 model (Table 1).

Research has also influenced these changes. They allowed us to take a different look at the learning process and paved the way from pedagogically oriented models to student-oriented paradigms. In addition, technological progress has changed the content and format of education. This has created new opportunities for personalization, expansion, and flexible organization of training. As a continuation of these changes, the concept of “Knowledge 6.0” was formed. This model characterizes the education system both in the present and in the future. Its main features are high-level personalization; immersive learning; and the wide use of new technologies such as AI, IoT, and blockchain (Gandawa, 2026; Gandawa, 2025). Such a “smart evolution” is carried out through high-quality technological programs and takes the education

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

system to a new level.

Table 1. Advantages and disadvantages of the Education 6.0 model

Advantages	Disadvantages
Through high-level personalization of AI and data analytics, the ability of each student to form an individual learning trajectory is expanded.	Data security and privacy risks exacerbate the problems of collecting large amounts of student data, protecting personal data, ethical use of data, cybersecurity, there is a need to implement the principle of "privacy by design".
The integration of technology and the human factor in the educational process, AI, Big Data and neurotechnologies are combined with supporting human cognitive and social development.	Ethical issues and personality the balance of technology risks that technology is not seen as an end, but as a means, and that humanistic values may be weakened.
Support for continuing education the education system will focus on continuous training and retraining.	Effects on psychological well-being increased digital workload, the influence of AI tools on motivation, and an increase in stress and emotional distress.
Experiential and real - life oriented learning the integration of theory and practice on the basis of experiential learning is enhanced.	Digital inequality inequality of infrastructure and resources, an increase in the gap between students.
The development of metacritics develops critical thinking, adaptability, self-regulation and the ability to solve complex problems.	The negative effects of generative AI are reduced cognitive effort, dependence on ready-made responses, and the risk of weakening critical thinking skills.
Inclusive and accessible educational opportunities digital technologies provide a wide range of access to education, which reduces geographical and physical restrictions.	The complication of the role of a teacher is the merger of the roles of a facilitator, mentor, analyst, an increase in the professional load, the need for continuous acquisition of new competencies.
Data-based management and decision-making it is possible to improve the quality of education through the analysis of learning outcomes.	The risk of hyper personalization is a weakening of the common educational base, fragmentation of educational content, a decrease in social cohesion.

"Education 6.0" is not only the continuation of the previous models or just another new version. It is a wider and more systematic approach to the overall development of the individual. Here, education is not a matter of giving out information. The emphasis is on leadership, entrepreneurship, critical thinking and the development of a personality designed to cope with future change. Nontraditional thinking, new ideas and innovative activities are important. In this model, the use of advanced technology plays a main role. However, technology is considered only as a tool, and in its application, it is considered important to respect the principles of ethics and sustainable development. This favor the social responsibility of the student and develops the ability to adapt to a changing environment

Results.

Educational models are constantly evolving. They are also subject to change with a changing society. If the economy is refreshed, new technology is introduced, or people's values move, and then education, the style and the content of its instruction become clear. At any stage of the education system, it was possible to prepare a

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

competitive specialist that society needed. Consequently, learning theories have also changed, with a focus on providing people with the knowledge, skills and competences needed to address complex social challenges. The old system, gained through the industrial revolution, was based on a fixed curriculum, the leading role of a teacher and a rigorous evaluation. It is designed primarily for the training of skilled personnel for manufacturing. Society needed something else later. The theories of learning have evolved, the pedagogical method has been reformed, different technologies have become widespread and the economy has gradually shifted to a knowledge-based paradigm. These developments called for the reform of the school system (Shaji, 2025). A stepwise evolution was developed from the "Education 1.0" model, where the teacher plays a key role in the "Education 6.0" model, where the student is an active participant (Fig. 2). Figure 2 describes the evolution of educational paradigms from "Education 1.0" to "Education 6.0".

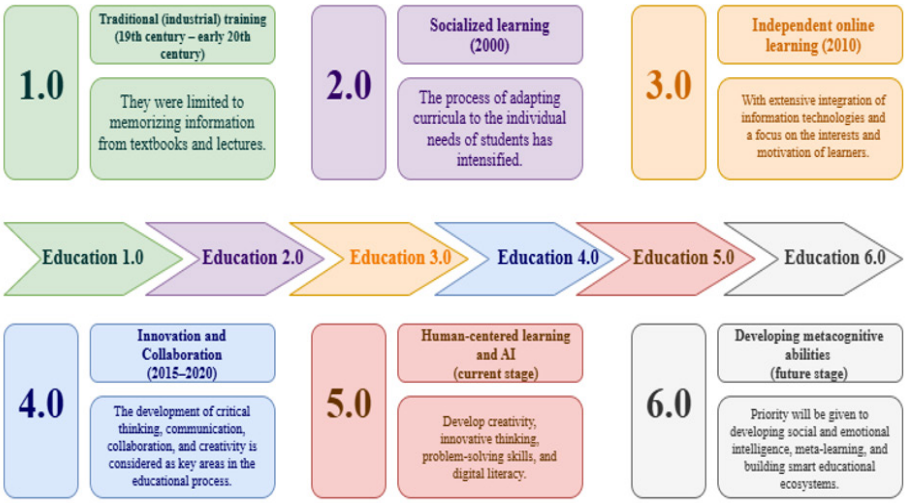


Fig. 2. Evolution of educational models

The development of educational models does not stop with changes in the content and methods of training. It is aimed at identifying and revising the role of the student in the process. The "Education 1.0" model, which was formed at the beginning of the XX century under the influence of the industrial revolution, relied on strict discipline, standardized content and conveyor logic. At that time, the learner was often considered a passive person who received ready-made information. The main goal of the system is to train the personnel needed by industry. However, over time, this model began to be criticized since it did not consider the personality, interests and thinking of a person. During this period, scientists such as John Dewey, Jean Piaget, and Lev Vygotsky reported that learning should be based on experience, personal development and social interaction. Thus, the idea was formed that knowledge is not transmitted ready-made but is actively compiled by a person himself. By the middle and end of the XX century, the amount of data related to the analysis of educational results increased,

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

and the internet became widespread. This influenced the formation of the "Education 2.0" model. Although elements of personalization were introduced during this period, the main role of the teacher in teaching remained. Even if the technology was used, it did not radically change the learning system.

At the beginning of the XXI century, the development of digital technologies and the spread of constructivist theories paved the way for the "Knowledge 3.0" model. In this model, more attention was given to the interests of the student and the individual educational trajectory. However, excessive emphasis on personalization in some cases weakens the skills of communication between people. After that, the "Education 4.0" model was formed. Joint reading, problem-oriented tasks and the application of knowledge in a life situation came to the fore here. However, too much attention to a specific practical skill inhibits the full development of metacritics. To reduce this limitation, the "Education 5.0" model began to be developed. It focuses on personal development, creativity, critical thinking, innovation, problem solving, digital literacy, and continuing education. In addition, it aims to create an intelligent educational environment that relies on the interaction between humans and AI. Here, learning is adapted to the cognitive, emotional, and social identities of the learner through AI, big data and adaptive algorithms. Metalearning, digital ethics and algorithmic thinking are considered important competencies. In this way, the evolution of education was accompanied by a change in teaching approaches, technologies and educational philosophies. These changes are important for understanding the possibility of the concept of "Education 6.0" (Table 2).

Table 2. Comparative knowledge analysis 1.0 – 6.0

Criterion	Education 1.0	Education 2.0	Education 3.0	Education 4.0	Education 5.0	Education 6.0
Pedagogical paradigm	Teacher-oriented	Standardization and control	Student-oriented	Competence-oriented	Personality-oriented	Anthropo-digital synergy
The role of the teacher	Main source of knowledge	Observer, appraiser	Mentor, facilitator	Moderator	Curator, consultant	Cognitive partner (with AI)
The role of the student	Passive receiver	Performer	Active participant	Problem solver	Personal trajectory owner	Coevolution subject
Learning approach	Lecture, memorization	Standardized tasks	Project-based	Problem-based learning	Personalized learning	Hyper-personalized, adaptive
Educational content	Fixed	Standardized	Flexible, inquisitive	Interdisciplinary	Interdisciplinary creativity	Dynamic, self-renewing
Rating system	Written tests	National standardized tests	Formative assessment	Based on competence	Personalized assessment	Predictive and data-driven
The role of technology	No or minimum	Additional tool	Integrated	Active use	AI and analytics	AI, XR, neurotechnologies
Training format	Classy, traditional	Mixed	Online, offline	Hybrid	Digital system	Virtual environment, MR
Basic skills	Memory	Performance	Creativity, collaboration	Critical thinking, communication	Meta-learning, innovation	Cognitive flexibility, conscious thinking
Cooperation	Low	Limited	High	Teamwork	Global cooperation	Human-AI collaboration
Personalization level	No	Low	Average	Average	High	Hyper-personalization
Training environment	Physical class	Organizational	Digital space	Smart environment	Intelligent system	Digital Without Borders-bioenvironment
The main purpose	Education	Compliance with standards	Self-study	Formation of competence	Disclosure of personal potential	Transformation of consciousness

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

These transformations are proof that the teacher-student connection is constructed in a very different manner than before. The main role was played by the teacher, but now the activities and independence of the student are coming to the fore. Digital technologies and AI are becoming basic elements of the educational process at the same time. If we see the evolution of the education system, it is quite evident that the aim of the education system is not only to provide information, as it was earlier. Today, it is about the attitude of a person, his ingenuity and ability to swiftly adapt to a new circumstance. Technology and AI have fundamentally changed the learning space. As the trend of EdTech becomes more extreme, future training systems will likely evolve to a more flexible style with more intelligence. Thus, the main task of the education system of today is to teach a person who is easily adaptable to change, thinks independently, and does not rely on a ready-made answer and freely masters it in a digital environment.

At the heart of the concept of "Education 6.0" is the idea of complementarity between man and machine. That is, technology is not used to replace a person but to strengthen his or her creativity, problem-solving ability and innovation ability. In addition, this model aims to create an environment that supports continuing education, interdisciplinary solidarity, and ethical decision-making (Nieto, 2026). Owing to this, the education system strives to become more flexible, more efficient and accessible to everyone. The main features of this model are as follows (Fig. 3):

- Adaptive AI-based learning has been widely applied. AI can customize the learning process automatically depending on the student's abilities and needs;

- The application of blockchain in the preservation and validation of educational achievements has already started. This enhances the security, openness and credibility of educational certifications;

- Immersive and interactive learning has developed. Through technologies such as VR, AR, and MR, the learning process is organized in a format that is closer and more interesting to practice;

- More emphasis was placed on personal skills. Universal abilities such as creativity, empathy, and critical thinking have become among the core competencies;

- The principle of continuous education has been strengthened. As the labor market is changing rapidly, people's need to constantly reread and acquire new skills is seen as an important factor. Figure 3 presents the main components of "Education 6.0".

The development of "Education 6.0" has led to the increasing integration of technology and pedagogy. This makes the educational process more adaptable than ever and adaptable to the needs of all (Moleka, 2023). Six guiding concepts underpin the model. There is one final goal. Completion of the development of personality. To prepare it for the complicated requirements of today's society (Fig. 4). Figure 4 presents the key principles of the Education 6.0 system that support the holistic development of learners.

The core ideas of the "Education 6.0" model suggest the all-round growth of the student. They are:

- Holistic learning develops the student's mind and emotions and social and

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ethical dimensions in conjunction. Therefore, he develops a feeling of sympathy, responsibility, and the ability to communicate with a person.

- The multidisciplinary approach involves blurring the boundaries between different areas. Afterward, the student will be able to learn the full knowledge and see a complicated topic from multiple sides.

- The premise of sustainability and ethics is the concept of environmental responsibility and sustainable development. Here, the pupil is taught to estimate what trail his doings leave to society and nature.

- Awareness of the world's civic cultural diversity, not apathetic to global problems, the ability to operate freely in varied contexts.

- The idea of lifelong learning encourages continuous learning. Technology and the labor market are changing so fast that a person has to keep up with it.

- Advanced technologies in the learning process. This includes such tools as AI, VR, AR and blockchain. With their guidance, you can view a complex concept with the naked eye and feel it with your hands.

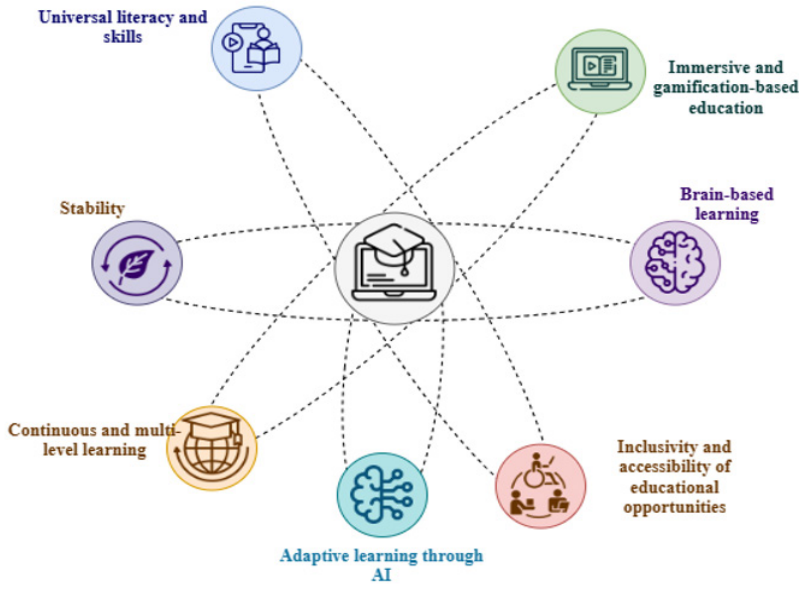


Fig. 3. Components of "Education 6.0"

Thus, these principles educate the learner not only with knowledge but also with the skills and values needed in the current world. In an age of rapid technological progress, education will need to be drastically altered (Swargiary, 2026).

Brain-based learning (educational neuroscience) is a pedagogical approach that involves organizing learning on the basis of the functioning of the brain. The main idea here is to create a learning environment that meets the natural perception feature of the human brain. For this reason, it is not enough to simply provide information in this teaching. In contrast, active action, experience, and the parallel connection of several senses come to the fore. This approach helps students to better master knowledge and skills. In brain-based learning, understanding brain function is central. This is because

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

it allows the teacher to compose the educational material not just as a theory but also by taking into account how the person perceives it. For example, a task involving several senses, joint work, or an activity that requires active participation keeps information firmly in memory. Then, a person will be able not only to perceive the material but also to apply it later in practice. The development of this direction is associated with the deepening of the scientific understanding of the brain.

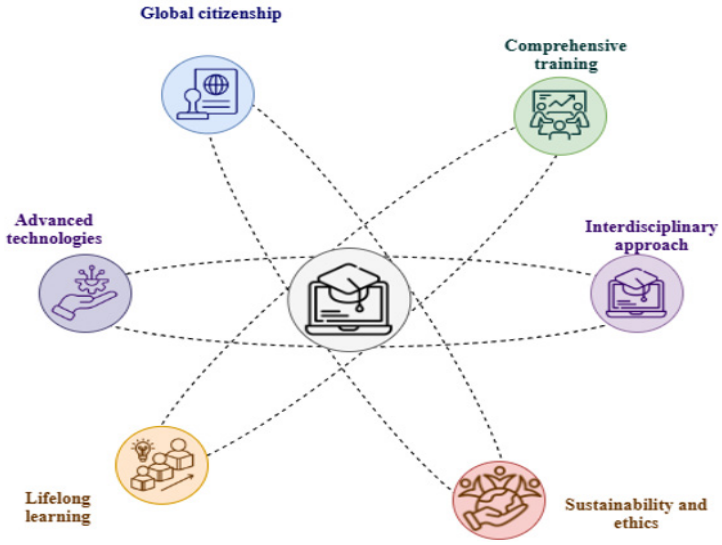


Fig. 4. Principles of "Education 6.0"

Brain-based learning is based on several fundamental ideas. One way to do this is to consider the specificity of the memory system; that is, information must be organized according to the ability of a person to recall. It would also be good if the educational content is structured so that attention is kept. New information together with past expertise help make a person understand the content better. If there is much knowledge, and it is disorganized, the cognitive load increases, and learning becomes more difficult. Hence, the proper amount of information is also among the main factors. Additionally, the role of emotions in learning and memory is considered: material that touches the soul is usually stored for a long period. One advantage of this method is that it is free of "neuromyths", for which there is no scientific basis. Not so here. It is not unusual, but it is a concept not explained by science. Brain activity was first explained at a very basic level. Later, the theories of the right and left cerebral hemispheres began to propagate. The triple-brain notion of Paul McLean also became widespread. The paradigm suggested that the bottom of the brain was for life, the middle for emotions and the top for thinking. Although these theories are now subject to critique, they have shaped the development of educational neuroscience.

The parallel use of such a strategy allows the results of training to be consistently improved. In the modern education system, neuroeducation, brain-based learning and AI-based cognition are closely related (Table 3). The ultimate goal of all three is to

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

increase the cognitive effectiveness of one study, but the theoretical basis, method, and technology used are separate. Moreover, the intensive introduction of AI into education has raised the problem of cognition to a completely new level. Now, a surprise question came to the fore: will AI support human thinking, or will it make it dependent on it? The balance between these two extremes has become an urgent problem.

Table 3. Neuroeducation, brain-based learning and AI-based communication

Criterion	Neuroeducation	Brain-based learning	AI-cognition
Purpose	Implementation of neuroscience data in the educational process	Aligning learning with the natural principles of brain functioning	Support and modeling cognitive processes through AI
Scientific basis	Cognitive neuroscience, psychology, pedagogy	Neuropsychology and learning science	AI, Cognitive Science, Learning Analytics
Main focus	Neural mechanisms of learning and memory	Attention, emotion, memory and motivation	The cognitive interaction between human and AI
Methods	Neuroeducation, active recall, intermediate repetition	Multi sensory learning, storytelling, collaboration	Adaptive learning, AI pedagogue, predictive analytics
Cognitive impact	Strengthening neural connections	Deep assimilation of information and increased motivation	Cognitive reinforcement and information automation
Advantages	Strengthens long-term memory	Increases training efficiency	Creates a personalized learning environment
Risks	Cognitive overload	Separation of attention	AI dependency and cognitive load reduction
The role of the student	Active cognitive subject	Emotional and cognitive participant	Hybrid learner (personality + AI)
The result	Cognitive flexibility	Deep learning and involvement	Formation of neuro-digital cognition

AI-based cognition has become one of the street areas of Science today. It is not limited to the automation of education. AI began to influence the cognitive activity of a person itself (Table 4).

Table 4. Cognitive implications of AI cognition

AI-the phenomenon of cognition	Description	Cognitive implications
Reducing cognitive load	Transfer of the functions of thinking and memory to AI	Weakening of independent cognitive activity
AI-generated memory	Memory support via AI	Quick search and organization of information
Attention fragmentation	Fragmentation of attention due to digital stimuli	A decrease in deep focus and analytical thinking
Algorithmic thinking	Thinking through algorithmic logic	Structural analysis intensifies, but creativity may be limited
AI dependency	Excessive dependence on AI tools	The weakening of metacognition and critical thinking

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Cognitive load theory has a separate place in the interpretation of neuroeducation and AI-supported learning. It allows you to distinguish what load the brain is facing in a digital learning environment where AI is used (Fig. 5). Figure 5 describes the cognitive load structure in digital learning supported by AI.

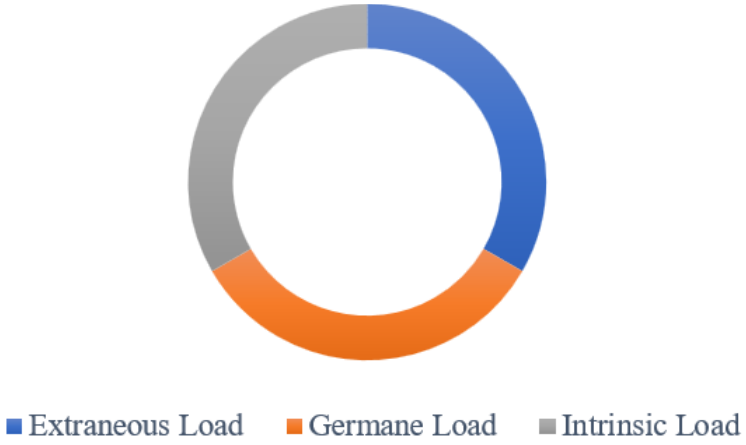


Fig. 5. The structure of the cognitive load in digital learning

The internal load accounts for the largest share. It depends on a few variables. Complexity of the educational material and initial pupil preparation. AI systems can alleviate this burden slightly by organizing information, analyzing it for everyone and providing individual feedback. An overload not directly related to learning content is an overload. In the digital world, it is due to the collision of several activities, a variety of notifications, an excess of information and the dispersion of attention. Rather, AI systems are more prone to overclock this overload if they are not pedagogically adequately developed. Exercise that uses valuable mental energy for meaningful thought, deep learning and the creation of new neural pathways. This load is closely related to neuroplasticity and is crucial for improving analytical thinking, metacognition and problem-solving skills (Table 5).

Table 5. Cognitive load theory and AI relationship

Cognitive load component	Content	AI communication
Intrinsic Load	Natural complexity of the material	AI interprets complex content step by step
Extraneous Load	Excess information load	Excessive digital stimuli increase cognitive load
Germane Load	Cognitive resource aimed at structuring knowledge	Strengthens the concept of AI through mapping and reflexive learning

Discussion.

Research shows how necessary it is to properly balance the cognitive load in AI-supported learning. AI is not a tool that replaces all human thinking. It should be used as a system that supports, and strengthens cognitive ability. In the modern era

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

of AI, the education system is not limited to digitalization but is being rebuilt on a neurocognitive basis. Neuroeducation, the relationship between brain-based learning and cognitive load theory, formed the basis for the formation of a new anthropo-digital pedagogy supplemented by AI. In the framework of, the proposed neuro-digital learning model sees the learner as an active participant in the cognitive partnership between humans and AI and not as a mere user of AI. This is one of the new paradigm directions in modern educational science. Today, the education system is changing significantly under the influence of neuroscience, cognitive psychology and AI. Thus, neuroeducation and brain-based learning have become the leading scientific directions for understanding the changeable and adaptive nature of the learner's brain. In the framework of neuroeducation, the learning process affects the formation of new neural connections in the brain. Brain-based learning involves aligning learning approaches to the natural cognitive mechanisms of the brain.

The development of AI technology has also brought the problem of cognition to a new level. Now several concepts describe the change in human cognition. Reduced cognitive load, memory with AI, distraction, algorithmic thinking, and AI dependency. In reducing the cognitive load, a person transfers part of the work of memory to a digital device and an AI system. This, at first glance, allows you to focus more on a complex analytical task, but also carries the risk of weakening long-term memory. External memory systems created by AI make it easier to store and rediscover information. Only in this case may the activity of independent thinking decrease. In addition, the continuous flow of information in the digital environment destabilizes attention and dulls the ability to read deeply and focus for a long time. Therefore, in the modern education system, honing deep thinking and constant attention has become one of the main tasks. In the era of AI, algorithmic thinking is a new type of logical, structured thinking. It helps to solve the problem systematically, although if it is exceeded, it can introduce creativity and critical reflection. AI dependency is the risk that a person relies excessively on AI in decision-making. Therefore, the main pedagogical principle of using AI is not to replace human intelligence but to support and strengthen it. Cognitive load theory plays an important role in explaining these problems.

AI technology can optimize the quality of training by tuning and changing the major component of the load. For example, picturing a complex idea decreases the internal load, filtering out unnecessary information lessens the overload, and reflection and knowledge mapping increase the useful cognitive load. The integration of AI and neurotechnology into higher education raises ethical issues that extend beyond the effectiveness of digital learning. Within the proposed neuro-digital learning model, ethical governance should therefore be considered an integral component of human AI cognitive interaction rather than an additional technological requirement (Fig. 6). Figure 6 shows a neuro-digital learning model that integrates neuroplasticity, brain-based learning, cognitive load, and AI cognition into an anthropo-digital pedagogical system aimed at supporting AI cognitive synergy.

This model is an anthropo-digital pedagogical system that combines neuroplasticity, brain-based learning, cognitive load theory and AI cognition in one

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

channel. His original idea is simple. AI should not replace students' thinking but should again support and strengthen their cognitive abilities. In the model, while neuroplasticity is the adaptive backbone of the learner's brain, brain-based learning involves creating a learning environment that is compatible with the natural cognitive process. The cognitive load optimization mechanism attempts to strike a balance between internal, external and useful cognitive load. In the AI-cognition phase, AI systems function as tools for metacognitive support, reflection, and cognitive partnership. Then a cognitive synergy is born between a person and the AI: the learner ceases to be an indifferent user in the digital environment and becomes an active subject who himself or herself controls his or her cognitive development. In short, the neuro-digital learning model combines neuroscience, cognitive psychology and AI, a new paradigm trend in modern educational science. His ultimate idea is that AI does not replace human intelligence but rather expands its cognitive potential.

The major contribution of this study is the neuropedagogical learning model, which merges four previously separate research areas, namely, neuroscience (neuroplasticity), brain-based learning, cognitive load theory, and AI-driven cognitive processes, into one operational framework. Most of the present digital learning models make use of only one of the areas or are neutral with respect to the cognitive architecture of the learner. However, the proposed model sees them as interdependent layers within a single process, where neuroplasticity forms the adaptive base, brain-based learning forms the learning environment, and cognitive load theory provides the regulation of the distribution of mental effort and cognitive processes driven by AI to provide metacognitive support.

The second aspect of innovation is the explicit technique of artificial intelligence intervention. Instead of describing AI as a general tool of "adaptation", the model describes its function from the perspective of cognitive load theory: AI alleviates internal load by structuring complex information, reduces external load by filtering irrelevant stimuli, and increases relevant load through reflection and matching of knowledge. The learner is not conceived as an object that can be optimized by an algorithm but as a codeveloping entity that maintains agency in its own cognitive development. The model structurally embeds meta-awareness and reflection rather than having to add them as a way of dealing with the known dangers of using AI, such as reduced independent cognition, fragmentation of attention, and excessive dependence.

These features distinguish this model from the approaches that currently dominate in this field. Adaptive learning optimizes content delivery on the basis of performance data but does not consider either brain function or student independence; intelligent education focuses on technologically saturated environments; and AI-enabled learning complements the performance of specific tasks, while largely ignoring the risk of addiction; and "Education 6.0", although broad and humanistic, functions as a social paradigm rather than an operational learning model. The proposed model connects the microlevel neurocognitive processes with the macrolevel digital infrastructure and management in the context of the Kazakh program Smart.EDU.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

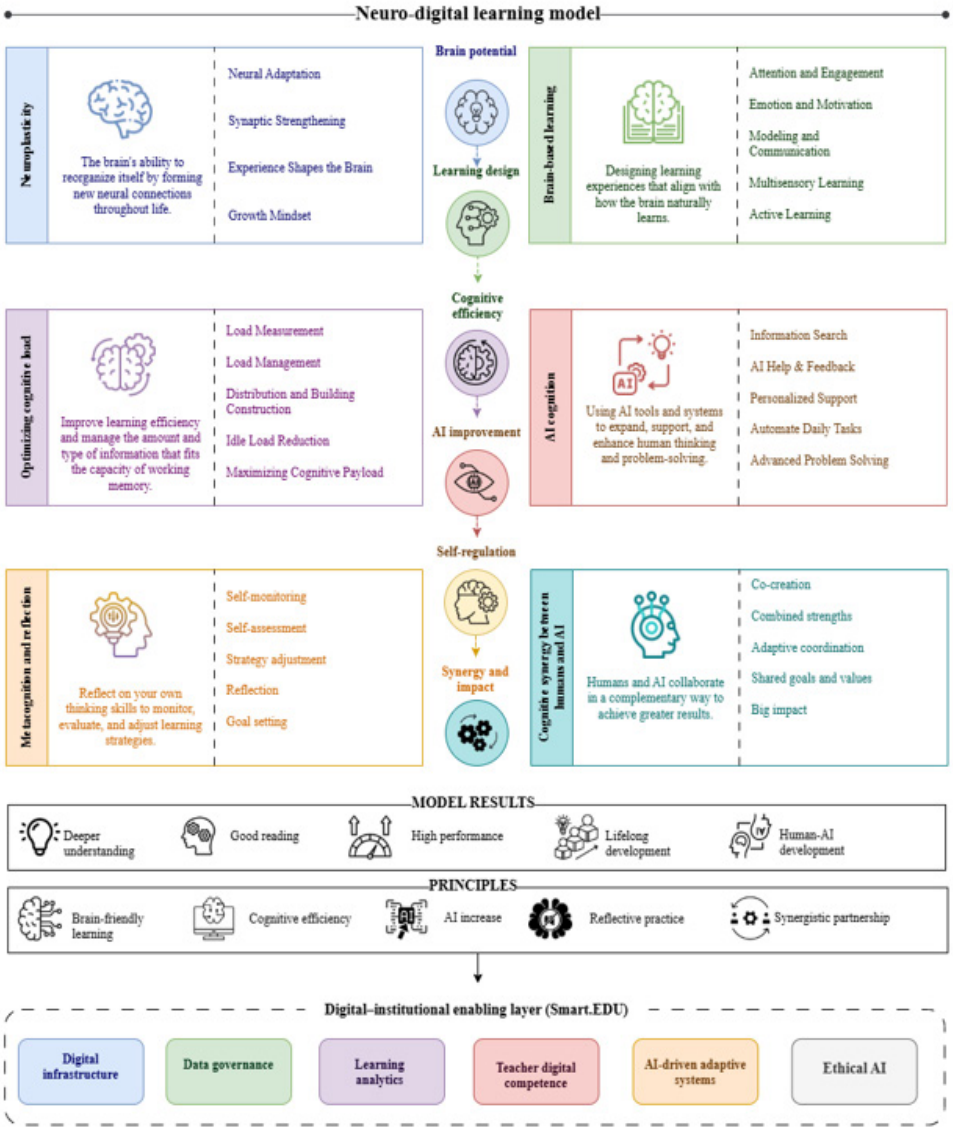


Fig. 6. Neuro-digital learning model

Conclusion.

This study examined changes in educational paradigms during the presingularity period. The results of the research provide a description of the singularity as a continuing, multidimensional evolutionary process that is taking place in the present, not a single event that is to take place in the future. From this point of view, it was demonstrated that considering the development of AI not in the limited technological context but in the broad evolutionary context of human cognition and the education system is necessary.

The analysis of the changes in educational models from "Education 1.0" to "Education 6.0" demonstrates a legal movement from the traditional teacher-oriented

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

system to the anthropo-digital synergy. The reason for this transition is a fundamental change in the role of the teacher and the learner. The learner becomes a subject of coevolution, not a passive recipient, and the pedagogue evolves into a cognitive partner working together with AI rather than a sole supplier of knowledge. The concept of "Education 6.0" is an expanded systematic approach that does not limit education to the supply of information but is focused on the overall development of the individual and the formation of critical thinking and meta-processes.

This study revealed the synergy of neuroeducation and brain-based learning with AI cognition. All of them are different from each other in terms of theory and tactics, but the final purpose is to improve the cognitive effectiveness of shared learning. Cognitive load theory helps explain this relationship. AI systems can reduce the internal burden by parsing complex content in phases, minimize overload by filtering out extraneous information, and increase useful cognitive load by mapping reflection and knowledge. Moreover, the results of this study revealed that the negative effects of AI cognition should not be ignored, as they involve a weakening of cognitive activity, fragmentation of attention and a dependence on AI.

The proposed neuroscience learning model is not a universal technological solution, but a conceptual and methodological framework that adapts to a specific educational environment. The model combines neuroplasticity, brain-based learning, cognitive load regulation, and AI supported learning in a single educational system. Its main principle is that AI should not replace the student's thinking, but complement and contribute to the development of cognitive capabilities. This approach maintains the cognitive autonomy of the learner and sees him as an active subject who controls his learning and cognitive development, and not as a passive consumer of the digital environment. The implementation of the model requires pedagogical support of the teacher, technological accessibility and compliance with ethical requirements. At the present stage, due to the fact that its educational effectiveness has not been empirically confirmed, widespread generalization of the results is limited. In this regard, at the next stage of the study, it is necessary to pilot the model in higher education and assess its impact on cognitive autonomy, metacognitive consciousness, critical thinking, learning activity, cognitive load, as well as the risks of over-dependence on AI. Thus, the developed model is aimed at combining the principles of neuroscience, cognitive psychology and AI in the process of higher education and is of theoretical and practical importance amid the ongoing digital transformation of higher education in Kazakhstan (Smart.EDU).

REFERENCES

Acuña, P. (2021). von Neumann's Theorem Revisited. Foundations of Physics. — Vol. 51. — No. 3. DOI: https://doi.org/10.1007/s10701-021-00474-5 [In Eng.].

Cunningham, I. (2019). A new educational paradigm for the 21st century. Development and Learning in Organizations // An International Journal. — Vol. 34. — No. 2. — Pp. 5–7. DOI: https://doi.org/10.1108/dlo-10-2019-0253 [In Eng.].

de Chardin, T. (2021). the Omega Point // Debating Faith: Essays on Science, Religion and Secularism. — Pp. 317. [In Eng.].

Gandawa, G. (2025). The STEMMA Classroom: Engineering Emotionally Literate Pedagogy in Secondary Technical Education // Journal of Advanced Engineering and Technology (JAET). [In Eng.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Gandawa, G. (2026). Digital Transformation and Quality Assurance in Universities in the Age of Education 6.0 Author // Journal of Education and Learning Sciences. [In Eng.].

Good, I. J. (1966). Speculations concerning the first ultraintelligent machine // In Advances in computers. — Vol. 6. — Pp. 31-88. Elsevier. [In Eng.].

Goodwin, L. (2022). Educational neuroscience, artificial intelligence, and technology in higher education. University of Pennsylvania. [In Eng.].

Kurzweil, R. (2026). The singularity is nearer: When we merge with AI. Penguin Group. [In Eng.].

Luksha, P. (2025, March 16). Careers in the pre-singularity age: Is everything already lost, or are we on the verge of reinvention? https://pavel-luksha-protopia.medium.com/careers-in-the-pre-singularity-age-is-everything-already-lost-or-are-we-on-the-verge-of-b4899e5ef28e [In Eng.].

Moleka, P. (2023). Exploring the Role of Artificial Intelligence in Education 6.0: Enhancing Personalized Learning and Adaptive Pedagogy. DOI: https://doi.org/10.20944/preprints202309.0562.v1 [In Eng.].

Nieto Jalil, J. M. (2026). Emerging Technologies and Regenerative Pedagogies in the Era of Education 6.0. Learning Analytics. — Pp. 141–159. DOI: https://doi.org/10.1201/9781003678250-9 [In Eng.].

Shaji George, A., Baskar, T., & Siranchuk, N. (2025). The evolution of Education 5.0 in the innovation era: A review of the progression from teacher-centered learning to student-driven models // Partners Universal International Innovation Journal. — Vol. 3. — No. 1. — Pp. 19–28. [In Eng.].

Swargiary, K. (2026). Education 6.0: Envisioning the Symbiotic Future of Learning and Innovation. Atlantis Publication House (Russia). [In Eng.].

Turner, C., & Astin, F. (2021). Grounded theory: what makes a grounded theory study? // European Journal of Cardiovascular Nursing. — Vol. 20. — No. 3. — Pp. 285–289. DOI: https://doi.org/10.1093/eurjcn/zvaa034 [In Eng.].

Ulam, S. M. (1980). Von Neumann: The Interaction of Mathematics and Computing // A History of Computing in the Twentieth Century. — Pp. 93–99. DOI: https://doi.org/10.1016/b978-0-12-491650-0.50014-9 [In Eng.].

Vinge, V. (2008). Signs of the singularity // IEEE Spectrum. — Vol. 45. — No. 6. — Pp. 76–82. DOI: https://doi.org/10.1109/mspec.2008.4531467 [In Eng.].

Vinge, V. (2017). The coming technological singularity: How to survive in a post-human era // Science Fiction Criticism. DOI: https://doi.org/10.5040/9781474248655.0037 [In Eng.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 241–257
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.013

МРНТИ 20.53.00

MULTIMODAL SEMANTIC MANIPULATION IN FAKE VIDEOS AND ITS INFLUENCE MECHANISM ON USERS' PERCEIVED TRUST

T. Daoyun[1], G.A. Amirkhanova[1], Fu Yanwei[2], Y. Faruzkyzy[1]*
[1]Al-Farabi Kazakh National University, Almaty, Kazakhstan;
[2]Fudan University, Shanghai, China.
E-mail: elnura.dauletkhan@gmail.com

Tang Daoyun — PhD student, Al-Farabi Kazakh National University, Almaty, Kazakhstan
https://orcid.org/0009-0000-7015-5276;
Gulshat A. Amirkhanova — PhD, Al-Farabi Kazakh National University, Almaty, Kazakhstan
https://orcid.org/0000-0003-3933-5476;
Fu Yanwei — Professor, School of Data Science, Fudan University, Shanghai, China
https://orcid.org/0000-0002-6595-6893;
Yenlik Faruzkyzy — Bachelor, Al-Farabi Kazakh National University, Almaty, Kazakhstan
E-mail: elnura.dauletkhan@gmail.com, https://orcid.org/0009-0001-1030-8266.

© T. Daoyun, A. Gulshat, Fu Yanwei, F. Yenlik

Abstract. In the contemporary digital information environment, fake videos have become an important challenge to online information ecology. Unlike text-only misinformation, video content integrates textual, audio, and visual modalities, which may shape users' trust judgments through different semantic cues. Drawing on Information Manipulation Theory, this study examines how five types of semantic manipulation—fabrication, concealment, misinterpretation, ambiguity, and incitement—affect users' perceived trust across textual, audio, and visual modalities. A within-subject experiment was conducted with 48 valid participants under 15 manipulation conditions, complemented by semi-structured interviews with 24 participants. A pilot manipulation check was conducted before the formal experiment to validate whether the experimental stimuli were perceived as intended. The results show that users' perceived trust varies significantly across manipulation types and modalities. Textual fabrication and audio fabrication were associated with higher perceived trust, whereas textual concealment, textual ambiguity, textual incitement,

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

and audio incitement reduced perceived trust. Visual fabrication also increased perceived trust, while the effects of other visual manipulations were weaker or statistically non-significant. The interview results further suggest that authority-like cues, information completeness, emotional tone, and attention allocation help explain users' modality-specific trust judgments. This study extends Information Manipulation Theory to multimodal fake-video communication and provides empirical implications for platform governance and media-literacy interventions.

Keywords: fake video, multimodal semantic manipulation, perceived trust, information manipulation theory, within-subject experiment, mixed methods

For citation: Daoyun T., Gulshat A., Yanwei Fu, Yenlik F. (2026). Multimodal semantic manipulation in fake videos and its influence mechanism on users' perceived trust // International journal of information and communication technologies. Vol. 7. No. 27. Pp. 241–257. https://doi.org/10.54309/IJICT.2026.27.3.013. (In Eng.).

Conflict of interest: The authors declare that there is no conflict of interest.

Acknowledgements. *The article was supported by the project from the Ministry of Science and Higher Education of the Republic of Kazakhstan, No.BR24992975 «Development of a digital twin of a food processing enterprise using artificial intelligence and IIoT technologies» (2024-2026).*

ФЕЙК ВИДЕОЛАРДАҒЫ МУЛЬТИМОДАЛЬДЫ СЕМАНТИКАЛЫҚ МАНИПУЛЯЦИЯНЫҢ ПАЙДАЛАНУШЫЛАРДЫҢ ҚАБЫЛДАЙТЫН СЕНІМІНЕ ӘСЕР ЕТУ МЕХАНИЗМІ

Т. Даоюн[1], Г. Амирханова[1], Фу Яньвэй[2], Е. Фарузқызы[1]*

[1]Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан;
[2]Фудань университеті, Шанхай, Қытай.
E-mail: elnura.dauletkhan@gmail.com

Даоюн Тан — Докторант, Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан
https://orcid.org/0009-0000-7015-5276;
Амирханова Гульшат Аманжоловна — PhD, Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан
https://orcid.org/0000-0003-3933-5476;
Яньвэй Фу — профессор, Фудан университетінің Деректерді өңдеу мектебі, Шанхай, Қытай
https://orcid.org/0000-0002-6595-6893;
Фарузқызы Еңлік — Бакалавр, Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан
E-mail: elnura.dauletkhan@gmail.com, https://orcid.org/0009-0001-1030-8266.

© Т. Даоюн, Г. Амирханова, Фу Яньвэй, Е. Фарузқызы

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Аннотация. Қазіргі цифрлық ақпараттық ортада жалған бейнелер онлайн-ақпараттық экология үшін үлкен проблемаға айналды. Мәтіндік жалған ақпараттан айырмашылығы, бейне мазмұны әртүрлі семантикалық сигналдар арқылы пайдаланушылардың сенім туралы пікірін қалыптастыра алатын мәтіндік, аудио және визуалды құралдарды біріктіреді. Ақпаратты манипуляциялау теориясына сүйене отырып, бұл зерттеу семантикалық манипуляцияның бес түрі — бұрмалау, жасыру, қате түсіндіру, екіұштылық және қоздыру — мәтіндік, аудио және визуалды құралдар арқылы пайдаланушылардың сенімін қабылдауға қалай әсер ететінін қарастырады. 15 манипуляция жағдайында 48 нақты қатысушымен 24 қатысушымен жартылай құрылымдалған сұхбаттармен толықтырылған пәнішілік эксперимент жүргізілді. Ресми экспериментке дейін эксперименттік тітіркендіргіштердің ойлағандай қабылданғанын растау үшін эксперименттік манипуляциялық тексеру жүргізілді. Нәтижелер пайдаланушылардың сенімін қабылдау манипуляция түрлері мен әдістеріне байланысты айтарлықтай өзгеретінін көрсетеді. Мәтінді және аудиожазбаны бұрмалау сенімділіктің жоғары деңгейімен байланысты болды, ал мәтінді жасыру, мәтіннің түсініксіздігі, мәтінді қоздыру және аудионы қоздыру қабылданған сенімді азайтты. Көрнекі бұрмалау сонымен қатар қабылданған сенімділікті арттырды, ал басқа визуалды манипуляциялардың әсері әлсіз немесе статистикалық тұрғыдан шамалы болды. Сауалнама нәтижелері сонымен қатар беделді сигналдар, ақпараттың толықтығы, эмоционалды тон және зейіннің таралуы пайдаланушылардың белгілі бір модальділікке тәуелді сенім туралы пікірлерін түсіндіруге көмектесетінін көрсетеді. Бұл зерттеу ақпаратты манипуляциялау теориясын мультимодальды жалған бейне байланысқа таратады және платформаны басқару және медиа сауаттылықты арттыру шаралары үшін эмпирикалық қорытындылар береді.

Түйінді сөздер: жалған бейне, мультимодальды семантикалық манипуляция, қабылданған сенім, ақпаратты манипуляциялау теориясы, пән ішіндегі эксперимент, аралас әдістер

Дәйексөздер үшін: Даоюн Т., Амирханова Г., Яньвэй Фу, Фарузқызы Е. (2026). Фейк видеолардағы мультимодальды семантикалық манипуляцияның пайдаланушылардың қабылдайтын сеніміне әсер ету механизмі // Халықаралық ақпараттық және коммуникациялық технологиялар журналы. Т. 7. No 27. Б. 241–257. https://doi.org/10.54309/IJICT.2026.27.3.013. (Ағыл. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

Алғыс. *Мақала Қазақстан Республикасы Ғылым және жоғары білім министрлігінің № BR24992975 «Жасанды интеллект және IIoT технологияларын пайдалана отырып, тамақ өнімдерін өңдейтін кәсіпорынның цифрлық егізін әзірлеу» (2024–2026 жж.) жобасы аясында қолдау тапты.*

МЕХАНИЗМ ВЛИЯНИЯ МУЛЬТИМОДАЛЬНОЙ СЕМАНТИЧЕСКОЙ МАНИПУЛЯЦИИ В ФЕЙКОВЫХ ВИДЕО НА ВОСПРИНИМАЕМОЕ ДОВЕРИЕ ПОЛЬЗОВАТЕЛЕЙ

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Т. Даоюн[1], Г. Амирханова[1], Фу Яньвэй[2], Е. Фарузкызы[1]*
[1]Казахский национальный университет имени аль-Фараби, Алматы, Казахстан;
[2]Фуданьский университет, Шанхай, Китай.
E-mail: elnura.dauletkhan@gmail.com

Даоюн Тан — докторант, Казахский национальный университет имени аль-Фараби, Алматы, Казахстан
https://orcid.org/0009-0000-7015-5276;
Амирханова Гульшат Аманжоловна — PhD, Казахский национальный университет имени аль-Фараби, Алматы, Казахстан
https://orcid.org/0000-0003-3933-5476;
Яньвэй Фу — профессор, Школа науки о данных Фуданьского университета, Шанхай, Китай https://orcid.org/0000-0002-6595-6893;
Фарузкызы Енлик — бакалавр, Казахский национальный университет имени аль-Фараби, Алматы, Казахстан
E-mail: elnura.dauletkhan@gmail.com, https://orcid.org/0009-0001-1030-8266.

© Т. Даоюн, Г. Амирханова, Фу Яньвэй, Е. Фарузқызы

Аннотация. В современной цифровой информационной среде поддельные видеоролики стали серьезной проблемой для информационной онлайн-экологии. В отличие от текстовой дезинформации, видеоконтент объединяет текстовые, аудио- и визуальные средства, которые могут формировать мнение пользователей о доверии с помощью различных семантических сигналов. Опираясь на теорию манипулирования информацией, в этом исследовании рассматривается, как пять типов семантических манипуляций — фальсификация, сокрытие, неверное толкование, двусмысленность и подстрекательство — влияют на восприятие доверия пользователей с помощью текстовых, аудио- и визуальных средств. Был проведен внутрипредметный эксперимент с 48 действительными участниками в условиях 15 манипуляций, дополненный полуструктурированными интервью с 24 участниками. Перед формальным экспериментом была проведена экспериментальная проверка манипуляций, чтобы подтвердить, были ли экспериментальные стимулы восприняты так, как предполагалось. Результаты показывают, что восприятие доверия пользователей существенно различается в зависимости от типов и способов манипулирования. Фальсификация текста и аудиозаписи была связана с более высоким уровнем воспринимаемого доверия, в то время как сокрытие текста, двусмысленность текста, подстрекательство к тексту и подстрекательство к аудио снижали воспринимаемое доверие. Визуальная фальсификация также повышала воспринимаемое доверие, в то время как эффекты других визуальных манипуляций были слабее или статистически незначимы. Результаты опроса также свидетельствуют о том, что авторитетные сигналы, полнота информации, эмоциональный тон и распределение внимания помогают объяснить суждения пользователей о доверии, зависящие от конкретной модаль-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ности. Это исследование распространяет теорию манипулирования информацией на мультимодальную поддельную видеосвязь и предоставляет эмпирические выводы для управления платформой и мер по повышению медиаграмотности.

Ключевые слова: поддельное видео, мультимодальная семантическая манипуляция, воспринимаемое доверие, теория манипулирования информацией, внутрипредметный эксперимент, смешанные методы

Для цитирования: Даоюн Т., Амирханова Г., Яньвэй Фу, Фарузқызы Е. (2026). Механизм влияния мультимодальной семантической манипуляции в фейковых видео на воспринимаемое доверие пользователей // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 27. Стр. 241–257. https://doi.org/10.54309/IJICT.2026.27.3.013. (На англ.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Благодарности. *Статья подготовлена при поддержке проекта Министерства науки и высшего образования Республики Казахстан № BR24992975 «Разработка цифрового двойника предприятия по переработке пищевой продукции с использованием технологий искусственного интеллекта и промышленного интернета вещей (IIoT)» (2024–2026 гг.).*

Introduction.

With the rapid development of mobile Internet technology and the popularization of intelligent terminals, online video platforms have gradually become an important channel for the public to obtain information, express opinions, and enjoy leisure and entertainment. According to a statistical report released by the China Internet Network Information Center (CNNIC), as of December 2025, the number of online video users in China has reached about 1.07 billion, accounting for about 96.6% of the total number of netizens. With its intuitive, vivid and large information carrying capacity, video plays an increasingly important role in the information dissemination system.

However, the lowering of the threshold for video content production and the popularization of video editing technology have significantly enhanced the processability and manipulability of video information (Leszczuk et al., 2025). In order to gain attention and traffic, some content producers tamper with or reconstruct real scenes through technical means such as editing, special effects processing, dubbing and mixing, and even create false information through fictional events, splicing clips or taking out of context. Such fake videos will not only mislead public perception, but may also spread rapidly on social media platforms, negatively impacting the online information ecology and social opinion environment.

Unlike traditional text information, video information is usually composed of multiple information modalities such as text, audio and visual images, and is a typical multimodal information carrier. In this multimodal communication environment, users' trust in information content often directly affects their information acceptance, comment interaction, and forwarding and dissemination). Therefore, user trust has become an important precursor factor affecting the effectiveness of video information

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

dissemination.

However, the influence of fake videos cannot be fully explained by technical manipulation alone. Whether manipulated content produces persuasive effects also depends on how users cognitively process semantic cues embedded in different modalities. In short-video communication, textual titles, audio narration, and visual frames may not contribute equally to trust formation. Therefore, examining users' modality-specific trust judgments is necessary for understanding the social impact of fake videos beyond algorithmic detection.

Literature Review

In recent years, a lot of research has been conducted on the identification of disinformation and the trust of social media users.

Multimodal Fake-Video and Disinformation Detection

In the field of disinformation detection, technology-oriented research dominates. Many scholars have tried to use deep learning, machine learning, and multimodal fusion technologies to improve the automation level of disinformation identification by extracting text semantic features, visual features, and cross-modal consistency features. For example, Xi et al. proposed a multimodal consistent neural network model to detect false information by analyzing features such as image tampering and graphic semantic inconsistency (Xi et al., 2021). Liu et al. constructed a multimodal comparative classification model to improve the detection ability of face fake videos by capturing local correlation representations (Liu et al., 2023). Al-Shamayleh research pointed out that the silence feature in fake audio will significantly affect the recognition effect of the detection model (Al-Shamayleh et al., 2025). These studies provide important methodological support for disinformation identification at the technical level. In addition to specific feature extraction, recent comprehensive surveys emphasize the critical role of multimodal approaches in deepfake forensics (El-Shafai et al., 2024; Wang, 2025). Researchers are continuously proposing advanced frameworks that integrate machine learning and AI to enhance overall detection accuracy (Gandhi et al., 2024; Joshi et al., 2025; Reddy et al., 2025). Furthermore, innovative models have begun incorporating user information and emotional recognition into audio-visual analysis, bridging the gap between technical detection and user engagement dynamics (Alsaeedi et al., 2025; Hao et al., 2025).

User Trust and Source Credibility in Social Media

At the same time, in terms of social media user trust research, scholars in the fields of library and information science, communication science and management information systems have also conducted in-depth discussions on the key factors affecting user trust. Existing studies have shown that factors such as information presentation, information quality, and information source have a significant impact on user trust. For example, Ginsca, A. L. et al. found that factors such as system generation clues, publisher identity sources, official certification logos, and link forms will positively affect user trust levels (Ginsca et al., 2015). Li et al.'s research pointed out that the objectivity and professionalism of WeChat official account content will significantly affect users' evaluation of the overall credibility of information (Li et al.,

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

2022).

Information Manipulation Theory in Multimodal Video Communication

Information Manipulation Theory (IMT) was proposed by McCornack, which systematically explains how information disseminators can achieve the specific purpose of deceiving audiences by changing the quantity, authenticity, expression, and relevance of information. In the evolution of social media, common types of false semantic manipulation have gradually solidified into fabricated facts, false claims, impersonation of authority, concealment of the truth, misinterpretation of original meaning, ambiguous scene, and incitement (Tufchi et al., 2023) . Combined with the data characteristics of existing popular science rumor refutation platforms (such as TikTok rumor refutation, scientific rumor refutation, etc.) and expert verification, this study selects the most representative five types of semantic manipulation techniques in the video ecosystem: fabrication, concealment, misinterpretation, ambiguity and incitement. On this basis, its influence on user trust in three modalities: text, audio and image is discussed. The five types of semantic manipulation in the video are shown in Figure 1.

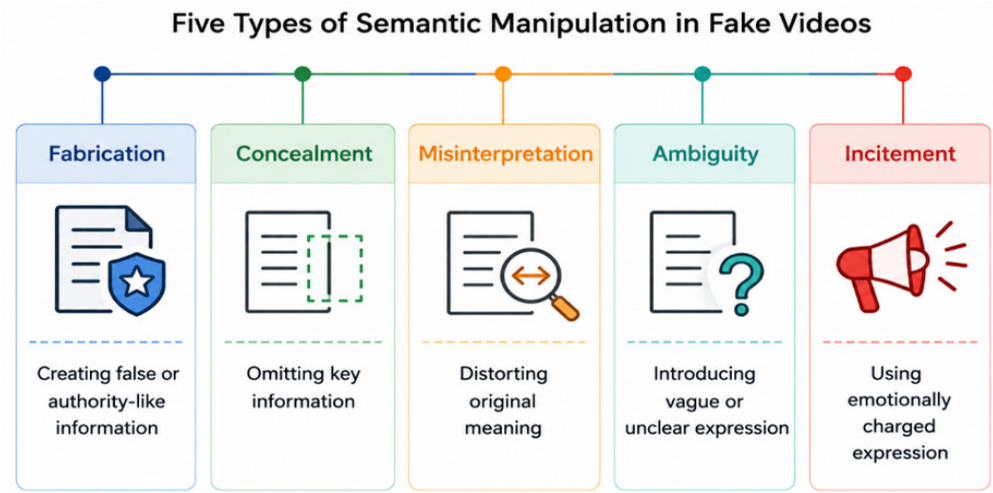


Fig. 1. Five types of semantic manipulation in fake videos

Research gaps

Although previous studies have provided important insights into disinformation detection and social media trust, three limitations remain. First, most fake-video studies have emphasized algorithmic detection accuracy, feature extraction, or multimodal consistency modeling, whereas users' cognitive responses to manipulated video semantics have received less attention. Second, existing user-trust research has mainly focused on text-based social media platforms, while video communication involves higher dynamics, stronger sensory immersion, and more complex multimodal cue integration. Third, few studies have compared how different types of semantic manipulation operate across textual, audio, and visual modalities within the same experimental framework. When the same topic content presents different

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

degrees of semantic bias or information manipulation in different modalities, the user's comprehensive judgment of the authenticity of the information may change significantly (Li et al., 2026).These gaps indicate the need to examine fake videos not only as technical artifacts but also as multimodal persuasive stimuli that may shape users' perceived trust.

Research Questions

Based on the above research background and research gaps, this paper introduces information manipulation theory into the video communication context, focusing on the influence mechanism of multimodal semantic manipulation on user trust perception. Specific research questions include:

- RQ1: In the context of video information dissemination, are there significant differences in user trust perception between different information modes (text, audio, and images)?
- RQ2:Do different types of information manipulation (fabrication, concealment, misinterpretation, ambiguity, incitement) have different effects on user trust?
- RQ3: How do users explain their trust judgments when exposed to different types of semantic manipulation across textual, audio, and visual modalities?

Research Contributions

The main contributions of this study are threefold. First, this study extends Information Manipulation Theory to the context of multimodal fake-video communication by operationalizing five semantic manipulation types across textual, audio, and visual modalities. Second, it shifts the analytical focus from algorithmic fake-video detection to users' modality-specific trust judgments, thereby complementing detection-oriented research with a user-cognition perspective. Third, by combining within-subject experimental evidence with interview-based explanations, the study provides practical implications for platform-level content governance and media-literacy interventions targeting authority-like cues, incomplete information, and emotionally charged expressions.

Materials and methods.

Theoretical Operationalization of Semantic Manipulation

Information Manipulation Theory explains deceptive communication as a process in which communicators strategically alter the quantity, quality, relevance, or manner of information. In the context of fake videos, such manipulation may appear not only in verbal statements but also in textual titles, audio narration, and visual frames. Based on the logic of Information Manipulation Theory and the characteristics of short-video misinformation, this study operationalizes semantic manipulation into five categories: fabrication, concealment, misinterpretation, ambiguity, and incitement. Fabrication refers to the creation of false or authority-like information; concealment refers to the omission of key information; misinterpretation refers to the distortion of original meaning; ambiguity refers to vague or unclear semantic presentation; and incitement refers to emotionally charged or provocative expression.

Hypothesis Development

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

·Text Semantic Manipulation And User Trust

Text is an important symbolic carrier for video communication, especially in video titles. The title not only highly summarizes the video content, but also triggers a series of information behaviors such as clicks, views, and comments on users. When users are exposed to new information, they often prioritize the reliability of the source and content direction conveyed by the title (Sun et al., 2023). Fictional official publishers (fabrications) may use the halo effect to mislead users (Kim et al., 2019); Concealing and obscuring key information will destroy the integrity of the information and raise doubts (Zarsky, 2003); Texts that incorporate inflammatory expressions directly intervene in the user's emotional tendencies and interfere with the judgment of their objective authenticity. Based on this, the following hypotheses are proposed:

H1a-H1e: Falsification, concealment, misinterpretation, ambiguity, and incitement of text in fake videos can significantly affect users' perceived trust.

·Audio Semantic Manipulation and User Trust

In this study, «audio» is defined as a speech expression that conveys the semantics of video content (excluding non-semantic elements such as background music and ambient noise floor). Compared with static text, audio can convey more delicate and rich emotions and details through speech speed and intonation, which is the core channel for users to understand video intentions. Especially in the face of information that is difficult to immediately falsify through common sense, the semantic guidance role of audio content will be sharply amplified. When audio information is added with irrelevant authoritative elements (fabrication), pinching (concealment), or emotional output (incitement), the user's cognitive defense line is highly vulnerable. Based on this, the following hypotheses are proposed:

H2a-H2e: Fabrication, concealment, misinterpretation, obscurity, and incitement of audio in fake videos can significantly affect users' perceived trust.

·Image Semantic Manipulation and User Trust

Images in video refer specifically to dynamically changing frames. Dynamic visual elements capture the user's visual attention in an instant due to their intuitiveness and vividness. Compared with pure graphics and text, dynamic pictures often give the audience the illusion of «seeing is believing» due to the more three-dimensional information dimension (Chang, 2023). Studies have shown that posts with rich visual elements are 11 times more viral than posts without visual elements. Therefore, after the images in fake videos are manipulated by splicing exaggeration (fabrication), missing keyframes (concealment), or visual impact (incitement), it will inevitably reshape users' evaluation criteria for the overall authenticity of the video. Based on this, the following hypotheses are proposed:

H3a-H3e: Falsification, concealment, misinterpretation, ambiguity, and incitement of images in fake videos can significantly affect users' perceived trust.

Variable measurement

The core dependent variable of this study is perceived trust, which is used to measure users' subjective evaluation of the authenticity and trustworthiness of video information after watching it. Referring to the mature scale in the existing information

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

credibility and media trust research, and adjusting appropriately in combination with the video information dissemination situation, this study uses the Likert 7-point scale for measurement, where «1» means «completely disagree» and «7» means «completely agree». User-perceived trust consists of the following three metrics:

1. I think the content in the video is real.
2. I think the content in the video is persuasive.
3. I think the content in the video is trustworthy.

After watching each experimental video, the subjects rated the above questions according to their own perception, and the average of the three measurement items was used as a comprehensive index of user perceived trust.

To examine the preliminary reliability of the perceived-trust scale, a pilot test was conducted with 20 participants before the formal experiment. The pilot sample produced a Cronbach's α of 0.87, which exceeded the commonly accepted threshold of 0.70 and indicated acceptable internal consistency. Exploratory factor analysis further showed that the KMO value was 0.82 and Bartlett's test of sphericity was significant ($p < 0.001$), indicating that the three-item scale was suitable for subsequent statistical analysis. It should be noted that this pilot reliability coefficient was calculated from the pilot sample, whereas the reliability coefficient reported in the Results section was calculated from the formal experimental sample.

Experimental Design and Stimulus Preparation

This study adopted a mixed-methods strategy combining a within-subject experiment, questionnaire measurement, and semi-structured interviews. The experimental design included three modalities—textual, audio, and visual information—and five manipulation types—fabrication, concealment, misinterpretation, ambiguity, and incitement—resulting in 15 manipulation conditions.

To reduce the potential influence of prior knowledge, three low-familiarity scientific and technological misinformation topics were selected: "substations produce radiation and noise affecting health," "a super cold wave is related to the triple La Niña phenomenon," and "current technology cannot realize face recognition between twins." Based on the original rumor-refuting materials for these topics, the textual title, audio narration, and visual frames were independently manipulated using professional audio and video editing software. When one modality was manipulated, the other modalities were kept unchanged to ensure single-variable control. In total, 48 experimental videos were produced: 3 topics × (1 original false video + 15 manipulated versions).

Participants were recruited via online advertisements on university forums and social media platforms.

Manipulation Check and Stimulus Validation

To verify whether each semantic manipulation was perceived as intended, a manipulation check was conducted before the formal experiment. Twenty participants who did not take part in the formal experiment viewed the stimulus videos and identified the perceived manipulation type. The check items were designed according to the operational definitions of fabrication, concealment, misinterpretation, ambiguity, and incitement. For example, participants were asked whether the title appeared to

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

contain an authority-like source cue for fabrication, whether key information appeared to be omitted for concealment, and whether the audio narration contained emotionally charged language for incitement. The recognition accuracy of all manipulation conditions exceeded 85%, suggesting that the experimental stimuli were sufficiently distinguishable and suitable for formal testing.

Participants and data collection

Participants were recruited via online advertisements on university forums and social media platforms. The recruitment criteria required that participants had active video-watching habits (at least three times per week) and had not been previously exposed to the three specific rumor-refuting topics used in this study ("substations produce radiation and noise affecting health", "super cold wave relates to triple La Niña phenomenon", and "current technology cannot realize face recognition between twins"). A screening questionnaire was administered to exclude those who reported prior exposure to these topics.

A total of 52 participants initially enrolled. After the experiment, 4 participants were excluded due to incomplete responses or failing attention-check questions (e.g., selecting a specific option as instructed), leaving 48 valid experimental datasets. Among the 48 valid participants, the proportion of men and women was balanced (50.0% each). The age distribution is dominated by young people aged 18–25 (accounting for 72.9%), and their education is concentrated at the undergraduate and master's levels. More than half (56.3%) of the subjects used video more than three times a day, representing a typical group of high-frequency active users. Because this study used a within-subject design, each participant evaluated all manipulation conditions, which improved statistical efficiency by reducing between-participant variance. A priori power considerations for paired comparisons indicated that a sample of approximately 34–41 participants would be sufficient to detect medium effects at $\alpha = 0.05$ and statistical power = 0.80. Therefore, the final valid sample of 48 participants was considered acceptable for detecting medium effects. Nevertheless, the sample size may still be insufficient for detecting small effects and for conducting more complex interaction analyses; this limitation is acknowledged in the Conclusion.

To reduce fatigue and order effects, the experimental procedure adopted block presentation, randomization, and interval rests. Each participant watched 48 videos divided into three topic-based blocks. Within each block, one original false video and 15 manipulated videos were presented in random order. The order of the three topic blocks was counterbalanced across participants using a Latin square design. After each block, participants took a mandatory 5-minute rest. The entire experimental session lasted approximately 90 minutes.

After the experiment, 24 participants who showed relatively high or low trust responses under specific manipulation conditions were selected for semi-structured interviews. The purpose of the interviews was to further explain the cognitive reasons behind the quantitative patterns rather than to generate statistically generalizable conclusions.

The interview data were analyzed using thematic analysis. Interview recordings

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

were transcribed and repeatedly read to identify recurring explanations of trust judgment. Initial codes were generated around perceived authority cues, information completeness, emotional tone, visual credibility, and attention allocation. These codes were then grouped into higher-level themes and compared with the quantitative results. Representative quotations were selected to illustrate the cognitive mechanisms behind the observed statistical patterns.

Results.

Descriptive statistics and reliability and validity tests

Among the 48 valid participants, the proportion of men and women was balanced (50.0% each). The age distribution is dominated by young people aged 18–25 (accounting for 72.9%), and their education is concentrated at the undergraduate and master's levels. More than half (56.3%) of the subjects used video more than three times a day, representing a typical group of high-frequency active users.

For the formal experimental sample (N = 48), the perceived-trust scale showed excellent internal consistency, with Cronbach's $\alpha = 0.956$. Composite reliability (CR) was 0.957, and the average variance extracted (AVE) was 0.880, indicating good convergent validity. The KMO value was 0.776, and Bartlett's test of sphericity was significant ($p < 0.001$), suggesting that the scale was suitable for factor-based validity assessment. These formal-sample results should be distinguished from the pilot reliability coefficient reported in the Methods section.

Research hypothesis testing

Because the same participants evaluated both the control condition and the corresponding manipulation conditions, paired-samples t-tests were used to compare perceived trust between each manipulation condition and its corresponding control condition. Exact p-values were reported, and Cohen's dz was calculated as the standardized effect size for within-subject comparisons. The hypothesis test results are shown in Table 1.

As shown in Table 1, 8 of the 15 hypotheses were supported under the two-tailed significance criterion. The results indicate that semantic manipulation affects users' perceived trust in a modality-specific manner rather than producing a uniform effect across all video modalities.

For the textual modality, fabrication significantly increased perceived trust compared with the control condition ($t = -5.214$, $p < 0.001$, $dz = -0.753$). By contrast, concealment ($t = 3.567$, $p = 0.001$, $dz = 0.515$), ambiguity ($t = 2.489$, $p = 0.016$, $dz = 0.359$), and incitement ($t = 3.415$, $p = 0.001$, $dz = 0.493$) significantly reduced perceived trust. Misinterpretation did not produce a significant difference ($t = 1.623$, $p = 0.111$). Therefore, H1a, H1b, H1d, and H1e were supported, whereas H1c was not supported.

For the audio modality, fabrication ($t = -6.942$, $p < 0.001$, $dz = -1.002$) and concealment ($t = -3.308$, $p = 0.002$, $dz = -0.477$) significantly increased perceived trust. Audio incitement significantly reduced perceived trust ($t = 3.122$, $p = 0.003$, $dz = 0.451$). However, audio misinterpretation and audio ambiguity did not produce significant effects. Therefore, H2a, H2b, and H2e were supported, whereas H2c and

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

H2d were not supported.

For the visual modality, fabrication significantly increased perceived trust (t = -3.456, p = 0.001, dz = -0.499). However, concealment, misinterpretation, ambiguity, and incitement in the visual modality did not reach statistical significance under the two-tailed criterion. Therefore, only H3a was supported, whereas H3b–H3e were not supported.

Table 1. Results of paired-samples t-tests for hypothesis testing

Modality	Hypothesis	Manipulation type	t	p	Cohen's dz	Result
Text	H1a	Fabrication	-5.214	<0.001	-0.753	Supported
Text	H1b	Concealment	3.567	0.001	0.515	Supported
Text	H1c	Misinterpretation	1.623	0.111	0.234	Not supported
Text	H1d	Ambiguity	2.489	0.016	0.359	Supported
Text	H1e	Incitement	3.415	0.001	0.493	Supported
Audio	H2a	Fabrication	-6.942	<0.001	-1.002	Supported
Audio	H2b	Concealment	-3.308	0.002	-0.477	Supported
Audio	H2c	Misinterpretation	-0.391	0.698	-0.056	Not supported
Audio	H2d	Ambiguity	-1.054	0.297	-0.152	Not supported
Audio	H2e	Incitement	3.122	0.003	0.451	Supported
Visual	H3a	Fabrication	-3.456	0.001	-0.499	Supported
Visual	H3b	Concealment	0.782	0.438	0.113	Not supported
Visual	H3c	Misinterpretation	0.319	0.751	0.046	Not supported
Visual	H3d	Ambiguity	0.897	0.374	0.129	Not supported
Visual	H3e	Incitement	1.893	0.065	0.273	Not supported

Paired-samples t-tests were used because the same participants evaluated the control and manipulation conditions. Cohen's dz indicates the within-subject standardized effect size. Negative t values indicate higher perceived trust in the manipulation condition than in the control condition, whereas positive t values indicate lower perceived trust in the manipulation condition than in the control condition. *p < 0.05, **p < 0.01, ***p < 0.001; n.s. = not significant.

Post-hoc test and demographic differences

Further differentiation analysis shows that gender factors play a role in information processing. Male users are more likely to have a negative impact on text blurring (t=-2.286, p<0.05), text incitement (t=-2.184, p<0.05), and audio concealment (t=-2.335, p<0.05) was significantly lower than that of female users. However, no significant differences were observed in the one-way ANOVA for the education level and age dimensions. These demographic results should be interpreted cautiously because the sample was concentrated among young and highly active video users. The absence of significant age and education effects may be related to the relatively homogeneous sample structure and the low public familiarity of the selected scientific and technological topics. Future studies should use larger and more heterogeneous samples to further examine demographic differences in fake-video trust judgments.

Discussion.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

The results indicate that user trust in fake videos is not shaped merely by the presence of manipulation, but by the specific combination of manipulation type and information modality. Fabricated authority-like cues in textual and audio modalities were associated with higher perceived trust, whereas emotionally inciting expressions tended to reduce trust, especially in textual and audio forms. These findings suggest that users do not process fake videos as a single integrated object; rather, they rely on modality-specific cues such as source-like textual signals, voice tone, semantic completeness, and visual salience.

Authority-Cue Heuristic and Logic-Based Resistance at the Textual Level

Textual fabrication produced a relatively strong positive effect on perceived trust. This result may be explained by an authority-cue heuristic: when fabricated titles contain official-looking publisher names, institutional labels, or source-like cues, users may transfer perceived credibility from the apparent source to the content itself. This interpretation is supported by the interview data. For example, participant P4 stated that an official-looking source made the video feel "more real." By contrast, textual concealment and ambiguity reduced trust because they disrupted information completeness and semantic clarity. Participants often described such videos as "incomplete," "unclear," or "arbitrary," suggesting that users were sensitive to logical gaps in textual information. Textual incitement also reduced perceived trust, likely because strongly emotional wording increased users' perception of sensationalism and persuasive intent.

Audio Coherence, Concealment, and Emotional Tone

The audio results show that both fabrication and concealment increased perceived trust, whereas incitement reduced trust. Audio fabrication may be persuasive because voice tone, rhythm, and formal narration can create an impression of professionalism or authority. Participant P12, for example, noted that the voice "sounded official," which increased trust in the video. The positive effect of audio concealment may reflect the temporal nature of audio processing. Because audio information is presented continuously and disappears quickly, users may find it difficult to notice omitted details unless they replay the content. As participant P10 indicated, the narration still sounded coherent even when key information was missing. In contrast, audio incitement reduced trust because drastic emotional tone and exaggerated wording made persuasive intent more salient.

Comparison with prior research

Our findings both align with and extend previous work in several ways. First, consistent with the technological detection literature (Liu et al., 2023; Al-Shamayleh et al., 2025), we confirm that multimodal features are critical carriers of deception. However, unlike most technical studies that focus on detection accuracy, our results reveal that users' trust judgments are not simply a function of whether manipulation is present, but rather vary systematically by modality and manipulation type. For example, fabrication in text and audio increases trust, whereas incitement decreases trust across all modalities. This nuanced pattern has not been highlighted in prior detection-oriented research.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Second, our finding that text-level ambiguity and concealment reduce trust supports the information manipulation theory (McCornack, 1992) in the video context, extending earlier applications in text-only social media (Tufchi et al., 2023). Yet, we also find that image-level concealment does not significantly affect trust, which diverges from the assumption that "seeing is believing" (Chang, 2023). This suggests that in dynamic video viewing, users may allocate limited attention unevenly across modalities, a cognitive mechanism rarely modeled in existing multimodal fake news detection frameworks (Xue et al., 2021).

Third, our qualitative interview data provide explanatory depth to quantitative patterns, echoing calls in recent literature to integrate user perspectives into disinformation research (Hao et al., 2025). The observed "authority superstition" effect for fabricated text and audio aligns with source credibility theory (Kim et al., 2019), but also highlights that users rarely verify sources when sensory cues appear official — a blind spot that technical detection alone cannot address.

Visual Evidence Dependence and Attention Allocation

Visual fabrication significantly increased perceived trust, suggesting that users may treat official-looking visual markers, watermarks, or scene elements as evidence of authenticity. However, visual concealment, misinterpretation, ambiguity, and incitement did not show significant effects in the revised statistical analysis. This pattern may indicate that subtle visual manipulations are less likely to be noticed when users' attention is simultaneously occupied by titles and audio narration. In short-video viewing, visual information may function as background evidence unless the manipulation is highly salient. Therefore, the visual modality should not be assumed to influence trust uniformly; its effect depends on the visibility and interpretability of the manipulated cue.

Conclusion.

This study examined how multimodal semantic manipulation in fake videos influences users' perceived trust. Drawing on Information Manipulation Theory, the study tested five manipulation types—fabrication, concealment, misinterpretation, ambiguity, and incitement—across textual, audio, and visual modalities. The results show that semantic manipulation affects perceived trust in a modality-specific manner. Textual fabrication and audio fabrication were associated with higher perceived trust, while textual concealment, textual ambiguity, textual incitement, and audio incitement reduced perceived trust. Visual fabrication also increased perceived trust, whereas the remaining visual manipulation types did not produce statistically significant effects under the revised analysis. The interview findings further suggest that authority-like cues, information completeness, emotional tone, and attention allocation help explain these modality-specific trust patterns.

The findings suggest that media-literacy interventions and platform-level warning mechanisms should pay particular attention to authority-like textual and visual cues, emotionally charged audio expressions, and incomplete or ambiguous information presentation. These implications are directly derived from the observed modality-specific trust patterns.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

This study has several limitations. First, the experimental scenario mainly referred to short-video communication, and the cross-platform generalizability of the findings requires further verification. Second, the study did not include real-video control conditions, which limits direct comparison between users' trust judgments of real and fake videos. Third, the experiment examined single-modality manipulation effects and did not test the superimposed or interactive effects of simultaneous multimodal manipulation. Fourth, although the within-subject design improved statistical efficiency, the sample size remains limited for detecting small effects and for conducting subgroup analyses. Finally, the sample was concentrated among young and highly active video users, which may limit the generalizability of the findings to broader populations. Future studies should expand the sample size, include real-video comparison conditions, and examine multimodal interaction effects in more diverse platform contexts.

REFERENCES

Al-Shamayleh, A. S., Riasat, H., Alluhaidan, A. S., Raza, A., El-Rahman, S. A., & AbdElminaam, D. S. (2025). Novel transfer learning based acoustic feature engineering for scene fake audio detection. Scientific Reports. — Vol. 15. — No. 1. — Pp. 8066. [in Eng].

Alsaeedi, A., AlMansour, A., & Jamal, A. (2025). Audio-visual multimodal deepfake detection leveraging emotional recognition // International Journal of Advanced Computer Science and Applications. — Vol. 16. — No. 6. — Pp. 22. [in Eng].

Chang, B. G. (2023). Seeing… What? Four Images and Their (In) Visibility // Cultural Critique. — Vol. 118. — No. 1. — Pp. 23-55. [in Eng].

El-Shafai, W., Fouda, M. A., & El-Salam, M. (2024). Deepfake forensics: a survey of digital forensic methods for multimodal deepfake identification on social media // Multimedia Tools and Applications. — Vol. 83. — Pp. 1–35. [in Eng].

Gandhi, K., Kulkarni, P., Shah, T., Chaudhari, P., Narvekar, M., & Ghag, K. (2024). A multimodal framework for deepfake detection // Journal of Electrical Systems. — Vol. 20. — Pp. 1–22. [in Eng].

Ginsca, A. L., Popescu, A., & Lupu, M. (2015). Credibility in information retrieval. foundations and trends® // in information retrieval. — Vol. 9. — No. 5. — Pp. 355–475. [in Eng].

Hao, X., Xu, W., Huang, X., Sheng, Z., & Yan, H. (2025). MFUIE: A fake news detection model based on multimodal features and user information enhancement // EAI Endorsed Transactions on Scalable Information Systems. — Vol. 12. — No. 1. — Pp. 7517. [in Eng].

Joshi, L. K., & Joshi, S. (2025). Deepfake detection using multimodal AI // International Journal of Research and Innovation in Applied Science. — Vol. 10. — No. 5. — Pp. 355–357. [in Eng].

Kim, A., Moravec, P. L., & Dennis, A. R. (2019). Combating fake news on social media with source ratings: The effects of user and expert reputation ratings // Journal of Management Information Systems. — Vol. 36. — No. 3. — Pp. 931–968. [in Eng].

Leszczuk, M., & Juszka, D. (2025). Unveiling the art of video enhancement: a comprehensive examination of content selection and sequencing for optimal quality in conventional and AR/VR environments // International Journal of Electronics and Telecommunications. — Pp. 1–14. [in Eng].

Li, G., Tang, D., Huang, J., Zhu, S., & Cao, J. (2026). Dual-scale adaptive attention-based Vision transformer with iterative refinement for clarity and consistency in multi-focus image fusion // Engineering Applications of Artificial Intelligence. — Vol. 163. — Article 112777. [in Eng].

Li, Z., Wang, M., Zhong, J., & Ren, Y. (2022). Improving the communication and credibility of government media in response to public health emergencies: analysis of tweets from the WeChat official accounts of 10 Chinese health commissioners // Frontiers in Public Health. — Vol. 10. — Article 900776. [in Eng].

Liu, X., Yu, Y., Li, X., & Zhao, Y. (2023). Mcl: multimodal contrastive learning for deepfake detection // IEEE Transactions on Circuits and Systems for Video Technology. — Vol. 34. — No. 4. — Pp. 2803–2813. [in Eng].

Reddy, N. R., Dinesh, P., Venkateswaralu, N., Rani, S., Kumar, A. V., & Rekha, D. (2025). Enhancing fake news detection: a multimodal approach integrating machine learning // International Journal of Latest Technology in Engineering, Management & Applied Science. — Vol. 14. — No. 3. — Pp. 272–281. [in Eng].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Sun, Y., & Liang, C. (2023). How headline strategies and news sources affect number of user clicks on WeChat. // Journalism Practice. — Vol. 17. — No. 8. — Pp. 1696–1721. [in Eng].

Tufchi, S., Yadav, A., & Ahmed, T. (2023). A comprehensive survey of multimodal fake news detection techniques: advances, challenges, and opportunities // International Journal of Multimedia Information Retrieval. — Vol. 12. — No. 2. — Pp. 28. [in Eng].

Wang, M. (2025). Deepfake detection: a multimodal survey. ITM Web of Conferences. — Vol. 78. — Article 02027. [in Eng].

Xi, D., Xu, W., Chen, R., Zhou, Y., & Yang, Z. (2021). Sending or not? A multimodal framework for Danmaku comment prediction // Information Processing & Management. — Vol. 58. — No. 6. Article 102687. [in Eng].

Xue, J., Wang, Y., Tian, Y., Li, Y., Shi, L., & Wei, L. (2021). Detecting fake news by exploring the consistency of multimodal data // Information Processing & Management. — Vol. 58. — No. 5. Article 102610. [in Eng].

Zarsky, T. Z. (2003). Thinking outside the box: considering transparency, anonymity, and pseudonymity as overall solutions to the problems in information privacy in the internet society. U. Miami L. Rev. — Vol. 58. — Pp. 991. [in Eng].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 258–278
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.014
UDC 004.89:004.94:637.1

ENSEMBLE AND HYBRID LEARNING FOR DIGITAL TWIN MILK CLASSIFICATION

Zh. Doumchariyev[1,3], J. Tussupov[1]*, A. Akynbekova[2], M. Khassanova[2], M. Kassim[4]
[1]L.N. Gumilyov Eurasian National University, Astana, Kazakhstan;
[2]Taraz University named after M.H Dulati, Taraz, Kazakhstan;
[3]Turan – Astana University, Astana, Kazakhstan;
[4]Universiti Teknologi MARA (UiTM), Shah Alam, Selangor, Malaysia.
E-mail: tussupov@mail.ru

Zhanagul Doumchariyeva — PhD Student, Senior Lecturer of the Department "Information Systems", L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
https://orcid.org/0009-0006-4085-8409;
Jamalbek Tussupov — Doctor of Physical and Mathematical Sciences, Professor of the Department "Information Systems", L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
E-mail: tussupov@mail.ru, https://orcid.org/0000-0002-9179-0428;
Aiman Akynbekova — Master, Senior Lecturer of the Department "Information Systems", Taraz University named after M.H Dulati, Taraz, Kazakhstan
https://orcid.org/0009-0003-0280-095X;
Munaram Khassanova — Master, Senior Lecturer of the Department "Applied informatics and programming", Taraz University named after M.Kh. Dulaty, Taraz, Kazakhstan
https://orcid.org/0009-0006-4833-4351;
Murizah Kassim — Professor, Head of Research and Innovation, Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM), Shah Alam, Selangor, Malaysia
https://orcid.org/0000-0002-8494-4783.

© Zh. Doumchariye, J. Tussupov, A. Akynbekova, M. Khassanova, M. Kassim

Abstract. Changes in milk composition and quality affect consumer health, production efficiency, and supply chain safety, which is why fast and reliable assessment is essential. Lately, there's increasing interest in digital models for food systems. However, the number of studies comparing ensemble, hybrid, and classical

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

machine learning models to directly classify milk quality is still limited. The aim of this study was to evaluate and compare the performance of classical, ensemble, and hybrid machine learning models for milk quality classification within a Digital Twin framework. The Kaggle Milk Quality dataset (1,059 samples, seven physicochemical attributes) was used. Six conventional algorithms, three ensemble methods (Random Forest, XGBoost, LightGBM), and three hybrid models (RF+KNN, XGBoost+RF, and Stacking) were assessed using stratified train–test splitting and five-fold cross-validation. The results showed that ensemble and hybrid methods consistently outperformed traditional machine learning methods. The RF+KNN hybrid model achieved perfect classification (100.0% across all metrics), while the Stacking hybrid model reached 99.69% due to a single misclassification. Both models substantially outperformed classical baselines. The findings indicate that ensemble and hybrid machine learning models are well suited for Digital Twin–based real-time milk quality monitoring and can support intelligent decision-making in smart dairy production.

Keywords: digital twin, milk quality classification, machine learning, ensemble learning, hybrid models

For citation: Doumchariye Zh., Tussupov J., Akynbekova A., Khassanova M., Kassim M. (2026). Ensemble and hybrid machine learning for real-time milk quality classification in digital twins // International journal of information and communication technologies. Vol. 7. No. 27. Pp. 258–278. https://doi.org/10.54309/IJICT.2026.27.3.014. (In Eng.).

Conflict of interest: The authors declare that there is no conflict of interest.

ЦИФРЛЫҚ ЕГІЗДЕГІ СҮТ САПАСЫН АНСАМБЛЬДІК ЖӘНЕ ГИБРИДТІ ОҚЫТУ АРҚЫЛЫ ЖІКТЕУ

Ж. Доумчариева[1,3], Д. Тусупов[1]*, А. Акынбекова[2], М. Хасанова[2], М. Кассим[4]
[1]Л. Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан;
[2]М.Х. Дулати атындағы Тараз университеті, Тараз, Қазақстан;
[3]«Тұран-Астана» университеті, Астана, Қазақстан;
[4]MARA технологиялық университеті (UiTM), Шах-Алам, Селангор, Малайзия.
E-mail: tussupov@mail.ru

Доумчариева Жанагул — магистр, Ақпараттық жүйелер кафедрасының PhD докторанты, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Қазақстан https://orcid.org/0009-0006-4085-8409;
Тусупов Джамалбек — физика-математика ғылымдарының докторы, Ақпараттық жүйелер кафедрасының профессоры, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан
E-mail: tussupov@mail.ru, https://orcid.org/0000-0002-9179-0428;
Акынбекова Айман — магистр, Ақпараттық жүйелер кафедрасының аға оқытушысы, М.Х. Дулати атындағы Тараз университеті, Тараз, Қазақстан https://orcid.org/0009-0003-0280-095X;

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Хасанова Мунарам — магистр, Қолданбалы информатика және бағдарламалау кафедрасының аға оқытушысы, М.Х. Дулати атындағы Тараз университеті, Тараз, Қазақстан
https://orcid.org/0009-0006-4833-4351;
Кассим Муризах — Зерттеу және инновациялар бөлімінің басшысы, Үлкен деректерді талдау және жасанды интеллект институтының профессоры (IBDAAI), MARA технологиялар университеті (UiTM), Шах Алам, Селангор, Малайзия
https://orcid.org/0000-0002-8494-4783.

© Ж. Доумчариева, Д. Тусупов, А. Акынбекова, М. Хасанова, М. Кассим

Аннотация. Сүттің құрамы мен сапасындағы өзгерістер тұтынушылардың денсаулығына, өндіріс тиімділігіне және жеткізу тізбегінің қауіпсіздігіне тікелей әсер етеді. Сондықтан сүт сапасын жедел әрі сенімді бағалау ерекше маңызды. Соңғы жылдары азық-түлік өндірісінде цифрлық егіз (Digital Twin) технологияларын қолдануға қызығушылық артып келеді. Дегенмен, сүт сапасын тікелей жіктеуде классикалық, ансамбльдік және гибридті машиналық оқыту модельдерін салыстырмалы түрде зерттеген жұмыстар әлі де аз. Зерттеудің мақсаты – Digital Twin тұжырымдамасы аясында сүт сапасын жіктеуге арналған классикалық, ансамбльдік және гибридті машиналық оқыту модельдерінің тиімділігін бағалау және салыстыру. Зерттеуде Kaggle платформасындағы Milk Quality деректер жиынтығы (1059 үлгі, жеті физика-химиялық көрсеткіш) пайдаланылды. Алты классикалық алгоритм, үш ансамбльдік әдіс (Random Forest, XGBoost, LightGBM) және үш гибридті модель (RF+KNN, XGBoost+RF және Stacking) стратификацияланған оқыту/сынақ жиынтықтары мен бес еселік кросс-валидация арқылы бағаланды. Нәтижелер ансамбльдік және гибридті әдістердің дәстүрлі машиналық оқыту модельдерінен тұрақты түрде жоғары нәтиже көрсеткенін дәлелдеді. RF+KNN гибридті моделі барлық көрсеткіштер бойынша 100,0% дәлдікпен мінсіз классификация нәтижесін көрсетті, ал Stacking ансамблі бір ғана қате классификацияға байланысты 99,69% нәтижеге жетті. Екі модель де классикалық базалық модельдерден айтарлықтай жоғары нәтиже көрсетті. Қорытындысында, ансамбльдік және гибридті машиналық оқыту модельдері Digital Twin негізіндегі нақты уақыттағы сүт сапасын бақылау және интеллектуалды шешім қабылдау жүйелерін құру үшін тиімді тәсіл болып табылады.

Түйінді сөздер: цифрлық егіз, сүт сапасын жіктеу, машиналық оқыту, ансамбльдік оқыту, гибридті модельдер

Дәйексөздер үшін: Доумчариева Ж., Тусупов Д., Акынбекова А., Хасанова М., Кассим М. (2026). Цифрлық егіздер негізінде сүт сапасын нақты уақыт режимінде жіктеуге арналған ансамбльдік және гибридті машиналық оқыту әдістері // Халықаралық ақпараттық және коммуникациялық технологиялар журналы. Т. 7. No 27. Б. 258–278. https://doi.org/10.54309/IJICT.2026.27.3.014. (Ағыл. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

жоқ деп мәлімдейді.

КЛАССИФИКАЦИЯ КАЧЕСТВА МОЛОКА В ЦИФРОВОМ ДВОЙНИКЕ АНСАМБЛЕВЫМИ И ГИБРИДНЫМИ МЕТОДАМИ

Ж. Доумчариева[1,3], Д. Тусупов[1]*, А. Акынбекова[2], М. Хасанова[2], М. Кассим[4]
[1]Евразийский национальный университет имени Л. Н. Гумилёва, Астана, Казахстан;
[2]Таразский университет имени М. Х. Дулати, Тараз, Казахстан;
[3]Университет «Туран-Астана», Астана, Казахстан;
[4]Университет технологий MARA (UiTM), Шах-Алам, Селангор, Малайзия.
E-mail: tussupov@mail.ru

Доумчариева Жанагул — магистр, PhD докторант кафедры Информационных систем, Евразийский национальный университет имени Л.Н. Гумилёва, Астана, Казахстан
https://orcid.org/0009-0006-4085-8409;
Тусупов Джамалбек — доктор физико-математических наук, профессор кафедры Информационных систем, Евразийский национальный университет имени Л.Н. Гумилёва, Астана, Казахстан
E-mail: tussupov@mail.ru, https://orcid.org/0000-0002-9179-0428;
Акынбекова Айман — магистр, старший преподаватель кафедры Информационных систем, Таразский университет имени М.Х. Дулати, Тараз, Казахстан
https://orcid.org/0009-0003-0280-095X;
Хасанова Мунарам — магистр, старший преподаватель кафедры прикладной информатики и программирования, Таразский университет имени М.Х. Дулати, Тараз, Казахстан
https://orcid.org/0009-0006-4833-4351;
Кассим Муризах — руководитель отдела Исследований и инноваций, профессор Института Анализа больших данных и искусственного интеллекта (IBDAAI), Технологический университет MARA (UiTM), Шах Алам, Селангор, Малайзия
https://orcid.org/0000-0002-8494-4783.

© Ж. Доумчариева, Д. Тусупов, А. Акынбекова, М. Хасанова, М. Кассим

Аннотация. Изменения состава и качества молока оказывают влияние на безопасность продукции, эффективность производства и здоровье потребителей, что делает актуальной разработку интеллектуальных методов оценки качества. Целью исследования являлось сравнение классических, ансамблевых и гибридных моделей машинного обучения для классификации качества молока в рамках концепции Digital Twin. В работе использован набор данных Milk Quality (Kaggle), содержащий 1059 образцов и семь физико-химических признаков. Были исследованы шесть классических алгоритмов, три ансамблевых ме-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

тода (Random Forest, XGBoost, LightGBM) и три гибридные модели (RF+KNN, XGBoost+RF и Stacking) с применением стратифицированного разделения выборки и пятикратной кросс-валидации. Результаты показали преимущество ансамблевых и гибридных моделей над традиционными алгоритмами. Гибридная модель RF+KNN продемонстрировала идеальную классификацию с результатом 100,0% по всем метрикам, тогда как ансамбль Stacking достиг 99,69% из-за единственной ошибочной классификации. Обе модели значительно превзошли классические базовые модели. Полученные результаты подтверждают перспективность использования ансамблевых и гибридных моделей в системах мониторинга качества молока на основе Digital Twin и интеллектуальной поддержки принятия решений.

Ключевые слова: цифровой двойник, классификация качества молока, машинное обучение, ансамблевое обучение, гибридные модели

Для цитирования: Доумчариева Ж., Тусупов Д., Акынбекова А., Хасанова М., Кассим М. (2026). Ансамблевые и гибридные методы машинного обучения для классификации качества молока в режиме реального времени на основе цифровых двойников // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 27. Стр. 258–278. https://doi.org/10.54309/IJICT.2026.27.3.014. (На англ.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Introduction.

Milk is one of the most important foods in the human diet. This is because the composition and nutritional value of milk is rich in proteins, fats, carbohydrates, vitamins and minerals. It plays a fundamental role for human health, especially for the growth and development of children, as well as for maintaining the health of adults. However, the quality of milk can deteriorate due to improper storage, lack of pasteurization and microbiological contamination. Drinking poor quality milk can lead to foodborne illnesses, bacterial infections, exposure to toxins, and long-term health complications. Therefore, a quick and reliable assessment of milk quality is an important issue in modern food safety systems (Sioneket al., 2023).

This study examines the interactions between food science, industrial automation, and Data-Driven Decision Support in the context of the Industry 4.0. The dairy industry, especially the production of fermented milk products, is facing an increasing need to introduce intelligent control systems that can reduce waste and operating costs while ensuring product quality. The integration of dual digital technologies with machine learning will allow predictive quality control, real-time monitoring and provide the necessary early warning capabilities for a modern intelligent production environment.

Traditionally, the quality of milk is checked in the laboratory through tests that examine its physical and chemical properties, as well as its microbiology. Although these methods give tangible results, they are often time-consuming, labor-intensive, expensive, and unsuitable for continuous monitoring in a production environment

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

(Aydogdu et al., 2023). New technologies such as advanced sensors, Internet of things and artificial intelligence will help us create intelligent systems capable of checking the quality of milk in real time. These technologies allow you to automatically check the quality of milk and quickly make decisions in the production of food products (Rajeswari et al., 2024). By combining sensors that can transmit data via the internet with artificial intelligence, we can now regularly check the quality of milk and achieve quick results even at large industrial enterprises. This is a big step forward in ensuring the safety and quality of our food (Rajeswari et al., 2024).

Milk quality problems have serious economic consequences: product recall, risk to consumer health, and closing production can lead to significant financial losses for dairy farmers. While traditional laboratory quality assessment methods are accurate, they work slowly and cannot provide the prompt feedback needed to manage the workflow. In recent years, advances in sensor technology, the Internet of Things (IoT), and artificial intelligence (AI) have opened up new opportunities for Automated Continuous Quality Control. However, integrating these technologies into real production environments remains a challenge that requires creating models that combine high accuracy and low computational latency.

One of the most promising developments in this area is the concept of Digi tal Twins. A Digital Twin is a copy of a physical system that always sends and receives information with the real thing through sensors and computer models (Huang et al., 2024). In factories and industries Digital Twin technology has shown it can be very useful for watching what is happening fixing problems before they start and helping people make decisions (Rasheed et al., 2025; Rasheed et al., 2020). In the food business this idea has become Digital Food Twins, which use sensor information, computer models and machine learning to check the quality of food guess what will happen and help people make decisions (Krupitzer et al., 2022; Verboven et al., 2020). Recent studies have also demonstrated the application of Digital Twin concepts for modeling fermentation kinetics and supporting intelligent operator decision-making in dairy production (Alimardanova et al., 2026; Doumcharuyeva et al., 2025).

Machine learning is becoming very important for checking the quality of milk and guessing what it will be like. Some people named Sharma and Kumar looked at all the research. Found that machine learning is a good way to check food quality because it is fast and accurate. They also found that using methods together works better than using just one old method (Liakos et al., 2024).

People participating in the study can divide their work into three groups. The first group uses such machine learning methods as reference vector machines, close neighbors, decision trees, and the naive Bayesian method. Some studies have shown that these methods allow you to correctly classify the quality of milk in more than 95% of cases. Solution trees and support vector machines are especially good for this. They work very well with milk quality data. Digital twin and digital food Twin technologies use these methods to achieve results. Machine learning and digital twin technologies help the food industry make decisions and improve milk quality (Bhavsar et al., 2023; Zhao, 2025). The second category consists of ensemble learning methods, including

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Random Forest, Extreme Gradient Boosting, Light Gradient Boosting Machine, and Gr adient Boosting approaches. When you combine several simple models, you can create even more advanced ones. This is the so-called an ensemble Model. This can make the forecast more accurate and reliable. One of the most popular types of Group models is XGBoost. It is fast and can be used in different situations (Chen et al., 2016). Some studies have shown that the use of random forest and extreme gradient enhancement models can be really effective in predicting milk quality and analyzing dairy products. For example, such models will help determine whether milk is good or not, what it contains. By combining multiple models, you can better understand the data and make more accurate predictions. This is important in many areas, such as food safety and quality control (Mhapsekar et al., 2025; Goyal et al., 2024).

The third category includes hybrid machine learning models that combine the advantages of several algorithms. Random Forests and k-Nearest Neighbors showed really good results. These methods can be combined with other methods, for example, deep learning, the creation of more powerful tools. In fact, some studies have shown that these hybrid models can be 99% accurate, which is very important for systems such as food management systems. However, there is still disagreement about which approach is better (Keong et al., 2024; Rao et al., 2025). Some people consider deep learning to be the right path because it can be really accurate. Others believe that simple models, such as assembly methods, are better because they are faster and easier to understand. This is especially important for things like the Internet of Things, where speed and clarity are important (Neto et al., 2019; Ehsani et al., 2025). A great example of this is the dairy industry. With the development of the IoT and artificial intelligence, there is a great need for models that can combine accuracy and computational efficiency. This means that the models must be fast, but at the same time be able to explain what you are doing. Most of the research done so far has focused on obtaining correct predictions, but it has not paid enough attention to things like real-time constraints and the ability to interpret models (Alimardanova et al., 2026). It is important not only to get the right answer, but also to get it quickly and understand why. This is where Digital Twin architectures come to the rescue-they help to integrate these models into the big picture and make them more useful. Therefore, despite the fact that we have made some progress, there is still a lot of work to be done to create accurate and practical models for use in the real-world.

Despite these advances, several important research gaps remain. First, we need to look at how different machine learning models work to determine the quality of milk. We need to see if these models can work together and work with models. Secondly, in most cases, people who study this do not think that everything should happen quickly, whether these models can work in real life or not. Third, we must understand how machine learning models work and be able to see what is happening inside them. This is very important if we want to use them in life. Finally, we need to figure out how to use machine learning models in Digital Food Twin systems. We should also consider how to share models to check the quality of food. It still works. This is important for the future of food systems (Doumchariyeva et al., 2026).

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Based on these observations, the research hypothesis of this study shows that hybrid and group machine learning models can significantly improve the accuracy of milk quality classification while maintaining appropriate computational performance for real-time digital twin applications. In addition, it is assumed that some complex methods can provide a good balance between prediction accuracy and output speed, which makes it more suitable for the Industrial Internet of things.

The main goal of this study is to compare the effectiveness of Classical, Group and hybrid machine learning models to classify milk quality within the framework of the digital twin concept. To achieve this goal, a publicly available milk quality data set consisting of 1,059 samples and seven quality-related characteristics was analyzed. Six classical machine learning algorithms, three cluster models (Random Forest, Extreme Gradient Boosting, and Light Gradient Boosting Machine) and three hybrid models (Random Forest plus k-Nearest Neighbors, Extreme Gradient Boosting plus Random Forest, and Stacking) were evaluated.). The results obtained provide practical recommendations for the selection of suitable machine learning models and Digital Food Twin applications for double food for real-time monitoring of milk quality. The results showed that the overlapping cluster system achieved the highest classification accuracy, and the Light Gradient Boosting Machine showed the fastest prediction delay of 1.89 MS, which is a useful solution for delay-sensitive industrial systems.

Materials and methods.

Dataset Description

The empirical analysis is based on a collection of publicly available data for predicting the quality of milk hosted on the Cagle platform (https://www.kaggle.com/datasets/cpluzshrijayan/milkquality). The collection includes 1,059 milk samples, each characterized by seven physico-chemical and organoleptic indicators that are often used in assessing the quality of milk: pH, temperature, taste, aroma, fat content, turbidity and color. The score variable encodes quality scores as one of three ordinal categories — Low, Medium, and high. This data set was selected on the basis of a heterogeneous combination of quantitative and categorical characteristics reflecting the main indicators of milk quality and their role recognized as a reference in the literature on the classification of food quality.

The characteristics of the input data, as well as the data types and their range of values, are presented in Table 1.

Table 1. Input characteristics of milk quality data sets

Feature	Type	Range / Values
pH	Numerical	3.0–9.5
Temperature (°C)	Numerical	34–90
Taste	Categorical	0 (unpleasant), 1 (pleasant)
Odor	Categorical	0 (unpleasant), 1 (pleasant)
Fat	Categorical	0, 1
Turbidity	Categorical	0, 1
Color	Numerical	240–255

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Data Preprocessing

Before training the model, the data set went through several stages of pre-processing to improve the quality of the data and ensure reliable operation of the model. First, duplicate records were detected and deleted. With the help of statistical analysis, abnormal observations were identified that can negatively affect Model training.

Categorical variables were encoded using One- Hot Encoding to convert them into machine-readable digital representations. In order to ensure that all variables make the same contribution during model training, quantitative characteristics were normalized using the StandardScaler method.

Data Partitioning and Validation Strategy

The data sets were divided into training and testing subgroups in a 70: 30 ratio. The training subgroup was used to develop the model and adjust the Superparameters, while the testing subgroup was reserved for final performance assessments.

To increase the reliability of the results, five-fold validation was used during model training. This approach reduces the risk of overconformity and provides a more representative assessment of the effectiveness of Model generalization.

Digital Twin Architecture

The proposed research framework was designed for integration into a Digital Food Twin environment. The architecture combines sensor-based data acquisition, machine learning prediction, and intelligent monitoring functionalities.

The workflow consists of the following stages:

IoT Sensors → Feature Extraction → Machine Learning Classifier → Digital Twin → Dashboard → Process Optimization

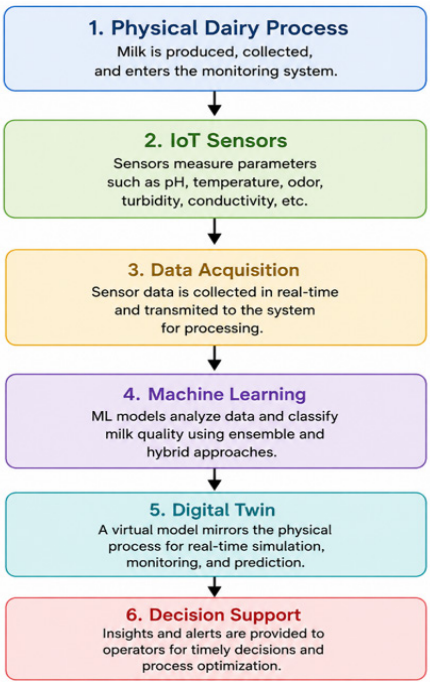


Fig. 1. General conceptual framework of Digital Twin–based milk quality monitoring.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Figure 1 shows the proposed Digital Twin framework for milk quality monitoring and optimization. Real-time data collected from IoT sensors are preprocessed through feature extraction techniques and analyzed using machine learning classifiers. The prediction results are integrated into the Digital Twin environment, where the virtual model is continuously updated. The generated information is visualized through a dashboard and used to support monitoring, quality assessment, and process optimization decisions.

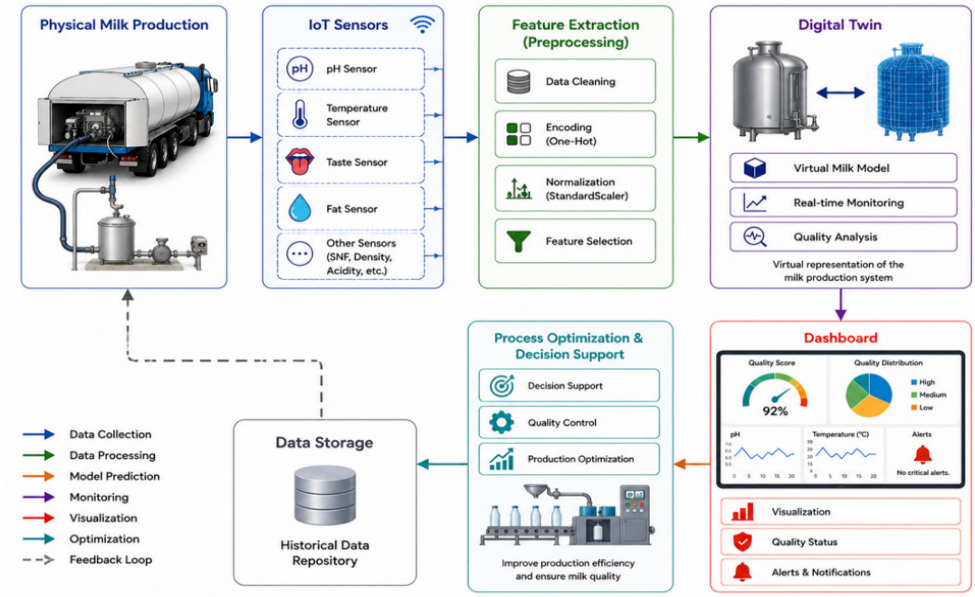


Fig. 2. Proposed Digital Twin architecture integrating machine learning models for real-time milk quality monitoring and decision support.

Figure 2 illustrates the proposed six-stage Digital Twin architecture for real-time milk quality monitoring. IoT sensors acquire physicochemical milk parameters, which are then preprocessed and classified by ensemble and hybrid machine learning models (Random Forest, XGBoost, LightGBM, and Stacking) to predict quality grades. The predictions are transferred to the Digital Twin environment, continuously updating the virtual representation for real-time monitoring. An interactive dashboard visualizes quality status and alerts, while the Process Optimization module translates analytical insights into quality control actions. Historical data are stored and fed back to support continuous model refinement.

Machine Learning Models

Three categories of machine learning models were investigated in this study: classical models, ensemble models, and hybrid models.

Classical Machine Learning Models

The baseline models included:

- Logistic Regression;
- k-Nearest Neighbors (k = 5);

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

- Decision Tree;
- Support Vector Machine with Radial Basis Function kernel;
- Gaussian Naïve Bayes;
- Multi-Layer Perceptron Neural Network.

These models were selected because they are commonly used benchmarks in food quality classification research.

Ensemble Learning Models

Although Random Forest, XGBoost, and LightGBM are commonly treated as individual algorithms, each internally combines multiple decision trees to generate a final prediction. Therefore, these approaches belong to the ensemble learning family, where multiple weak learners are aggregated to improve classification accuracy and robustness.

Three ensemble methods were evaluated:

- Random Forest (100 trees);
- Extreme Gradient Boosting (100 boosting iterations);
- Light Gradient Boosting Machine (100 boosting iterations).

Hybrid Models

Unlike ensemble methods, which combine multiple learners within the same algorithmic framework, hybrid models integrate two or more distinct machine learning algorithms. This approach allows the exploitation of complementary strengths of different methods and may improve predictive performance in complex classification tasks.

To investigate whether combinations of algorithms can outperform individual models, three hybrid approaches were developed:

RF + KNN, where feature importance scores generated by Random Forest were incorporated into the k-Nearest Neighbors classifier;

XGBoost + RF, combining gradient boosting and bagging principles;

Stacking Ensemble, using Random Forest, XGBoost, and Support Vector Machine as base learners and Logistic Regression as the meta-learner.

The hybrid models were designed to exploit complementary strengths of individual algorithms and improve classification performance within Digital Twin environments.

In addition, the Area Under the Receiver Operating Characteristic Curve (ROC-AUC) was used to evaluate the overall discriminative ability of the models independently of the classification threshold.

The region of the functioning characteristics of the receptors under the winter (rock-OK) is faking a «one-to-one» strategy to evaluate the multi-layer central recognition cabinet. The calculation efficiency was measured using:

- Exercise time (in seconds);
- Prediction time (milliseconds per 1000 samples).

These indicators are very important for digital twin systems, where real-time performance and low latency are the main requirements.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

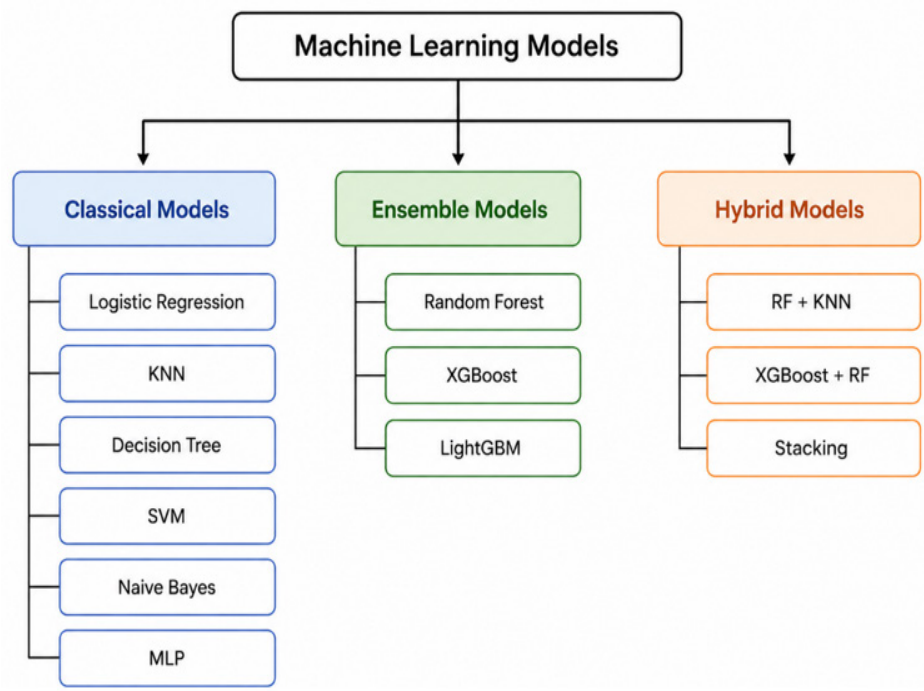


Fig. 3. Classification of classical, ensemble, and hybrid machine learning models used in this study.

Figure 3 shows the classification of machine learning models used in this study. Models are divided into three groups: Classic ensemble and hybrid methods, after which they are compared in terms of the effectiveness of the Milk Quality Rating.

Performance Evaluation Metrics.

The effectiveness of the model was evaluated using several additional evaluation indicators. Accuracy was used as the primary evaluation metric:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

To provide a comprehensive evaluation of the classification models, Precision, Recall, and F1-score were also calculated:

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision+Recall} \quad (4)$$

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Experimental Environment

All experiments were conducted using a workstation running Microsoft Windows 11 with an Intel Core i7 processor and 16 GB RAM.

Python 3.10, Scikit-learn 1.2.2, XGBoost 1.7.6, LightGBM 4.0.0, NumPy, Pandas, Matplotlib were used for the software environment.

The experimental installation is designed in such a way as to provide the possibility of reproduction and an objective comparison of all studied machine learning models.

The source code used to train and evaluate models is available at the reasonable request of the respective author.

Results and discussion.

Comparative Performance of Machine Learning Models

The effectiveness of the studied machine learning models was evaluated using the indicators of Accuracy, Precision, Recall, F1-score, and ROC-AUC. The results are presented in Table 2.

Table 2. Comparative efficiency of machine learning models to classify milk quality

Category	Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Hybrid	RF+KNN	1.0000	1.0000	1.0000	1.0000	1.0000
Hybrid	Stacking	0.9969	0.9969	0.9969	0.9969	1.0000
Hybrid	XGBoost+RF	0.9969	0.9969	0.9969	0.9969	1.0000
Ensemble	XGBoost	0.9969	0.9969	0.9969	0.9969	1.0000
Ensemble	Random Forest	0.9969	0.9969	0.9969	0.9969	1.0000
Ensemble	LightGBM	0.9969	0.9969	0.9969	0.9969	1.0000
Classical	SVM (RBF)	0.9969	0.9969	0.9969	0.9969	1.0000
Classical	Decision Tree	0.9969	0.9969	0.9969	0.9969	0.9969
Classical	KNN (k=5)	0.9937	0.9937	0.9937	0.9937	1.0000
Classical	MLP	0.9277	0.9284	0.9277	0.9284	0.9970
Classical	Naïve Bayes	0.8365	0.8387	0.8365	0.8387	0.9750
Classical	Logistic Regression	0.8239	0.8272	0.8239	0.8272	0.9270

As shown in Table 2, the RF+KNN hybrid model achieved perfect classification on the test set, with 100.0% across all five metrics (Accuracy, Precision, Recall, F1-score, and ROC-AUC). The Stacking model, XGBoost+RF, XGBoost, Random Forest, LightGBM, and SVM (RBF) each achieved 99.69%, reflecting a single misclassification out of 318 test samples. The Decision Tree and KNN (k=5) reached 99.69% and 99.37%, respectively. Among classical classifiers, SVM and Decision Tree were the strongest, while Logistic Regression (82.39%) and Naïve Bayes (83.65%) showed the lowest accuracy. Overall, hybrid and ensemble models consistently outperformed the classical baselines, confirming the benefit of combining multiple learning algorithms for milk quality classification.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Table 3. Average accuracy of classification by model class

Category	Models	Mean Accuracy
Classical	6 models	92.93%
Ensemble	3 models	99.69%
Hybrid	3 models	99.79%

As shown in Table 3, the average classification accuracy increased from 92.93% for basic models to 99.69% for group methods and 99.79% for hybrid models. The results show that combining multiple learning algorithms significantly increases the effectiveness and reliability of classification. In particular, hybrid methods have achieved the highest overall accuracy, demonstrating their potential for for Digital Twin–based quality control based on dual technologies and decision support applications.

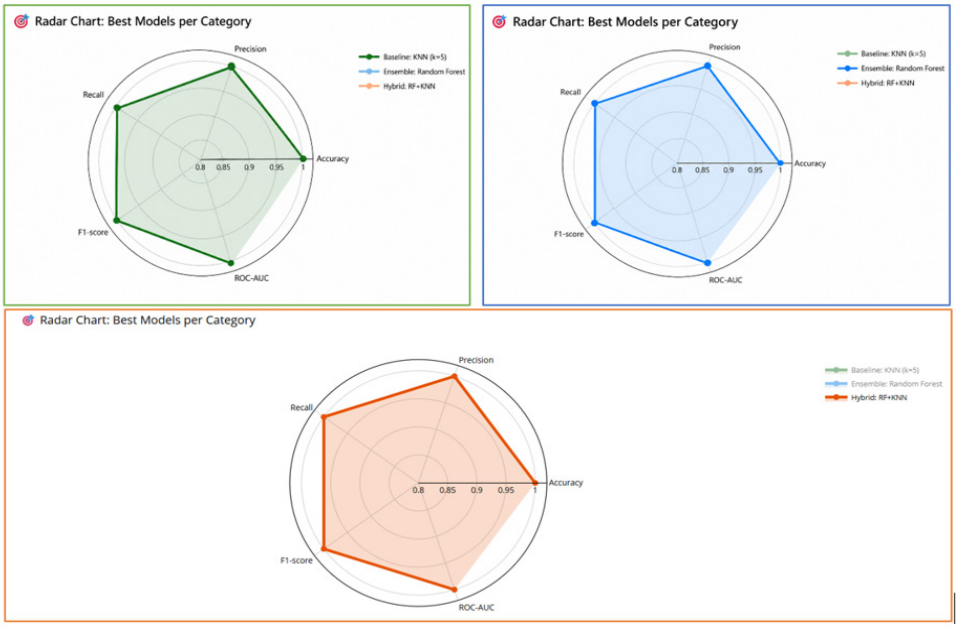


Fig. 4. Radar chart comparison of KNN, Random Forest, and RF+KNN models across five classification metrics.

Figure 4 presents radar chart visualizations comparing the best-performing model from each category across five classification metrics: Precision, Accuracy, ROC-AUC, F1-score, and Recall. The Ensemble model (Random Forest) demonstrates a consistently balanced and high-coverage profile across all metrics. The Hybrid model (RF+KNN) achieves the largest area coverage overall, reflecting superior performance particularly in Accuracy and ROC-AUC. In contrast, the Baseline model (KNN, k=5) exhibits a notably reduced profile in ROC-AUC and F1-score, indicating comparatively weaker generalization. These visualizations confirm that ensemble and hybrid approaches outperform the classical baseline across all evaluated dimensions.

Computational Performance

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

The time of training and forecasting of the most effective models is given in Table 4.

Table 4. Learning time and prediction of effective machine learning models

Model	Training Time (s)	Prediction Time (ms/1000 samples)
Logistic Regression	0.03	0.20
KNN (k=5)	0.01	0.50
Decision Tree	0.02	0.30
SVM (RBF)	2.84	19.60
Naive Bayes	0.01	0.15
MLP	0.50	0.40
Random Forest	0.39	2.98
XGBoost	0.42	2.15
LightGBM	0.21	1.89
RF+KNN	0.08	1.50
XGBoost + RF	1.15	4.20
Stacking	3.68	28.40

As shown in Table 4, XGBoost showed an estimated time of 2.15 ms per 1000 samples, followed by the speed of LightGBM (1.89 Ms), RF+KNN (1.50 ms) and Random Forests (2.98 ms). Among the models evaluated, the SVM (RBF) showed the highest computational cost, requiring 2.84 seconds to perform and 19.60 Ms per 1,000 samples to predict. The Stacking and XGBoost+RF also require additional computing resources compared to individual range models. These results show that XGBoost and LightGBM provide an effective balance between computational efficiency and classification performance.

The violin diagrams in Figure 5 show that ensemble and hybrid models consistently achieve high F1-scores with low contrast, which confirms their stability and reliability. In contrast, key models, especially Logistic Regression and the Naive Bayes method, show low estimates and low predictive power.

Confusion Matrix Analysis

To provide a more detailed assessment of classification performance, confusion matrices were constructed for three representative models (Figure 6). The RF+KNN hybrid model achieved perfect classification, with zero misclassifications across all 318 test samples. The Stacking model produced one misclassification (one high-quality sample classified as medium-quality), corresponding to 99.69% accuracy. The Logistic Regression baseline exhibited substantial confusion between adjacent classes, with 56 misclassifications, reflecting its lower accuracy of 82.39%.

These results confirm that the perfect performance of the RF+KNN model is not an artifact of metric calculation but reflects the genuine separability of the three milk quality classes in the feature space when heterogeneous base learners are combined. The single misclassification of the Stacking model is consistent with its 99.69% accuracy reported in Table 2.

The stability and robustness of the evaluated models were further assessed using

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

cross-validation. The distribution of F1-scores across validation folds is presented in Figure 5.

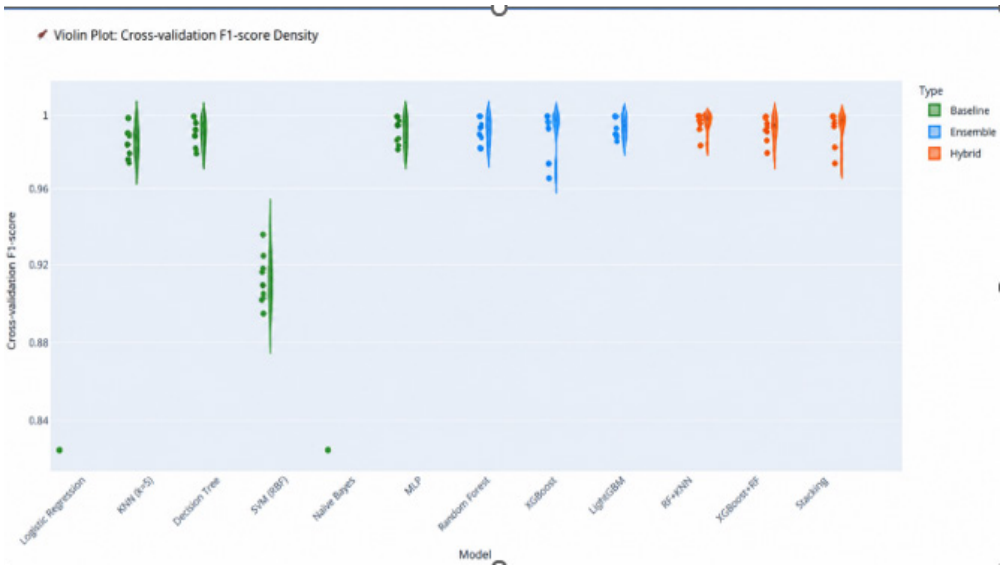


Fig. 5. Violin plot of cross-validation F1-score distributions for the evaluated machine learning models.

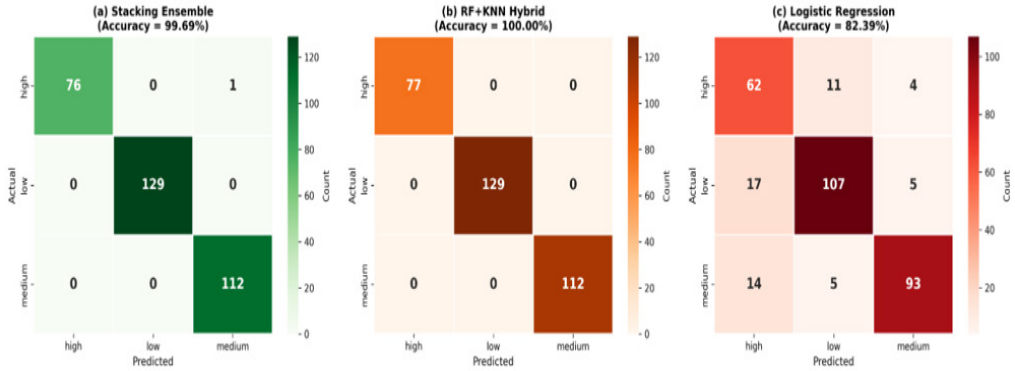


Fig. 6. Confusion matrices for three representative models.

(a) RF+KNN hybrid (Accuracy = 100.0%, 0 misclassifications); (b) Stacking model (Accuracy = 99.69%, 1 misclassification); (c) Logistic Regression baseline (Accuracy = 82.39%, 56 misclassifications). The matrices confirm the perfect classification of the RF+KNN model and illustrate the progressive degradation of performance in the Stacking and baseline models.

Digital Twin Deployment Readiness

A heat map has been developed to provide a visual comparison of the performance of the model for several evaluation indicators. The results are shown in Figure 7.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

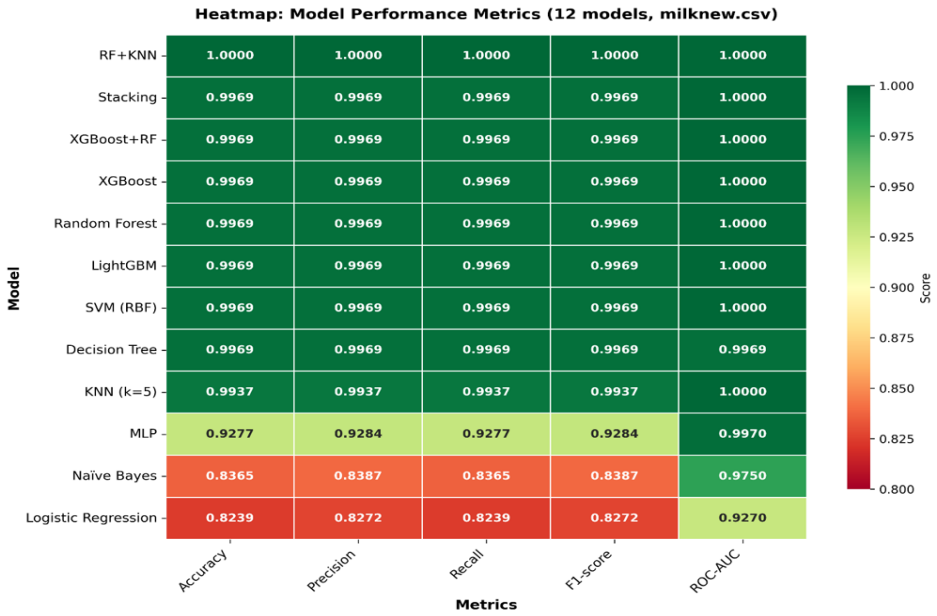


Fig. 7. Heatmap visualization of classification performance metrics for the evaluated machine learning models.

The heatmap shows that RF+KNN achieved the highest overall performance, with 1.0000 across all evaluated metrics, while Stacking and the other ensemble models also demonstrated near-perfect performance.

Interpretation of Classification Results

The high performance of the stacking model is due to its ability to combine the advantages of several classifiers. It is implemented at the meta-learning level. By combining predictions from different listeners, the model identifies complementary patterns in the data set and increases the reliability of the classification. This finding is consistent with previous studies (Liakos et al., 2025; Keong et al., 2024; Rao et al., 2025), which stated that integrated and hybrid methods often outperform individual classifiers in predicting food quality. In particular, our results are similar to those of Sharma and Kumar (Liakos et al., 2025), who showed that ensemble learning methods consistently provide high accuracy in food quality control applications.

Readiness to work with the Digital Twin and the ability to use it in Real-Time.

Digital twin systems provide continuous monitoring and intelligent decision-making. Digital dual systems require high prediction accuracy and minimal calculation delay (Su et al., 2025). As noted by Su et al. (Su et al., 2025), real-time data processing and predictive analysis are an important component of industrial dual digital architecture. In conclusion, among the high-performance models, the RF + KNN model has the lowest prediction delay, which is only 1.50 Ms per 1000 samples. The prediction time for LightGBM and XGBoost models is 1.89 MS and 2.15 ms, respectively. These results are consistent with the conclusions of Rasheed et al. (Rasheed et al., 2020). They particularly emphasize the importance of computational efficiency in implementing a Digital Twin in a production environment. Although the Stacking model showed

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

near-perfect classification performance, it required significantly more computational resources, resulting in a prediction delay of 28.40 ms.

The comparison of the evaluated models indicates a clear trade-off between predictive accuracy and computational efficiency. Although the Stacking model achieved near-perfect classification performance, its inference time was longer because predictions must be generated sequentially by multiple base learners and a meta-learner. In contrast, the RF+KNN hybrid model exhibited the lowest prediction latency, as Random Forest reduces feature complexity before the KNN classification stage, enabling faster inference while maintaining high accuracy. LightGBM also demonstrated an effective balance between prediction accuracy and computational efficiency due to its histogram-based learning and leaf-wise tree growth strategy. These findings suggest that the selection of a machine learning model for Digital Twin applications should consider not only predictive performance but also real-time computational requirements.

Comparison with Previous Studies

The results obtained in this study are consistent with previous studies. This study also demonstrated the effectiveness of group training in predicting and evaluating food quality, as in this case. Previous research has shown that Random Forest and XGBoost often outperform basic machine learning methods due to their ability to detect nonlinear relationships and complex interactions of objects (Liakos, 2025; Bhavsar et al., 2023; Mhapsekar et al., 2025). In this study, similar results were obtained: Ensemble and hybrid models consistently surpassed Logistic Regression, naive bays, MLB, KNN methods, decision-making trees.

In particular, a study by Vidal and Franco confirms the 99.69% accuracy obtained using the zgpost and litgbm methods (Mhapsekar et al., 2025), showing that gradient enhancement methods provide excellent predictive properties in tasks related to assessing milk quality. Our results also confirm the observations of Singh and Goyal (Alimardanova et al., 2026), who stressed the importance of computational efficiency in IoT-based dairy tracking applications.

In addition, the excellent results obtained using the Stacking model in this study (99.69% accuracy) confirm the latest results of Kumar and Sharma (Doumchariyeva et al., 2026). These researchers showed that combining heterogeneous learning patterns using the structure of the Stacking model can improve the accuracy and reliability of predictions in food quality classification tasks. However, unlike many previous studies that focused solely on classification efficiency, this study also evaluated computational efficiency and ease of use of Digital Twins. This means that it provides a comprehensive assessment of the industrial implementation in real-time.

The main objective of this study is to systematically compare the prediction latency of classical, ensemble, and hybrid models. Although the Stacking model provides near-perfect classification, its prediction time is long—28.40 ms per 1000 samples. This makes it less suitable for real-time IoT applications. In contrast, the prediction time of the RF+KNN model is 1.50 ms, the LightGBM model is 1.89 ms, and the XGBoost model is 2.15 ms, providing a good balance between accuracy and

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

speed. This meets the practical requirements of digital twin systems (Ehsani et al., 2025; Alimardanova et al., 2026).

Practical Implications

The results of the study provide practical recommendations on intelligent milk quality control systems. For applications where both classification accuracy and computational efficiency are critical, the RF+KNN model offers the most attractive solution, achieving 100.0% across all evaluated metrics with a prediction latency of only 1.50 ms per 1000 samples. In addition, the LightGBM and XGBoost models offer alternative deployment solutions in IoT and peripheral computing space. These recommendations contribute to the creation of effective digital doubles of nutrition. They support real-time quality assessment, process management, and intelligent decision-making in dairy production systems.

Research Limitations and Future Directions

Regardless of the expected results, it is worth noting a number of limitations. Most importantly, the study was conducted using a publicly available data set on the quality of Kaggle milk, which may not fully reflect the observed variability in industrial milk production. In addition, the data set contains relatively accurate and structured measurement results. Real-time transmitted IoT sensor data often includes interference, missing values, calibration errors, and connection latency.

In addition, the studied models were evaluated in stable conditions. They did not take into account the conceptual drift or online learning scenarios often used in long-term industrial implementations. Therefore, future research should focus on testing the proposed approach using sensory data collected in real-time in dairies. Additional features include the adaptive structure of digital twins, online learning algorithms, mechanisms for detecting model deviations, and the application of the proposed structure to kefir and other traditional dairy products.

Overall, the results show that collective and hybrid machine learning models are a viable solution. They emphasize that digital food products provide a solid foundation for the twin system, and can significantly improve automated milk quality control and support intelligent decision-making.

These findings are consistent with previous studies on Digital Twin-based fermentation modeling and Markov chain-based decision support, which demonstrated the feasibility of integrating predictive models into intelligent dairy production systems (Alimardanova et al., 2026; Doumcharieva et al., 2026).

Conclusion.

This study presented a Digital Twin-based framework for real-time milk quality classification by integrating classical, ensemble, and hybrid machine learning models into an intelligent decision support system.

The experimental results demonstrated that hybrid and ensemble models consistently outperformed classical baselines. The RF+KNN hybrid achieved perfect classification (100.0% across all metrics) with the lowest prediction latency (1.50 ms), offering the most attractive balance for real-time Digital Twin applications. The Stacking model reached 99.69% but required higher computational resources (28.40

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ms), while LightGBM provided an efficient ensemble alternative (1.89 ms) for edge computing environments.

The proposed framework provides a practical foundation for intelligent quality monitoring in smart dairy production and supports Digital Twin-driven decision making. Future research will focus on validating the proposed architecture using real industrial sensor data and deploying it within an operational Digital Twin environment.

The proposed framework extends previous research on Digital Twin modeling and intelligent decision support for dairy fermentation by applying ensemble and hybrid machine learning methods to real-time milk quality classification (Alimardanova et al., 2026; Doumchariyeva et al., 2026).

REFERENCES

Aydogdu, T., O'Mahony, J. A., & McCarthy, N. A. (2023). pH, the Fundamentals for Milk and Dairy Processing: A Review // Dairy. — Vol. 4. — No. 3. — Pp. 395–409. DOI: https://doi.org/10.3390/dairy4030026. [In Eng.].

Alimardanova, M., Doumchariyeva, Z., Zhumakhan, N., Togzhanova, K., & Tlevlessova, D. (2026). Comparative Kinetic Modeling of Ayran Fermentation with Functional Additives // Fermentation. — Vol. 12. — No. 7. — Pp. 323. DOI: https://doi.org/10.3390/fermentation12070323. [In Eng.].

Bhavsar, D., Jobanputra, Y., Swain, N. K., & Swain, D. (2023). Milk Quality Prediction Using Machine Learning // EAI Endorsed Transactions on Internet of Things. — Pp. 10. DOI: https://doi.org/10.4108/eetiot.4501. [In Eng.].

Chen, T., & Guestrin, C. (2016). XGBoost. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. — Pp. 785–794. DOI: https://doi.org/10.1145/2939672.2939785. [In Eng.].

Doumchariyeva, Z., Tussupov, J., Sambetbayeva, M., Zhukabayeva, T., Yessenaliyeva, M., Kalemshariva, B., Issayev, S., Khassanova, M. (2026). Reliability reserve: A Markov chain-based metric for real-time operator decision support in ayran fermentation // Informatics. — Vol. 13. — No. 7. — Pp. 105. DOI: https://doi.org/10.3390/informatics13070105. [In Eng.].

Ehsani, S., Weller, P., & Parastar, H. (2025). Revolutionizing food integrity: Harnessing the power of ensemble learning algorithms // TrAC Trends in Analytical Chemistry. — Vol. 193. — Article 118458. DOI: https://doi.org/10.1016/j.trac.2025.118458. [In Eng.].

Goyal, K., Kumar, P., & Verma, K. (2024). XAI-empowered IoT multi-sensor system for real-time milk adulteration detection // Food Control. — Vol. 164. — Article 110495. DOI: https://doi.org/10.1016/j.foodcont.2024.110495. [In Eng.].

Huang, Y., Ghadge, A., & Yates, N. (2024). Implementation of digital twins in the food supply chain: a review and conceptual framework // International Journal of Production Research. — Vol. 62. — No. 17. — Pp. 6400–6426. DOI: https://doi.org/10.1080/00207543.2024.2305804. [In Eng.].

Rajeswari K.A., M., & Priyadharshini, C. (2024). A Comprehensive Review of Recent Artificial Intelligence Techniques and IOT Applications in Dairy Farms. 2024 10th International Conference on Advanced Computing and Communication Systems (ICACCS). — Pp. 242–247. DOI: https://doi.org/10.1109/icaccs60874.2024.10717012. [In Eng.].

Krupitzer, C., Noack, T., & Borsum, C. (2022). Digital Food Twins Combining Data Science and Food Science: System Model, Applications, and Challenges. Processes. — Vol. 10. — No. 9. — Pp. 1781. DOI: https://doi.org/10.3390/pr10091781. [In Eng.].

Keong, T. W., Husin, Z., Ismail, M. A. H., & Yasruddin, M. L. (2024). Stacked ensemble learning based on deep transfer learning models for food ingredient classification and food quality determination // Neural Computing and Applications. — Vol. 36. — No. 30. — Pp. 18705–18725. DOI: https://doi.org/10.1007/s00521-024-10233-y. [In Eng.].

Liakos, K. G., Athanasiadis, V., Bozinou, E., & Lalas, S. I. (2025). Machine Learning for Quality Control in the Food Industry: A Review // Foods. — Vol. 14. — No. 19. — Pp. 3424. DOI: https://doi.org/10.3390/foods14193424. [In Eng.].

Mhapsekar, R., Kilbane, D., Davy, S., Abraham, L., Fenelon, M., & O'Shea, N. (2025). A systematic review of the internet of things and artificial intelligence applications in milk quality monitoring and analysis // International Journal of Dairy Technology. — Vol. 78. — No. 3. Portico. DOI: https://doi.org/10.1111/1471-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

0307.70049. [In Eng.].

Neto, H. A., Tavares, W. L. F., Ribeiro, D. C. S. Z., Alves, R. C. O., Fonseca, L. M., & Campos, S. V. A. (2019). On the utilization of deep and ensemble learning to detect milk adulteration // BioData Mining. — Vol. 12. — No. 1. DOI: https://doi.org/10.1186/s13040-019-0200-5. [In Eng.].

Rasheed, A., San, O., & Kvamsdal, T. (2020). Digital Twin: Values, Challenges and Enablers From a Modeling Perspective // IEEE Access. — Vol. 8. — Pp. 21980–22012. DOI: https://doi.org/10.1109/access.2020.2970143. [In Eng.].

Rao, S., & Neethirajan, S. (2025). Computational Architectures for Precision Dairy Nutrition Digital Twins // A Technical Review and Implementation Framework. DOI: https://doi.org/10.20944/preprints202506.2401.v1. [In Eng.].

Sionek, B., Szydłowska, A., Küçükgöz, K., & Kołożyn-Krajewska, D. (2023). Traditional and New Microorganisms in Lactic Acid Fermentation of Food // Fermentation. — Vol. 9. — No. 12. — Pp. 1019. DOI: https://doi.org/10.3390/fermentation9121019. [In Eng.].

Su, S., Zhong, R. Y., & Jiang, Y. (2025). Digital twin and its applications in the construction industry: A state-of-art systematic review // Digital Twin. — Vol. 2. — No. 1. DOI: https://doi.org/10.12688/digitaltwin.17664.3. [In Eng.].

Verboven, P., Defraeye, T., Datta, A. K., & Nicolai, B. (2020). Digital twins of food process operations: the next step for food process models? Current Opinion in Food Science. — Vol. 35. — Pp. 79–87. DOI: https://doi.org/10.1016/j.cofs.2020.03.002. [In Eng.].

Zhao, B. (2025). Machine Learning-based Milk Quality Prediction // Exploring Science Academic Conference Series. — Vol. 9. — Pp. 1–14. DOI: https://doi.org/10.70267/ic-aimees.202501. [In Eng.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 279–305
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.015

STUDY OF A DYNAMIC MIXING DEVICE IN A BIOREACTOR USING IOT TECHNOLOGIES

K.Z. Igibayev[1,2], A.Kh. Kozbakova[1,3]*, O. Auelbekov[1,4] , T. Merembayev[1], K. Yessentayev[1,4]

[1]Institute of Information and Computational Technologies (Committee Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan), Kazakhstan;
[2]International University Transport and Humanities, Almaty, Kazakhstan;
[3]Almaty Technological University, Almaty, Kazakhstan;
[4]Kazakh National Women's Teacher Training University, Almaty, Kazakhstan.
E-mail: ajnurkozbakova@gmail.com

Kuanyshbek Z. Igibayev — Master, Institute of Information and Computational Technologies Committee Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan, senior lecturer of the Department "Automation and electric power engineering in transport"International University Transport and Humanities, Almaty, Kazakhstan
https://orcid.org/0000-0001-6495-5066;
Ainur Kh. Kozbakova — PhD, Professor, senior research fellow Institute of Information and Computational Technologies Committee Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Professor of the Department of «Information systems», Almaty Technological University, Almaty, Kazakhstan
E-mail: ajnurkozbakova@gmail.com, https://orcid.org/0000-0002-5213-4882;
Omirlan Auelbekov — Candidate of physical and Mathematical Sciences Associate Professor of the Department of «Information Technologies and Library Affairs», Kazakh National Women's Teacher Training University, senior research fellow Institute of Information and Computational Technologies Committee Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Almaty, Kazakhstan
https://orcid.org/0000-0002-2903-9086;
Timur Merembayev — PhD, senior research fellow of the Institute of Information and Computational Technologies Committee Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan
https://orcid.org/0000-0001-8185-235X;
Kairat Yessentayev — PhD student, of the Department of «Information Technologies and Library Affairs», Kazakh National Women's Teacher Training University, software

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

engineer of the Institute of Information and Computational Technologies Committee Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Almaty, Kazakhstan
https://orcid.org/ 0000-0002-8452-5378.

© K.Z. Igibayev, A.Kh. Kozbakova, O. Auelbekov, T. Merembayev, K. Yessentayev

Abstract. This article presents a comprehensive study of the influence of a dynamic mixing device on the temperature regime of biomass in an anaerobic bioreactor integrated with Internet of Things (IoT)-based monitoring and control technologies. The relevance of the study is due to the need to improve the efficiency of anaerobic digestion by ensuring temperature uniformity, which is one of the key factors for stable methanogenesis and biogas production. An intelligent control system based on the ESP32 microcontroller is proposed, providing multi-point temperature measurements along the reactor height, real-time data transmission via Wi-Fi/MQTT protocols, and adaptive control of mixing parameters. The integration of the IoT system provides continuous monitoring of process parameters, analysis of temperature distribution, and automatic adjustment of the operating mode depending on the state of the biomass. The research results demonstrate that the proposed approach improves heat and mass transfer, increases the activity of methanogenic microorganisms, and enhances biogas yield by 15–25% while simultaneously reducing energy consumption. This made it possible to evaluate the influence of the dynamic mixing device on flow distribution, temperature fields, biomass circulation, and the intensity of biogas release in different zones of the bioreactor.

Keywords: bioreactor, biomass mixture, biogas, dynamic mixing, temperature regime, IoT, ESP32, anaerobic digestion, intelligent control

For citation: Igibayev K.Z., Kozbakova A.Kh., Auelbekov O., Merembayev T., Yessentayev K. (2026). Study of a dynamic mixing device in a bioreactor using iot technologies // International journal of information and communication technologies. Vol. 7. No. 27. Pp. 279–305. https://doi.org/10.54309/IJICT.2026.27.3.015. (In Eng.).

Conflict of interest: The authors declare that there is no conflict of interest.

Acknowledgments: *The work was supported by a grant and funding from the Ministry of Science and Higher Education of the Republic of Kazakhstan within the framework of the Project №AP23490744, Institute Information and Computational Technologies CS MSHE RK.*

БИОРЕАКТОРДА IoT ТЕХНОЛОГИЯЛАРЫН ПАЙДАЛАНА ОТЫРЫП ДИНАМИКАЛЫҚ АРАЛАСТЫРУ ҚҰРЫЛҒЫСЫН ЗЕРТТЕУ

Қ.З. Игибаев[1,2], А.Х. Козбакова[1,3]*, О. Әуелбеков[1,4], Т. Мерембаев[1], Қ. Есентаев[1,4]

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

[1]Ақпараттық және есептеу технологиялары институты (Қазақстан Республикасы Ғылым және жоғары білім министрлігі Ғылым комитеті), Қазақстан;
[2]Халықаралық көлік және гуманитарлық университет, Алматы, Қазақстан;
[3]Алматы технологиялық университеті, Алматы, Қазақстан;
[4]Қазақ ұлттық қыздар педагогикалық университет, Алматы, Қазақстан.
E-mail: ajnurkozbakova@gmail.com

Игибаев Қуанышбек З. — магистр, ҚР Ғылым және жоғары білім министрлігі Ғылым комитетінің Ақпараттық және есептеу технологиялары институты, Халықаралық көлік және гуманитарлық ғылымдар университетінің аға оқытушысы, Алматы, Қазақстан
https://orcid.org/0000-0001-6495-5066;
Козбакова Айнұр Х. — PhD, профессор, Республикалық Ғылым және жоғары білім министрлігі Ғылым комитетінің Ақпараттық және есептеу технологиялары институтының аға ғылыми қызметкері, «Ақпараттық жүйелер» кафедрасының профессоры, Алматы технологиялық университеті, Алматы, Қазақстан
E-mail: ajnurkozbakova@gmail.com, https://orcid.org/0000-0002-5213-4882;
Әуелбеков Омирлан — физика-математика ғылымдарының кандидаты, «Ақпараттық технологиялар және кітапхана ісі» кафедрасының доценті, Қазақ ұлттық әйелдер мұғалімдер даярлайтын университеті, ҚР Ғылым және жоғары білім министрлігі Ғылым комитетінің Ақпараттық және есептеу технологиялары институтының аға ғылыми қызметкері, Алматы, Қазақстан
https://orcid.org/0000-0002-2903-9086;
Мерембеев Тимур — PhD, ҚР Ғылым және жоғары білім министрлігі Ғылым комитетінің Ақпараттық және есептеу технологиялары институтының аға ғылыми қызметкері, Қазақстан
https://orcid.org/0000-0001-8185-235X;
Есентаев Қайрат — PhD студенті, бағдарламалық қамтамасыз ету инженері, ҚР Ғылым және жоғары білім министрлігі Ғылым комитетінің Ақпараттық және есептеу технологиялары институты, Алматы, Қазақстан
https://orcid.org/ 0000-0002-8452-5378.

© Қ.З. Игибаев, А.Х. Козбакова, О. Әуелбеков, Т. Мерембаев, Қ. Есентаев

Аннотация. Бұл мақала заттар интернетіне (IoT) негізделген мониторинг және басқару технологияларымен біріктірілген анаэробты биореактордағы биомассаның температуралық режиміне динамикалық араластыру құрылғысының әсерін жан-жақты зерттеуді ұсынады. Зерттеудің өзектілігі анаэробты дигестия тиімділігін арттыру қажеттілігімен байланысты, себебі температуралық біркелкілікті қамтамасыз ету тұрақты метан түзілуі мен биогаз өндірудің негізгі факторларының бірі болып табылады. ESP32 микроконтроллеріне негізделген интеллектуалды басқару жүйесі ұсынылған, ол реактор биіктігі бойымен көп нүктелі температура өлшеулерін, Wi-Fi/MQTT протоколдары арқылы нақты

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

уақыттағы деректерді жіберуді және араластыру параметрлерін бейімді басқаруды қамтамасыз етеді. IoT жүйесін интеграциялау процесс параметрлерін үздіксіз бақылауды, температура таралуын талдауды және биомасса күйіне байланысты жұмыс режимін автоматты түрде реттеуді жүзеге асырады. Зерттеу нәтижелері ұсынылған тәсілдің жылу мен масса алмасуын жақсартып, метан түзетін микроорганизмдердің белсенділігін арттырып, биогаз өнімділігін 15–25%-ға арттыратынын және бір уақытта энергия тұтынуын азайтатынын көрсетті. Бұл биореактордың әртүрлі аймақтарындағы ағынның таралуы, температуралық өрістер, биомасса айналымы және биогаз бөліну қарқындылығына динамикалық араластырғыш құрылғының әсерін бағалауға мүмкіндік берді.

Түйінді сөздер: биореактор, биомасса қоспасы, биогаз, динамикалық араластыру, температураны бақылау, IoT, ESP32, анаэробты қорыту, интеллектуалды басқару

Дәйексөздер үшін: Игибаев К.З., Козбакова А.Х., Әуелбеков О., Мерембаев Т., Есентаев К. (2026). Биореакторда iot технологияларын пайдалана отырып динамикалық араластыру құрылғысын зерттеу // Халықаралық ақпараттық және коммуникациялық технологиялар журналы. Т. 7. No 27. Б. 279–305. https://doi.org/10.54309/IJICT.2026.27.3.015. (Ағыл. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

Алғыс: *Бұл жұмыс Қазақстан Республикасы Ғылым және жоғары білім министрлігінің Ғылым комитетіне қарасты Ақпараттық және есептеуіш технологиялар институтының № AP23490744 жобасы аясында бөлінген грант пен қаржыландырудың қолдауымен орындалды.*

ИССЛЕДОВАНИЕ УСТРОЙСТВА ДИНАМИЧЕСКОГО ПЕРЕМЕШИВАНИЯ В БИОРЕАКТОРЕ С ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИЙ IoT

К.З. Игибаев[1,2], А.Х. Козбакова[1,3]*, О. Ауелбеков[1,4], Т. Мерембаев[1], К. Есентаев[1,4]

[1]Институт информационных и вычислительных технологий (Комитет науки Министерства науки и высшего образования Республики Казахстан), Казахстан;
[2]Международный университет транспорта и гуманитарных наук, Алматы, Казахстан;
[3]Алматинский технологический университет, Алматы, Казахстан;
[4]Казахский национальный женский педагогический университет, Алматы, Казахстан.
E-mail: ajnurkozbakova@gmail.com

Игибаев Куанышбек З. — магистр, Институт информационных и вычислительных технологий Комитета по науке Министерства науки и высшего образования Республики Казахстан, старший преподаватель кафедры Международного университета транспорта и гуманитарных наук, Алматы, Казахстан

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

https://orcid.org/0000-0001-6495-5066;
Козбакова Айнур Х. — PhD, профессор, старший научный сотрудник Института информационных и вычислительных технологий Комитета по науке Министерства науки и высшего образования Республики Казахстан, профессор кафедры «Информационные системы» Алматинского технологического университета, Алматы, Казахстан
E-mail: ajnurkozbakova@gmail.com, https://orcid.org/0000-0002-5213-4882;
Ауелбеков Омирлан — кандидат физико-математических наук, доцент кафедры «Информационные технологии и библиотечное дело» Казахского национального женского педагогического университета, старший научный сотрудник Института информационных и вычислительных технологий Комитета по науке Министерства науки и высшего образования Республики Казахстан, Алматы, Казахстан
https://orcid.org/0000-0002-2903-9086;
Мерембаев Тимур — PhD, старший научный сотрудник Института информационных и вычислительных технологий Комитета по науке Министерства науки и высшего образования Республики Казахстан
https://orcid.org/0000-0001-8185-55X;
Ессентаев Кайрат — PhD, студент кафедры «Информационные технологии и библиотечное дело» Казахского национального женского педагогического университета, инженер-программист Института информационных и вычислительных технологий Комитета науки Министерства науки и высшего образования Республики Казахстан, Алматы, Казахстан
https://orcid.org/ 0000-0002-8452-5378.

© К.З. Игибаев, А.Х. Козбакова, О. Ауэльбеков, Т. Мерембаев, К. Ессентаев

Аннотация. В данной статье представлено комплексное исследование влияния устройства динамического перемешивания на температурный режим биомассы в анаэробном биореакторе, интегрированном с технологиями мониторинга и управления на базе «Интернета вещей» (IoT). Актуальность исследования обусловлена необходимостью повышения эффективности анаэробного сбраживания за счет обеспечения равномерности температурного режима, что является одним из ключевых факторов стабильного метаногенеза и производства биогаза. Предложена интеллектуальная система управления на базе микроконтроллера ESP32, обеспечивающая многоточечные измерения температуры по высоте реактора, передачу данных в реальном времени по протоколам Wi-Fi/MQTT и адаптивное управление параметрами перемешивания. Интеграция системы IoT обеспечивает непрерывный мониторинг технологических параметров, анализ распределения температуры и автоматическую настройку режима работы в зависимости от состояния биомассы. Результаты исследования показывают, что предложенный подход улучшает тепло- и массообмен, повышает активность метаногенных микроорганизмов и увеличивает выход биогаза на 15–25 % при одновременном снижении энергопотребления. Это позволило оценить влияние

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

устройства динамического перемешивания на распределение потока, температурные поля, циркуляцию биомассы и интенсивность выделения биогаза в различных зонах биореактора.

Ключевые слова: биореактор, смесь биомассы, биогаз, динамическое перемешивание, температурный режим, IoT, ESP32, анаэробное сбраживание, интеллектуальное управление

Для цитирования: Игибаев К.З., Козбакова А.Х., Ауэльбеков О., Мерембаев О., Ессентаев К. (2026). Исследование устройства динамического перемешивания в биореакторе с использованием технологий IoT // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 27. Стр. 279–305. https://doi.org/10.54309/IJICT.2026.27.3.015. (На англ.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Благодарность: *Работа выполнена при грантовой поддержке и финансировании Министерства науки и высшего образования Республики Казахстан в рамках проекта № AP23490744, реализуемого Институтом информационных и вычислительных технологий Комитета науки Министерства науки и высшего образования Республики Казахстан.*

Introduction.

Modern studies demonstrate that the efficiency of anaerobic digestion is determined by the stability of the temperature regime, the intensity of mixing, and the level of process automation (Khan et al., 2017; Appels et al., 2008). Temperature is one of the key factors influencing the activity of methanogenic microorganisms and biogas yield, while deviations greater than 2–3°C may lead to a decrease in process efficiency (Khan et al., 2017). According to research findings, temperature non-uniformity inside the bioreactor results in the formation of zones with different biological activity, which worsens the kinetics of methanogenesis and reduces the overall gas yield (Ward et al., 2008; Chen et al., 2008). Studies (Lindmark et al., 2014; Karim et al., 2005) show that mixing plays an important role in ensuring uniform distribution of temperature and substrate; however, conventional mixers often fail to provide sufficient adaptability when the viscosity of the biomass mixture changes. In recent years, dynamic and intelligent mixing systems capable of adapting their operating modes depending on the state of the medium have been actively developed (Wu, 2010). Such systems improve heat and mass transfer, eliminate stagnant zones, and enhance the efficiency of biochemical processes (Stroot et al., 2001). At the same time, Internet of Things (IoT) technologies are being actively introduced to enable real-time monitoring of bioreactor parameters (Ray, 2016; Al-Fuqaha et al., 2015). The use of wireless technologies such as Wi-Fi, MQTT, and cloud platforms ensures continuous data collection and analysis, as well as remote process control (Banks et al., 2014; Thangavel et al., 2014). Studies (Calise, 2020; Gil-Carrera et al., 2019) demonstrate that the integration of IoT into biogas plants reduces the risk of emergency situations, improves the stability of the technological regime, and optimizes energy consumption. In addition, the application

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

of intelligent control algorithms and machine learning methods makes it possible to predict parameter deviations and prevent failures (Zhang et al., 2019; Li et al., 2020). From the perspective of practical efficiency, studies show that the combined application of mixing and digital monitoring can increase biogas yield by 10–30% (Mertins et al., 2021; Céileachair et al., 2023). This confirms the necessity of developing integrated systems that combine mechanical and digital technologies.

The literature analysis shows that:

- temperature uniformity is a critically important factor (Khan et al., 2017; Ward et al., 2008);
- dynamic mixing increases the efficiency of heat transfer (Wu, 2010; Stroot et al., 2001);
- IoT provides a qualitatively new level of control (Ray, 2016; Banks et al., 2014);
- integration of technologies leads to improved efficiency of biogas plants (Mertins et al., 2021).

The solution proposed in this article corresponds to modern scientific trends and advances them through the integration of dynamic mixing and an intelligent IoT-based control system. The aim of this study is to investigate the temperature regime of the biomass mixture in an anaerobic bioreactor using a dynamic mixing device and IoT-based monitoring and control technologies, as well as to evaluate the influence of this approach on the efficiency of the digestion process and biogas yield.

The rapid increase in global energy demand, coupled with growing environmental concerns and the depletion of fossil fuel reserves, has heightened interest in renewable and sustainable energy technologies. Among these, biogas technologies hold a particularly important role, as they address two critical challenges simultaneously: the utilization of organic waste and the production of renewable energy carriers (Angelidaki et al., 2011; Demirbas, 2009; Weiland, 2010). Anaerobic digestion is widely employed across agriculture, wastewater treatment plants, food-processing industries, and municipal waste management systems due to its ability to convert organic substrates into methane-rich biogas. This biogas can then be used for thermal energy, electricity generation, or upgraded to biomethane for injection into gas grids and transportation applications (Khan et al., 2017; Weiland, 2010). The efficiency of anaerobic digestion depends heavily on maintaining stable physicochemical and biological conditions within the bioreactor. Key factors influencing performance include temperature stability, substrate homogeneity, pH balance, mixing intensity, and the concentration of active microorganisms (Khan et al., 2017; Ward et al., 2008; Chen et al., 2008). Temperature, in particular, is a critical parameter affecting the metabolic activity of methanogenic microorganisms and the kinetics of biochemical reactions during anaerobic digestion (Chen et al., 2008). Even minor temperature fluctuations can significantly impact methane production efficiency, microbial stability, and overall biogas plant performance (Ward et al., 2008). Studies have shown that temperature variations exceeding 2–3 °C may inhibit methanogenesis, slow substrate degradation, and reduce gas yield (Khan et al., 2017; Chen et al., 2008). A major engineering challenge

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

in conventional anaerobic bioreactors is the formation of temperature stratification and stagnant zones within the biomass mixture (Ward et al., 2008; Lindmark et al., 2014). In many practical systems, inadequate mixing leads to uneven heat distribution throughout the reactor volume, resulting in localized cold or overheated areas. This unevenness disrupts biochemical processes and adversely affects microbial activity (Lindmark et al., 2014; Karim et al., 2005). Additionally, poor circulation promotes substrate sedimentation, accumulation of dense organic layers, impaired heat and mass transfer, and reduced contact between microorganisms and nutrients (Karim et al., 2005; Stroot et al., 2001). These issues ultimately lower biogas productivity and extend the processing time of organic waste. To address these challenges, various mechanical mixing systems have been developed to improve hydrodynamic conditions inside bioreactors. Conventional agitators, propeller systems, and recirculation pumps are commonly used to enhance substrate homogeneity and heat transfer (Lindmark et al., 2014; Wu, 2010). However, most traditional mixing systems operate at fixed settings and cannot adapt to the continuously changing properties of the biomass, such as viscosity, density, and organic loading rate (Lindmark et al., 2014; Hoffmann et al., 2008). Since the rheological characteristics of biomass can vary significantly throughout different stages of anaerobic digestion, fixed mixing systems often fail to maintain optimal circulation across the entire reactor volume. Recently, there has been growing interest in adaptive and intelligent mixing systems that can adjust their operating parameters based on the current state of the biomass mixture (Wu, 2010; Stroot et al., 2001). Dynamic mixing devices with adjustable blade angles and variable immersion depths offer significant advantages over conventional systems. These devices enhance circulation flows, eliminate stagnant zones, improve heat transfer, and promote uniform substrate distribution throughout the reactor (Wu, 2010). Furthermore, adaptive mixing enables more efficient reactor operation under varying substrate viscosities and loading conditions while reducing unnecessary energy consumption.

Materials and methods.

The relevance of this study is determined by the growing need to improve the energy efficiency of biogas plants and to expand the application of digital technologies in bioenergy systems (Angelidaki et al., 2011; Demirbas, 2009). Under conditions of increasing interest in alternative energy sources, the task of improving the stability and productivity of anaerobic bioreactors becomes particularly important (Appels et al., 2008). One of the main problems of existing bioreactors remains the insufficient control of the temperature field inside the reaction volume (Ward et al., 2008). Uneven temperature distribution within the biomass mixture creates zones with reduced biological activity, which worsens the conditions for methanogenesis and decreases the efficiency of organic waste processing (Chen et al., 2008). Traditional mixing methods in many cases do not provide the required intensity and depth of impact on the substrate, especially when its viscosity, density, and structural condition change (Lindmark et al., 2014; Karim et al., 2005). Additional relevance to the study is provided by the active development of Internet of Things technologies, which make it possible to transition from passive bioreactors to intelligent systems with capabilities

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

for continuous monitoring, analysis, and automatic regulation of process parameters (Ray, 2016; Banks et al., 2014; Thangavel et al., 2014). Therefore, the development and investigation of a system combining dynamic mixing and IoT-based monitoring is an important scientific and technical task aimed at improving the reliability, controllability, and productivity of biogas plants. The scientific novelty of this work lies in the proposed integrated approach to controlling the temperature regime of the biomass mixture, based on the combined use of a dynamic mixing device with adjustable blade angles and a vertical movement system, together with an IoT-based multipoint monitoring system. Unlike traditional solutions, the proposed approach implements not only temperature control at individual points, but also real-time analysis of the spatial distribution of the temperature field throughout the entire height of the bioreactor (Zanella et al., 2014). A key feature of the proposed approach is the use of an adaptive control algorithm in which the operating mode of the mixing device changes not only according to the blade inclination angle (0–90°), but also according to the height position of the blades with a discrete step of 20 cm. This makes it possible to create controlled воздействия on different layers of the biomass mixture, starting from the lower zones with increased density and ending with the near-surface layers characterized by active gas release. Unlike conventional systems, where mixing is performed at a fixed point, the proposed scheme provides layer-by-layer mixing and targeted elimination of temperature non-uniformities.

Description of the Dynamic Mixing Device

The proposed dynamic mixing device (Figure 1) is a mechanical system consisting of a rotating shaft with blades, an electric motor, a gearbox, a lifting and lowering mechanism, a threaded section, and a self-cleaning system.

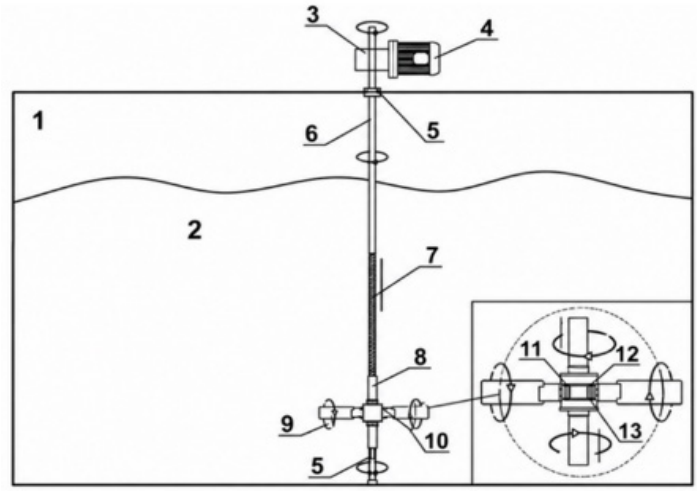


Fig. 1. Structural diagram of the dynamic mixing device installed in the bioreactor

1 - bioreactor body, 2 - biomass mixture, 3 - drive mounting unit, 4 - electric motor with gear reducer, 5 - support and guiding elements, 6 - vertical drive shaft, 7 - threaded section or movable adjustment mechanism, 8 - central mixing mechanism

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

unit, 9 -10 blades, 11-12-13 - additional detailed structural elements included in the mixing mechanism.

The design feature of the device is the ability to change the direction of rotation, regulate the immersion depth of the working elements, and adapt the mixing characteristics according to the state of the biomass mixture. The structural diagram (Figure 1) of the dynamic mixing device includes the bioreactor body (1), inside which the biomass mixture (2) is located and serves as the working medium of the anaerobic digestion process. In the upper part of the reactor, a drive mounting unit (3) is installed, on which an electric motor with a gear reducer (4) is mounted, providing rotational motion of the system. The torque from the motor is transmitted through the vertical drive shaft (6), which is stabilized by support and guiding elements (5), ensuring its alignment and reducing vibrations during operation. The shaft includes a threaded section or movable adjustment mechanism (7), allowing the position of the mixing device to be changed along the height of the reactor. In the lower part of the structure, the central mixing mechanism unit (8) is located, to which the side working elements - blades (9, 10) – are attached. These blades create circulation flows in the biomass mixture (2), ensuring intensive mixing and uniform distribution of temperature and substance concentration. Additionally, the mixing mechanism includes detailed structural elements (11, 12, 13), which are responsible for regulating the blade inclination angle and transmitting force from the shaft (6) to the working elements. This makes it possible to change the mixing mode depending on the properties of the biomass mixture (2), such as viscosity and density. Thus, the combined operation of all structural components - the body (1), the drive system (3, 4), the shaft (6), the regulating mechanism (7), and the blade system (9–13) – ensures effective mixing of the biomass mixture (2), improved heat transfer, and increased biogas yield. During operation of the dynamic mixing device, the shaft rotates under the action of the electric motor through the gearbox. The blades mounted on the shaft simultaneously:

- rotate around the axis,
- move along the threaded section of the shaft,
- change their position along the height of the reactor.

At the same time, an additional functional feature of the device is implemented, namely that during operation the blades rotate at a certain angle, which is automatically regulated depending on the biomass load, thereby ensuring effective mixing. This process is caused by the interaction between the gear mechanism and the movement mechanism, in which the resistance of the biomass mixture (its viscosity, density, and non-uniformity) affects the position of the blades and their angle of attack.

The device operates in a self-regulating mode:

- when the viscosity of the medium increases, the blade angle changes to increase the mixing effect;
- when the resistance of the medium decreases, the intensity of the effect is reduced.

The automatic blade inclination angle regulation mechanism is implemented based on monitoring the electric motor current consumption, which serves as a parameter

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

characterizing the biomass load. In the initial state, with minimum substrate viscosity and a blade inclination angle of 0°, a base motor current value of approximately 5 A is established. As the viscosity of the biomass mixture increases and the blades enter zones with increased resistance, the motor current gradually rises from 5 A to 10 A, depending on the physic mechanical properties of the medium.

With a further increase in load and when the blade inclination angle reaches its maximum value (up to 90°), a short-term peak current of 12–14 A is observed, indicating the maximum resistance of the medium. The control system is designed to automatically respond to changes in motor current and provides rapid regulation of:

- the blade inclination angle;
- the depth of blade immersion into the biomass mixture.

The blade inclination angle adjustment range is from 0° to 90° and is determined by the current load caused by the substrate viscosity. When the preset threshold current value corresponding to a critical load is reached, the electric motor, designed for reversible operation, automatically stops and switches to rotation in the opposite direction. As a result of this reverse mode, the blade inclination angle changes toward a decrease - from 90° to 0° - which leads to a reduction in medium resistance and restoration of the normal mixing regime.

The blade inclination angle changes automatically due to:

- changes in the load acting on the blades;
- corresponding changes in electric motor current;
- reversible drive control.

Figure 2 illustrates the design of a bioreactor engineered for optimal biomass processing. Key components include a dynamic mixing device (DMD) for vigorous agitation, a comprehensive temperature monitoring system (T1–T4), a level sensor, a pH sensor, and a water jacket for precise thermal control. The DMD's shaft (6), featuring a threaded section (7) and a blade mechanism (8) positioned at the reactor's base, drives active biomass mixing. This rotational action creates upward and downward currents, establishing a consistent circulation pattern throughout the entire reactor volume. Consequently, stagnant zones are eliminated, and the substrate is uniformly distributed. The dynamic mixing process also facilitates intensive heat transfer from the water jacket (hot water inlet – 18, return water outlet – 17), ensuring a uniform temperature profile across the reactor's height. Data from the temperature sensors (9–12) validates the significant reduction in the temperature gradient between the upper and lower sections. Furthermore, mass transfer is enhanced: biogas bubbles separate more efficiently from the liquid phase and rise rapidly towards the gas outlet (15). This leads to an increased biogas release rate and prevents gas buildup in the lower reactor layers. The integrated pH sensor (13) and level sensor (14) allow for continuous monitoring of crucial digestion parameters, thereby guaranteeing optimal conditions for methanogenesis. The biomass discharge outlet (17), located at the reactor's bottom, enables the removal of processed substrate and ensures sustained system performance.

The implementation of the dynamic mixing device leads to:

- uniform flow distribution throughout the reactor.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

- mitigation of temperature and concentration gradients.
- enhanced heat and mass transfer.
- improved rate and stability of biogas production.

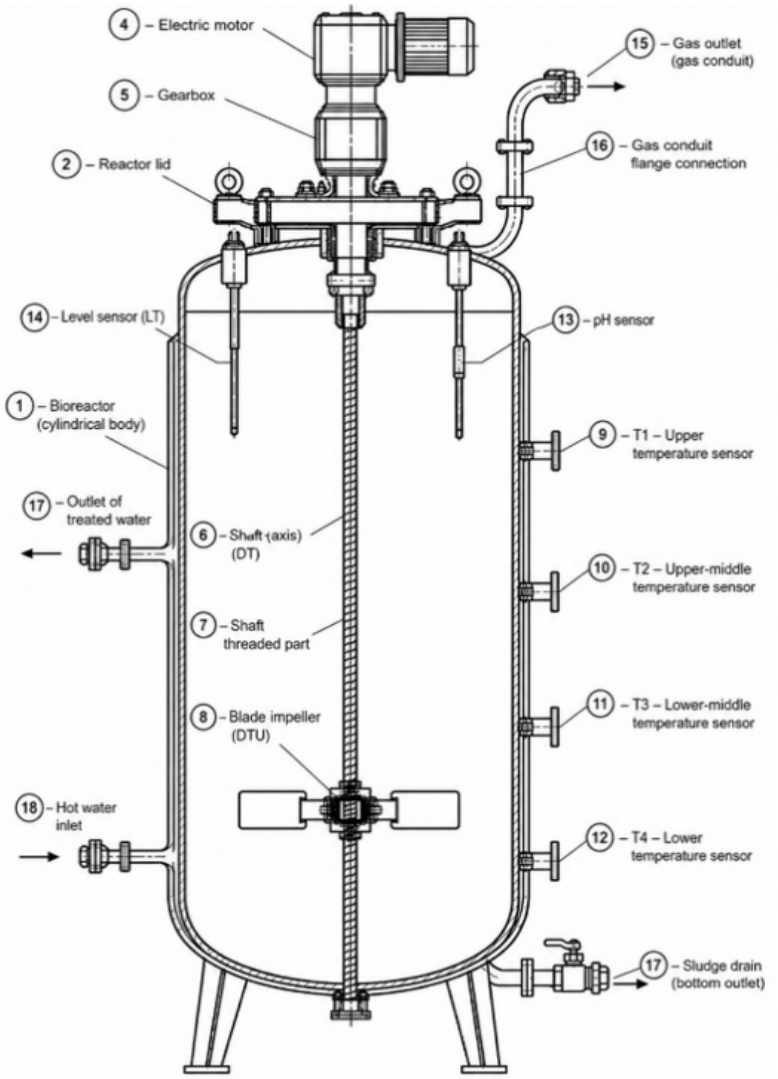


Fig. 2. Bioreactor with a Dynamic Mixing Device

This study addresses several interrelated objectives to achieve the stated goal. It begins with an analysis of current bioreactor mixing methodologies, specifically examining their effect on the thermal uniformity of the medium. Following this, a dynamic mixing device is engineered to promote more effective heat distribution and eradicate stagnant zones. An IoT-based monitoring system, built on the ESP32 platform, is then established for comprehensive multipoint temperature measurement and the real-time acquisition of operational data. Additionally, the research explores the temperature distribution within the biomass mixture under diverse reactor conditions, quantifies the impact of mixing on biogas output, and formulates an intelligent control

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

algorithm for adaptive activation and regulation of the mixing process based on the system's current state.

This principle ensures self-regulating adaptation of the device to changing properties of the biomass mixture without the need for external operator intervention. The use of the threaded section and the vertical movement mechanism makes it possible to change the working impact zone inside the reactor, which is especially important in the case of non-uniform distribution of temperature and substrate density along the reactor height. The self-cleaning system prevents the accumulation of deposits and ensures stable operation of the device under long-term operating conditions. Due to the formation of directed and turbulent flows, the proposed device contributes to the elimination of stagnant zones, equalization of temperature throughout the reactor volume, and intensification of heat transfer. As a result, more uniform mixing of the biomass mixture is achieved, the load on the drive system is reduced, conditions for biochemical processes are improved, and the overall efficiency of the bioreactor is increased. The novelty also lies in the development of an intelligent mixing control algorithm in which the motion parameters (blade angle α and height h) change depending on the magnitude of the temperature gradient and the local state of the biomass mixture (Zhang et al., 2019; Li et al., 2020). In mathematical form, this can be represented as follows:

$$\alpha = f(\nabla T),\ h = f(\nabla T, z) \tag{1}$$

where ∇T - is the temperature gradient, and z - is the coordinate along the height of the reactor.

This approach makes it possible not only to mix the substrate at predetermined time intervals, but also to purposefully influence zones with unfavorable thermal conditions, such as overheated or cooled regions. When significant temperature deviations are detected, the system automatically changes the position of the device and the blade angle, intensifying the mixing process specifically in the problematic areas.

Change in Biomass Mixing with Variation of Blade Angle from 0° to 90°

When the inclination angle of the blades of the dynamic mixing device changes from 0° to 90°, a gradual change occurs in the nature of the blade interaction with the biomass mixture. At small angles, the blades create low resistance to the flow; therefore, the mixing process occurs gently, and the electric motor current consumption remains close to its minimum value. As the inclination angle increases, the interaction area between the blade and the biomass mixture becomes larger, which leads to intensified circulation flows, increased shear impact, and more effective elimination of stagnant zones. At blade angles of 60–75°, the mixing process becomes most efficient, since active movement of the biomass is ensured in both the radial and vertical directions. When the angle reaches 90°, the resistance of the biomass mixture becomes maximal, and the motor current temporarily rises to its peak value. Therefore, this operating

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

mode is advisable only under conditions of high viscosity or when it is necessary to break up dense layers.

The blade inclination angle is denoted as:

$$\alpha = 0 \ldots 90°$$

where α - is the blade inclination angle relative to the horizontal position.

The effective interaction area of the blade with the biomass mixture can be expressed as:

$$A_{eff} = A \cdot sin\alpha \tag{2}$$

where: A_{eff} - is the effective interaction area of the blade, A- is the total blade area, α - is the blade inclination angle.

As the inclination angle increases, the hydrodynamic resistance also increases:

$$F = C_d \cdot \frac{\rho v^2}{2} \cdot A_{eff} \tag{3}$$

or:

$$F = C_d \cdot \frac{\rho v^2}{2} \cdot A_{eff} \cdot \sin(\alpha) \tag{4}$$

where: F - is the resistance force of the biomass mixture, C_d - is the drag coefficient, ρ - is the density of the biomass mixture, v - is the blade movement velocity, A - is the blade area.

The torque on the shaft can be determined as:

$$M = F \cdot r \tag{5}$$

where: M - is the torque, r - is the radius of the blade position.

Then, the dependence of the torque on the blade angle is given by:

$$M(\alpha) = C_d \cdot \frac{\rho v^2}{2} \cdot A_{eff} \cdot r \cdot \sin(\alpha) \tag{6}$$

The current consumption of an electric motor can be roughly expressed in terms of load torque:

$$I(\alpha) = I_0 + k \cdot \sin(\alpha) \tag{7}$$

where: I_0 - is the current of the idle or minimum mode, k - is the coefficient of coupling between the torque and the current of the motor.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

For a practical description, you can use a simplified dependency:

$$I(\alpha) = I_0 + (I_{max} - I_0) \cdot \sin(\alpha) \quad (8)$$

If you accept:

$$I_0 = 5\ A, I_{max} = 10\ A$$

then:

$$I(\alpha) = 5 + 5 \cdot \sin(\alpha)$$

Table 1. Changes in parameters when changing the angle of the blades.

Blade angle, deg	α	motor current, A	Mixing Pattern
0°	0	5.00	Minimal resistance, weak mixing
15°	0.259	6.30	Initial Flow Gain
30°	0.500	7.50	Moderate stirring
45°	0.707	8.54	Active Stirring
60°	0.866	9.33	Intensive circulation of the biomass mixture
75°	0.966	9.83	Very strong mixing, destruction of stagnant zones
90°	1.000	10.00	Maximum load, maximum resistance

When a short-term angle of 90° is reached in the high viscosity zone, the current can increase to a peak value:

$$I_{peak} = 12 \ldots 14\ A$$

Changing the operation of the stirring device during lifting in 20 cm increments

When the mixing device is lifted along the vertical axis with a step of 20 cm, the blades are sequentially affected by different layers of the biomass mixture. This mode allows to destroy dead zones not only in the lower part of the reactor, but also along the entire height of the working medium. As the blades rise, they move the dense layers upwards, improve the distribution of organic particles and increase the release of biogas. which are then combined into a common volume flow. This increases the uniformity of mixing, reduces the settling of solids and improves the contact of microorganisms with the culture medium. The height of the agitator position can be set as:

$$h_i = h_0 + i \cdot \Delta h \quad (9)$$

where: h_i - position of the blades at the i-th level, m; h_0 - initial height from the bottom of the reactor, m; Δh=0,20 м - lifting step; i=0,1,2,3,…,n - step number.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

If the working height of the biomass mixture is equal to (H), the number of positions shall be determined:

$$n = \frac{H - h_0}{\Delta h} \tag{10}$$

For example, if the height of the biomass mixture: H = 2 m, h_0 = 0,20 м
Then: $n = \frac{2-0,2}{0.2.} = 9$
That is, the device can occupy 10 positions: from 0.20 m to 2.00 m.

Relation of vertical movement of the device to the angle of inclination of the blades

When lifting the device, the angle of the blades can change depending on the resistance of the biomass mixture:

$$\alpha(h) = \alpha_{min} + (\alpha_{max} - \alpha_{min}) \cdot \frac{I(h) - I_0}{I_{max} - I_0} \tag{11}$$

where: *α(h)*-is the angle of the blades at a given height; $\alpha_{min} = 0°; \alpha_{min} = 90°$; $I(h)$ -motorcurrentatagivenaltitude. $I_0 = 5\ A; I_{max} = 10\ A$
If the biomass mixture is more viscous at a certain height, the motor current increases, and the system automatically increases the angle of inclination of the blades. If the resistance decreases, the angle of the blades is reduced so as not to create an unnecessary load on the engine.

Thus, changing the angle of inclination of the blades from 0° to 90° allows you to adjust the intensity of mixing of the biomass mixture depending on its viscosity and resistance. At small angles, the load on the motor is minimal, but the mixing efficiency is also low. With an increase in the angle to 45–75°, more pronounced circulation flows are formed, which provide an even distribution of organic mass and activation of gas emissions. The angle of 90° corresponds to the maximum resistance of the biomass mixture and is used as a short-term mode for compaction or increased viscosity of the substrate. The control of the angle of the blades by changing the motor current allows you to automatically adapt the operation of the mixing device to the real state of the biomass in the reactor. Thus, the lifting of the mixing device with a step of 20 cm ensures the layer-by-layer treatment of the biomass mixture along the entire height of the reactor. In the lower zone, the device destroys dense sludge, in the middle zone it forms stable circulation flows, and in the upper zone it contributes to the active release of biogas. This method of mixing is more efficient than the usual stationary arrangement of the blades, since the effect is not at one point, but throughout the entire volume of the reactor. Lift control, combined with changing the angle of the blades, adapts the operation of the device to the viscosity of the biomass and reduces the risk of stagnant zones. In addition, the paper shows that the combination of vertical movement of the device and regulation of the blade angle leads to the formation of

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

more intense circulation and turbulent flows throughout the reactor volume (Wu, 2010; Stroot et al., 2001). This ensures a reduction in temperature gradients, improved heat exchange and an increase in the homogeneity of the biomass mixture, which in turn contributes to an increase in the efficiency of anaerobic digestion and the stability of the biogas formation process.

IoT Monitoring and Control System

The developed IoT monitoring and control system is built on the basis of the ESP32 microcontroller, which acts as a central node for collecting, processing and transmitting data (Al-Fuqaha et al., 2015; Li et al., 2020). The system includes temperature sensors located on several levels inside the bioreactor, as well as additional humidity, gas leakage and pressure sensors that allow monitoring both the internal state of the environment and the parameters of safe operation of the facility. wireless communication channels using Wi-Fi and the MQTT protocol (Banks et al., 2014; Thangavel et al., 2014), which provides the possibility of remote monitoring and integration with external information systems.

To visualize the parameters, a web panel is used, which displays current temperature values, heat field distribution, actuator status, and alarm messages. Functionally, the system provides real-time data collection, analysis of temperature deviations, automatic control of the mixing device and transmission of notifications in case of emergency situations (Ray, 2016; Calise, 2020). The IoT component not only increases the information content of process monitoring, but also creates the basis for intelligent control of the bioreactor. Figure 3 shows an intelligent control system for the process of mixing the biomass mixture in the bioreactor, equipped with multipoint temperature monitoring. Inside the reactor vessel, a dynamic mixing device with a vertical shaft and adjustable blades is installed, which ensures the formation of circulation flows throughout the biomass mixture volume. In the upper part of the reactor, there is an active release of biogas, which indicates the intensive course of anaerobic processes fermentation. Four temperature sensors are installed to monitor the temperature regime by the height of the reactor: the upper T1 sensor, the middle upper T2 sensor, the middle lower T3 sensor and the lower T4 sensor. These sensors are placed at different levels (H1–H4), which allows obtaining information about the spatial distribution of temperature inside the bioreactor. Signals from the sensors are transmitted to the data acquisition and processing system based on the ESP32 microcontroller, where their primary processing is carried out. The data is then transmitted wirelessly (WiFi/MQTT) to the cloud system and displayed in the web interface as a real-time temperature field. Based on the analysis of the data obtained, the presence of temperature gradients and inhomogeneities in the biomass mixture is determined. If deviations are detected, the system automatically adjusts the agitator by changing the blade angle and the height of the mechanism. The presented system provides adaptive control of the mixing process, which contributes to the equalization of the temperature field, increasing the homogeneity of the biomass mixture and increasing the efficiency of biogas production.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

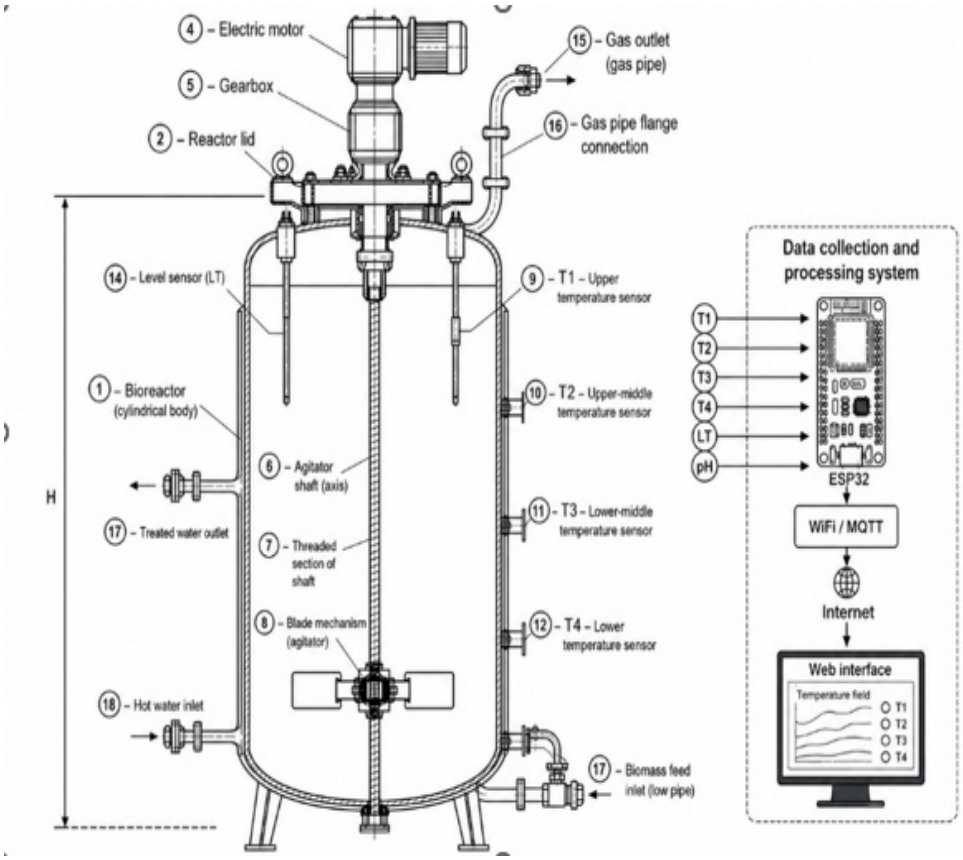


Fig. 3. IoT System Diagram for Bioreactor

Figure 4 shows a structural diagram of the algorithm for controlling the mixing of a biomass mixture in a bioreactor, implemented on the basis of IoT technologies and temperature field analysis. The algorithm consists of sequential steps that provide data collection, processing, analysis, and adaptive control of mixing parameters. In the first step, data is collected from the T1-T4 temperature sensors located at the height of the reactor, after which the information is transmitted via the ESP32 microcontroller via WiFi/MQTT to the cloud system and displayed in the web interface in real time. At the second stage, data processing is carried out, including the formation of a temperature field and the calculation of the temperature gradient as the difference between the temperature values along the height of the reactor. And the third stage is the analysis of the state of the system, which determines the presence of a temperature imbalance based on a comparison of the current gradient with the permissible value. If the threshold value is exceeded, the intensive mixing mode is activated, otherwise the current operating mode is maintained blade inclination in the range from 0° to 90° depending on the value of the local temperature gradient. Next, control actions are performed, including moving the device in height and changing the angle of the blades. After the control actions are performed, the result is monitored for a specified time, then the data is re-collected and the effectiveness of the regulation is checked. If a uniform temperature

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

distribution is achieved, the system switches to the minimum mode, and if deviations persist, the control cycle is repeated. As a result of the algorithm, temperature gradients are reduced, the temperature field is equalized by the height of the reactor, the anaerobic digestion process is stabilized, the intensity of biogas formation is increased and energy consumption is reduced. Event logging is also provided, recording system parameters such as time, agitator height, blade angle, and pre- and post-adjustment temperatures. The experimental study was carried out on a bioreactor in several modes of operation, which made it possible to compare the effectiveness of different methods of influencing the biomass mixture. Three main modes were considered: no mixing, conventional mechanical mixing and dynamic mixing using the proposed device. The study was carried out in two characteristic temperature ranges: mesophilic mode at a temperature of about 35 °C and thermophilic mode at a temperature of about 55 °C. Comparison of the data obtained made it possible to determine the degree of temperature homogeneity of the medium, to establish the influence of the type of mixing on the dynamics of the process and to assess the effectiveness of the proposed system in conditions close to real operation.

Results and discussion.

As a result of the application of the proposed algorithm, significant improvements were obtained in the temperature distribution and efficiency of the digestion process. The analysis was carried out using three-dimensional graphs and a table reflecting the dependence of temperature on reactor altitude and time. Analysis of the results presented in Table 2 shows that the change in the height of the dynamic mixing device (DPC) significantly affects the nature of mixing and the distribution of biomass mixture flows inside the bioreactor. At the initial position of the blades at a height of 0.20 m, the impact is concentrated in the lower zone of the reactor, where the dense sedimentary layer is destroyed and the accumulation of heavy organic deposits is prevented. This helps to reduce the local compaction of the substrate and improve the access of microorganisms to organic substances. When the blades rise to 0.40 m, the impact zone shifts to the lower-middle area of the reactor. In this mode, there is an active uplift of heavy particles, which reduces the probability of the formation of stagnant zones and improves the circulation of biomass. At a height of 0.60 m, the vertical flow increases, which contributes to a more intensive heat transfer and distribution of organic components throughout the reactor volume. where the main mixing of the biomass mixture is carried out.

In this area, the most stable flow distribution is achieved, temperature heterogeneity is reduced (Ward et al., 2008; Lindmark et al., 2014; Wu, 2010) and the uniformity of anaerobic digestion is increased. With a further rise of up to 1.00 m, the device acts on the central zone of the reactor, ensuring the equalization of the concentration of organic substances and increasing the efficiency of biochemical processes. At a height of 1.20 m, the impact zone passes into the central-upper part of the reactor.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

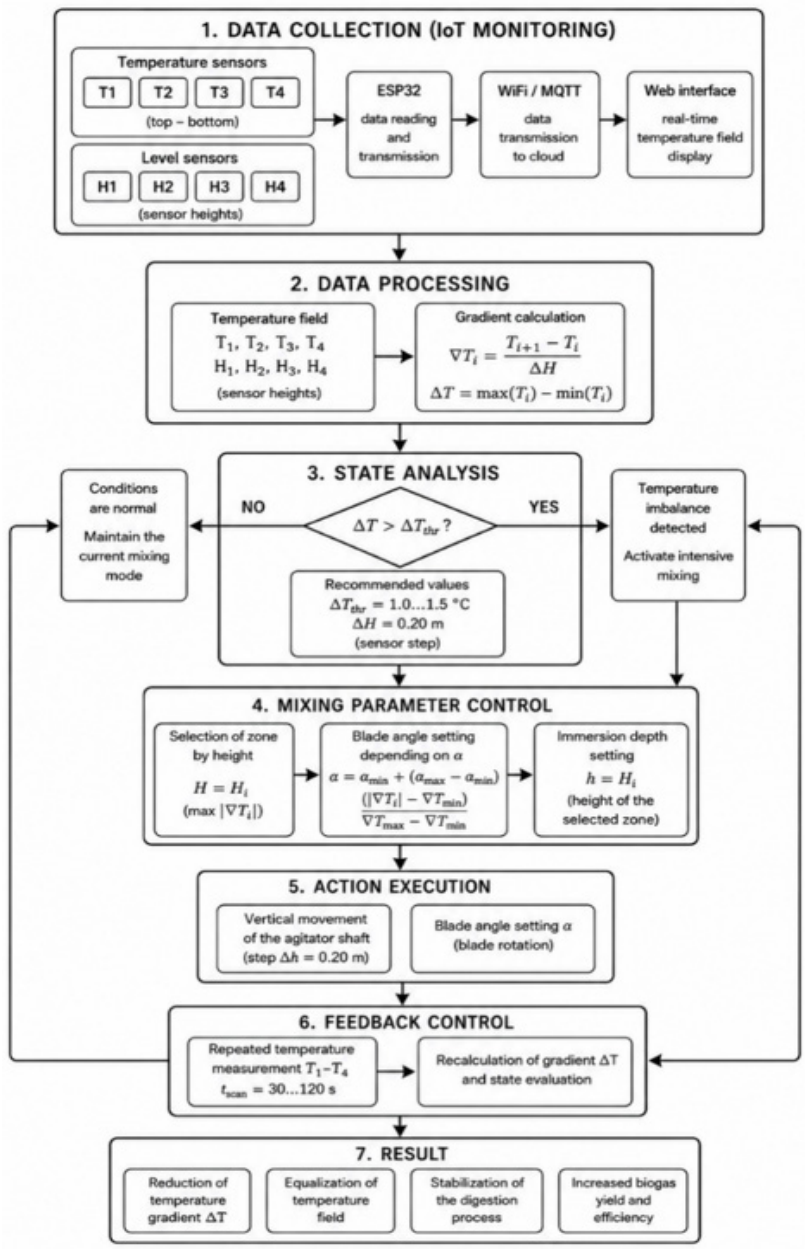


Fig. 4. Structural diagram of the algorithm for controlling the mixing of a biomass mixture in a bioreactor

Here there is an increase in the circulation of the biomass mixture and an acceleration of the movement of gas-liquid flows. At the 1.40 m position, the blades work in the upper zone, providing biomass to the area of active gas release, which improves the process of separating biogas bubbles from the substrate. When the DMD rises to 1.60 m, the release of gas bubbles is intensified (Wu, 2010; Stroot et al., 2001), which is associated with an increase in turbulence in the upper layers of the reactor. This helps to reduce gas accumulation within the biomass and reduce local

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

oversaturation zones. At a height of 1.80 m, the device acts on the near-surface area, ensuring the removal of accumulated gas and preventing the formation of stable gas pockets. The maximum position of the blades at a height of 2.00 m corresponds to the surface zone of the reactor. In this mode, the most intense gas emission and destruction of surface foam are observed. This improves biogas yield, stabilizes reactor operation and reduces the risk of clogging the top of the system. The results obtained confirm that the sequential change in the DMD height makes it possible to effectively control hydrodynamic flows, biomass distribution and the process of gas formation throughout the bioreactor.

Table 2. Change of the impact zone when lifting the DMD with a step of 20 cm

Step	Blade height, m	Height, cm	Area of Exposure	Mixing Pattern
0	0,20	20	Lower Zone	Destruction of sediment and dense layer
1	0,40	40	Lower - middle zone	Lifting of heavy particles upwards
2	0,60	60	Lower-middle zone	Vertical Flow Reinforcement
3	0,80	80	Middle Zone	Basic mixing of the biomass mixture
4	1,00	100	Central Zone	Leveling the concentration of organic matter
5	1,20	120	Central-upper zone	Increased circulation
6	1,40	140	Upper zone	Biomass supply to the gas release zone
7	1,60	160	Upper zone	Gas bubble release stimulation
8	1,80	180	Near-surface zone	Removal of accumulated gas
9	2,00	200	Surface Zone	Maximum Gassing and Foam Breakdown

The graph in Figure 5 shows the initial state of the system without intelligent management. The graph shows a pronounced temperature stratification: the temperature is higher in the upper part of the reactor, and much lower in the lower part. The graph surface has a slope and an inhomogeneous structure, which indicates weak mixing and poor heat transfer.

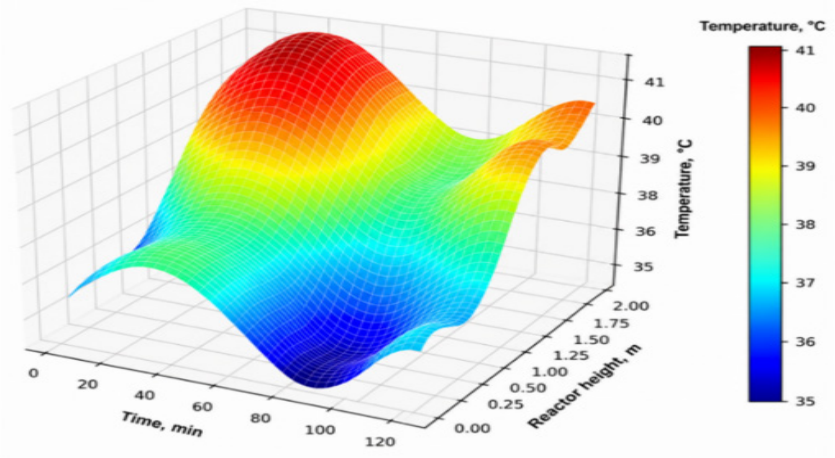


Fig. 5. Initial state of the system without intelligent control

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

The graph in Figure 6 shows the state after the algorithm has been applied. The temperature surface becomes more even, the differences between the layers are reduced, and the temperature distribution along the height of the reactor is evened out. This indicates effective mixing and the elimination of stagnant zones.

Figure 7 shows the dynamics of the temperature gradient over time. At the initial moment, there is a high temperature range, but as the algorithm works, the gradient gradually decreases and stabilizes. The surface of the graph becomes more uniform, which indicates the achievement of a stable regime. In addition, it has been established that a decrease in the temperature gradient leads to improved heat exchange and a more uniform distribution of microorganisms in the biomass mixture.

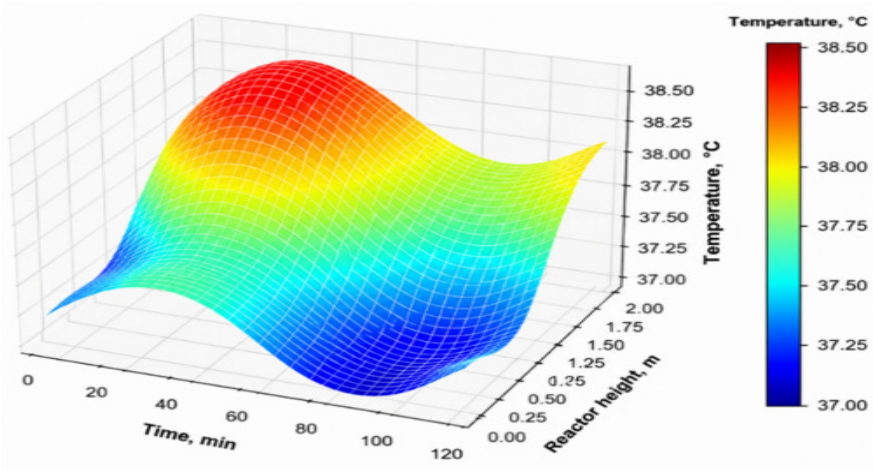


Fig. 6. Graph of the demonstration of the state after applying the algorithm

This contributes to an increase in the intensity of anaerobic digestion and an increase in biogas yield. The results of modeling and analysis confirm the effectiveness of the proposed algorithm. The use of adaptive mixing control not only stabilizes the temperature regime, but also improves the overall energy efficiency of the system.

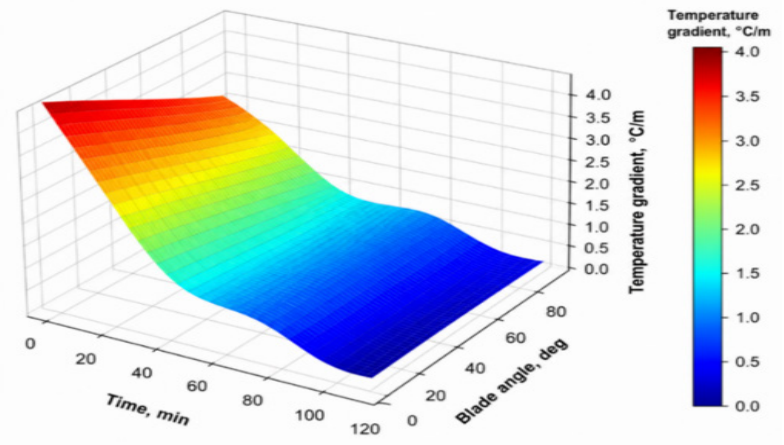
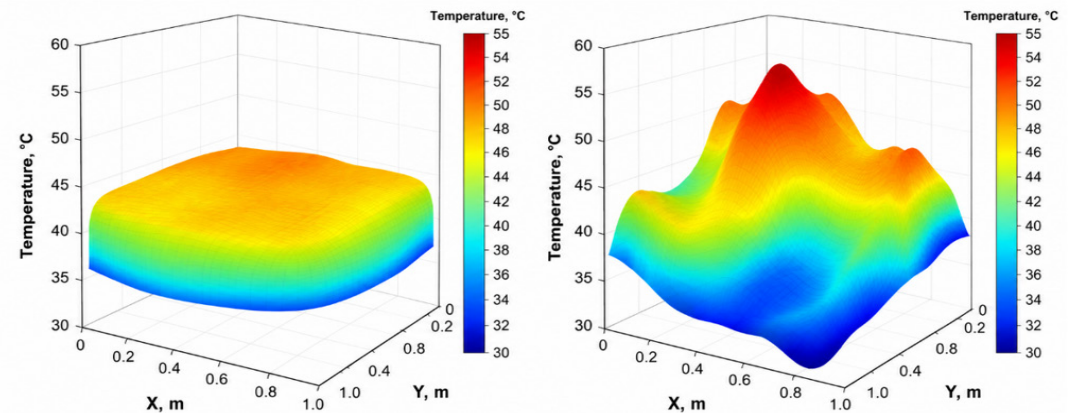


Fig. 7. Graph of reflection of the dynamics of temperature changes gradient in time.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Studies have shown that in the absence of mixing, pronounced temperature gradients are formed inside the bioreactor, and the difference between individual zones can reach 8–10 °C. Such unevenness adversely affects the stability of methanogenesis and reduces the efficiency of the use of the reaction volume, which is consistent with the data of other studies (Ward et al., 2008; Lindmark et al., 2014). When using a dynamic mixing device, the temperature field pattern changes significantly. Due to the active movement of the biomass mixture and the destruction of stagnant zones, a more uniform distribution of heat is ensured, and the temperature deviation along the volume of the reactor does not exceed 1–2 °C (Wu, 2010; Stroot et al., 2001). This indicates the high efficiency of the proposed design in ensuring thermal homogeneity and creating optimal conditions for anaerobic digestion (Chen et al., 2008; Karim et al., 2005).



a) with the use of DMD. b) with a conventional reaming device.

Fig. 8. Comparison of the temperature field in a bioreactor when using a DMD and a conventional mixing device

Figure 8 shows a comparative three-dimensional temperature distribution in a bioreactor under two mixing modes: with the use of a dynamic mixing device (DMD) and with a conventional mixing device. Figure 8(a) shows the temperature field when the DMD is operating at a speed of 30 rpm. The temperature varies in a narrow range (approximately 40-45 °C), without sudden local overheating or cooling zones. This indicates efficient mixing, in which there is intensive heat transfer and equalization of the temperature field. Figure 8 (b) shows the temperature distribution when using a conventional agitator. The surface has pronounced irregularities and peaks, which indicates the presence of significant temperature inhomogeneities. Temperatures vary over a wider range (30 to 55 °C), with local overheating zones (up to 55 °C) and low temperature zones (around 30 to 35 °C). This indicates insufficient mixing efficiency and the presence of stagnant areas.

A comparison of the two modes shows that the use of DMD leads to:

- reducing temperature gradients,
- elimination of local overheating,

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

- formation of a more homogeneous structure of the biomass mixture,
- increasing the efficiency of heat exchange.

The use of a dynamic agitator provides a more uniform temperature distribution over the volume of the bioreactor compared to traditional mixing systems, which contributes to increasing the stability and efficiency of the biogas formation process (Lindmark et al., 2014; Wu, 2010; Weiland, 2010).

In the course of the study, additional results were obtained using the MATLAB environment and the Simulink module, which made it possible to simulate the dynamics of the temperature field and mixing processes in the bioreactor. Simulations have shown that in the absence of mixing, pronounced temperature gradients of up to 8–10 °C are formed, while when using a dynamic mixing device, they are reduced to 1–2 °C throughout the reactor volume. The results of the simulation also confirmed that adaptive change in the angle of inclination of the blades in the range of 60-75° provides the most uniform temperature distribution and the lowest energy costs. In addition, 3D models of temperature and flow rate distribution were built at MATLAB, demonstrating the elimination of stagnant zones and the formation of stable circulation flows. The data obtained are in good agreement with the experimental results and confirm the effectiveness of the proposed method of dynamic mixing and intelligent control.

The results obtained confirm that the improvement of temperature uniformity has a direct impact on the kinetics of the digestion process. With the use of dynamic mixing, there is an acceleration of the decomposition of the organic substrate, an increase in the activity of microorganisms and a reduction in the overall processing time. This is due to the fact that a uniform temperature field contributes to maintaining a favorable environment for the life of methanogenic microflora and reduces the likelihood of zones in which biological activity is suppressed. In addition, more intensive heat and mass transfer contributes to better contact of microorganisms with organic matter, which increases the completeness of substrate processing. Thus, dynamic mixing has a positive effect not only on the physical state of the biomass mixture, but also on the biochemical efficiency of the entire process. The use of an IoT monitoring system made it possible to provide continuous monitoring of the process in real time and improve the accuracy of the bioreactor control. Thanks to multi-point temperature measurement, it is possible to detect deviations in heat distribution in a timely manner and react quickly to them by switching on agitation. This approach ensures the transition from periodic monitoring to continuous digital process support. An additional advantage of the system is the ability to automate decision-making. Based on the data coming from the sensors, the controller analyzes the state of the biomass mixture and activates the actuators only when it is really needed. This makes it possible to increase energy efficiency, reduce excessive equipment operation, and at the same time improve the stability of the digestion process (Ray, 2016; Al-Fuqaha et al., 2015). The combined use of a dynamic mixing device and IoT monitoring has shown that such a system provides a noticeable increase in the overall efficiency of the bioreactor. According to the results of the study, an increase in biogas yield within 15–25% was established, a

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

decrease in energy consumption due to a more rational mode of operation of actuators and an increase in the stability of the technological process. These results indicate that the proposed system is not limited to improving the local characteristics of the thermal regime, but has a complex positive effect on the functioning of the entire biogas plant. Increasing the stability and controllability of the process creates conditions for the practical implementation of the developed solution both at small farm plants and at industrial biogas plants (Mertins et al., 2021; Céileachair et al., 2023).

Conclusion.

The obtained results demonstrate that the integration of a dynamic agitator with an intelligent IoT monitoring system exerts a complex influence on the efficiency of anaerobic bioreactors. A key limitation in conventional systems is significant temperature stratification along the reactor's height, leading to gradients of 8–10 °C. This stratification impedes methanogenic microbial activity and fosters the development of stagnant zones. The introduction of dynamic mixing fundamentally reconfigures temperature distribution by generating directed circulation and turbulent flows. These phenomena ensure vigorous movement of the biomass mixture throughout the reactor, thereby reducing temperature deviations to an optimal range of 1–2 °C for methanogenesis. The adaptive control of the blade angle, adjustable from 0° to 90°, plays a pivotal role in process optimization. At lower angles, gentle mixing with reduced energy expenditure is achieved. Conversely, increasing the angle enhances hydrodynamic resistance and intensifies the process. The most advantageous operational mode is observed within the 60–75° range, striking an optimal balance between energy consumption and mixing intensity. A temporary adjustment to 90° is recommended for highly viscous substrates or when dense layers require disruption. Furthermore, the vertical displacement of the device, with a step of 0.20 m, imparts a layer-by-layer effect on the biomass mixture. This action dislodges precipitates in the lower zone, establishes stable circulation flows in the middle zone, and intensifies gas release in the upper zone. The ESP32-based IoT monitoring system facilitates multi-point temperature measurement, real-time data transmission, and the execution of adaptive control algorithms. This transition from periodic monitoring to continuous, intelligent process management allows agitation to be activated only when permissible temperature gradients are exceeded. Such a targeted approach minimizes energy consumption (Calise et al., 2020; Mertins et al., 2021) and enhances overall system efficiency. From a biochemical perspective, diminished temperature heterogeneity promotes increased activity of methanogenic microflora (Khan et al., 2017; Chen et al., 2008), accelerates organic substrate decomposition, and shortens digestion time. Enhanced heat and mass transfer ensures more effective contact between microorganisms and the nutrient medium, leading to a more complete processing of organic matter. Consequently, the synergistic effects of dynamic mixing and intelligent monitoring can lead to a 15–25% increase in biogas yields, coupled with reduced operational costs and mitigated risks associated with process parameter variability. The practical significance of this research lies in its potential for widespread adoption across biogas plants of varying scales, from farm-level operations to industrial complexes. The utilization

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

of readily available components, wireless communication technologies, and adaptive control algorithms ensures straightforward implementation, scalability, and versatility. This system is particularly valuable for facilities requiring remote monitoring, where minimizing personnel intervention and maximizing technological process reliability are paramount. The integration of the dynamic mixing device with the IoT control system represents a significant advancement in anaerobic bioreactor management, ensuring process stability, enhanced energy efficiency, and reduced operational risks».

REFERENCES

Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., Ayyash, M. (2015). Internet of Things: A survey // IEEE Communications Surveys & Tutorials. — Vol. 17. — No. 4. — Pp. 2347–2376. DOI: https://doi.org/10.1109/COMST.2015.2444095. [In Eng.].

Angelidaki, I., Ellegaard, L., Ahring, B.K. (2011). Applications of the anaerobic digestion process // Bioresource Technology. — Vol. 102. — No. 2. — Pp. 199–209. DOI: https://doi.org/10.1016/j.biortech.2011.09.104. [In Eng.].

Appels, L., Baeyens, J., Degrève, J., Dewil, R. (2008). Principles and potential of the anaerobic digestion of waste-activated sludgem // Progress in Energy and Combustion Science. — Vol. 34. — No. 6. — Pp. 755–781. DOI: https://doi.org/10.1016/j.pecs.2008.06.002. [In Eng.].

Banks, A., Gupta, R. (2014). MQTT Version 3.1.1. IETF RFC. — Pp. 6455. DOI: https://doi.org/10.17487/RFC5246. [In Eng.].

Calise, F., Dentice d'Accadia, M., Libertini, L. (2020). Integration of biomethane in energy systems // Energy Conversion and Management. — Vol. 205. — DOI: https://doi.org/10.1016/j.enconman.2020.113090. [In Eng.].

Chen, Y., Cheng, J.J., Creamer, K.S. (2008). Inhibition of anaerobic digestion process: A review // Bioresource Technology. — Vol. 99. — No. 10. — Pp. 4044–4064. DOI: https://doi.org/10.1016/j.biortech.2007.01.057. [In Eng.].

Demirbas, A. (2009). Biogas production technologies // Energy Conversion and Management. — Vol. 50. — No. 7. — Pp. 1786–1792. DOI: https://doi.org/10.1016/j.enconman.2009.03.011. [In Eng.].

Gil-Carrera, L., et al. (2019). Feasibility study of biomethane mobile solution // Applied Energy. — Vol. 239. — Pp. 471–481. DOI: https://doi.org/10.1016/j.apenergy.2019.01.141. [In Eng.].

Hoffmann, R.A., Garcia, M.L., Veskivar, M., Karim, K., Al-Dahhan, M.H., Angenent, L.T. (2008). Effect of mixing on biogas production // Bioresource Technology. — Vol. 99. — No. 13. — Pp. 5702–5710. DOI: https://doi.org/10.1016/j.biortech.2008.12.030. [In Eng.].

Khan, I.U., Othman, M.H.D., Hashim, H., et al. (2017). Biogas as a renewable energy fuel – A review of biogas upgrading, utilisation and storage // Energy Conversion and Management. — Vol. 150. — Pp. 277–294. DOI: https://doi.org/10.1016/j.enconman.2017.08.035. [In Eng.].

Karim, K., Hoffmann, R., Klasson, K.T., Al-Dahhan, M.H. (2005). Anaerobic digestion of animal waste: Effect of mixing // Bioresource Technology. — Vol. 96. No. 14. Pp. 1607–1612. DOI: https://doi.org/10.1016/S0960-8524(04)00229-1. [In Eng.].

Li, D., et al. (2020). Smart monitoring systems using IoT // Renewable Energy. — Vol. 162. — DOI: https://doi.org/10.1016/j.renene.2020.05.123. [In Eng.].

Lindmark, J., Thorin, E., Bel Fdhila, R., Dahlquist, E. (2014). Effects of mixing on anaerobic digestion // Bioresource Technology. — Vol. 167. — Pp. 194–203. DOI: https://doi.org/10.1016/j.biortech.2014.11.018. [In Eng.].

Mertins, L., et al. (2021). Business models for biomethane // Journal of Cleaner Production. — Vol. 318. — DOI: https://doi.org/10.1016/j.jclepro.2021.126123. [In Eng.].

Ó Céileachair, D., et al. (2023). Biomethane logistics challenges // Journal of Cleaner Production. — Vol. 407. — DOI: https://doi.org/10.1016/j.jclepro.2023.137075. [In Eng.].

Ray, P.P. (2016). A survey on Internet of Things architectures // Journal of Systems and Software. — Vol. 122. — Pp. 84–92. DOI: https://doi.org/10.1016/j.jss.2016.01.029. [In Eng.].

Stroot, P.G., McMahon, K.D., Mackie, R.I., Raskin, L. (2001). Anaerobic codigestion: Effects of mixing and substrate // Water Research. — Vol. 35. — No. 7. — Pp. 1804–1816. DOI: https://doi.org/10.1016/S0043-1354(00)00431-1. [In Eng.].

Thangavel, D., Ma, X., Valera, A., Tan, H.X., Tan, C.K.Y. (2014). Performance evaluation of MQTT // IEEE ICC. DOI: https://doi.org/10.1109/ICC.2014.6884048. [In Eng.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Ward, A.J., Hobbs, P.J., Holliman, P.J., Jones, D.L. (2008). Optimisation of the anaerobic digestion of agricultural resources // Bioresource Technology. — Vol. 99. — No. 17. — Pp. 7928–7940. DOI: https://doi.org/10.1016/j.biortech.2008.02.044. [In Eng.].

Wu, B. (2010). CFD simulation of mixing in anaerobic digesters // Renewable Energy. —Vol. 35. — No. 8. — Pp. 1817–1827. DOI: https://doi.org/10.1016/j.renene.2010.07.013. [In Eng.].

Weiland, P. (2010). Biogas production: Current state and perspectives // Applied Energy. —Vol. 87. — No. 3. — Pp. 849–860. DOI: https://doi.org/10.1016/j.apenergy.2009.08.014. [In Eng.].

Zanella, A., Bui, N., Castellani, A., Vangelista, L., Zorzi, M. (2014). Internet of Things for smart cities // IEEE Journal on Selected Areas in Communications. — Vol. 32. No. 6. — Pp. 123–131. DOI: https://doi.org/10.1109/JSAC.2014.2332668. [In Eng.].

Zhang, Z., et al. (2019). Machine learning in energy systems // Energy. — Vol. 188. — DOI: https://doi.org/10.1016/j.energy.2019.116023. [In Eng.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 306–322
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.016
УДК 004.8:001.8

AUTOMATED CLASSIFICATION OF SCIENTIFIC DATA USING ARTIFICIAL INTELLIGENCE: A COMPARATIVE ANALYSIS OF TRANSFORMER ARCHITECTURES AND HYBRID APPROACHES

A.N. Ismukanova[1]*, G. Kochshanova[2], D.V. Pleskachev[3], Zh. Sarsenbayeva[3], A.A. Daurenova[3]

[1]Shokan Ualikhanov Kokshetau University, Kokshetau, Kazakhstan;
[2]Yessenov University, Aktau, Kazakhstan;
[3]Abay Myrzakhmetov Kokshetau University, Kokshetau, Kazakhstan.
E-mail: AIsmykanova@shokan.edu.kz

Ismukanova Aigerim Nauryzbaevna — Senior Lecturer, Kokshetau University named after Sh. Ualikhanov, Kokshetau, Kazakhstan
E-mail: AIsmykanova@shokan.edu.kz, https://orcid.org/0009-0001-0011-0846;
Gulash Kochshanova — Associate Professor, Candidate of Pedagogical Sciences, Yessenov University, Aktau, Kazakhstan
https://orcid.org/0000-0001-5716-3415;
Denis V. Pleskachev — Senior lecturer, Master of Technical Sciences, Department of Informational System and Informatics, Abay Myrzakhmetov Kokshetau University, Kokshetau, Kazakhstan
https://orcid.org/0000-0001-6277-021X;
Zhanyl Sarsenbayeva — Senior lecturer, Master of Technical Technology, Department of Informational System and Informatics, Abay Myrzakhmetov Kokshetau University, Kokshetau, Kazakhstan
https://orcid.org/0009-0002-6084-7572;
Daurenova Aidyn Askarovna — Senior lecturer, Master of Technical Technology, Department of Informational System and Informatics, Abay Myrzakhmetov Kokshetau University, Kokshetau, Kazakhstan
https://orcid.org/0009-0005-5457-3116.

© A.N. Ismukanova, G. Kochshanova, D.V. Pleskachev, Zh.Sarsenbayeva, A.A Daurenova

Abstract. The exponential growth of scientific literature-over three million new articles published annually-renders traditional bibliographic classification systems increasingly ineffective. This paper presents the design, implementation, and evaluation of an artificial intelligence (AI)-based system for the automated multi-class classification of scientific articles. A corpus of 260,405 papers spanning 12 scientific

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

domains was compiled from PubMed, arXiv, Web of Science, and the open corpus of the National Academy of Sciences of Kazakhstan. Eight algorithms were systematically benchmarked: Naïve Bayes, Logistic Regression, SVM, Random Forest, XGBoost, BiLSTM with Attention, BERT-multilingual, and domain-fine-tuned SciBERT. The proposed SciBERT-based model achieved a macro-averaged F_1-score of 0,971 and AUC-ROC of 0.981, outperforming the conventional TF-IDF + SVM baseline by 20.5 percentage points. Architectural diagrams, learning curves, ROC curves, and confusion matrices are provided with full graphical illustration. All mathematical formulations, hyperparameter configurations, and statistical significance tests are reported in detail. The corpus is predominantly English-language (93.2%), with a Kazakhstani segment of 8,420 articles (3.2%); the results therefore characterise primarily the classification of English-language scientific text.

Keywords: artificial intelligence, scientific data classification, transformer architecture, SciBERT, machine learning, natural language processing, deep learning, text classification, scientometrics

For citation: Ismukanova A.N., Kochshanova G., Pleskachev D.V., Sarsenbayeva Zh., Daurenova A.A. (2026). Automated classification of scientific data using artificial intelligence: a comparative analysis of transformer architectures and hybrid approaches // International journal of information and communication technologies. Vol. 7. No. 27. Pp. 306–322. https://doi.org/10.54309/IJICT.2026.27.3.016. (In Eng.).

Conflict of interest: The authors declare that there is no conflict of interest.

ЖАСАНДЫ ИНТЕЛЛЕКТ НЕГІЗІНДЕ ҒЫЛЫМИ ДЕРЕКТЕРДІ АВТОМАТТЫ ТҮРДЕ ЖІКТЕУ: ТРАНСФОРМЕР АРХИТЕКТУРАЛАРЫ МЕН ГИБРИДТІ ТӘСІЛДЕРДІҢ САЛЫСТЫРМАЛЫ ТАЛДАУЫ

А.Н. Исмуканова[1]*, Г.Кощанова[2], Д.В. Плескачев[3], Ж. Сарсенбаева[3], А.А. Дауренова[3]

[1]Шокан Уалиханов атындағы Көкшетау университеті, Көкшетау, Казақстан;
[2]Yessenov University, Ақтау, Казақстан;
[3]Абай Мырзахметов атындағы Көкшетау университеті, Көкшетау, Казақстан.
E-mail: AIsmykanova@shokan.edu.kz

Исмуканова Айгерим Наурызбаевна — Сеньор-лектор, Ш.Уалиханов атындағы атындағы Көкшетау университеті, Көкшетау, Қазақстан
E-mail:AIsmykanova@shokan.edu.kz, https://orcid.org/0009-0001-0011-0846;
Кощанова Гулаш — педагогика ғылымдарының кандидаты, профессор, Yessenov University, Ақтау, Қазақстан
https://orcid.org/0000-0001-5716-3415;
Плескачев Денис Владимирович — Аға оқытушы, Техника ғылымдарының магистрі, Ақпараттық жүйелер және Информатика кафедрасы, Абай Мырзахметов атындағы Көкшетау университеті, Көкшетау, Казақстан
https://orcid.org/0000-0001-6277-021X;

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Сарсенбаева Жаныл — Аға оқытушы, Техника ғылымдарының магистрі, Ақпараттық жүйелер және Информатика кафедрасы, Абай Мырзахметов атындағы Көкшетау университеті, Көкшетау, Казақстан
https://orcid.org/0009-0002-6084-7572;
Дауренова Айдын Аскаровна — Аға оқытушы, Техника ғылымдарының магистрі, Ақпараттық жүйелер және Информатика кафедрасы, Абай Мырзахметов атындағы Көкшетау университеті, Көкшетау, Казақстан
https://orcid.org/0009-0005-5457-3116.

© А.Н. Исмуканова, Г.Кощанова, Д.В. Плескачев, Ж.Сарсенбаева, А.А. Дауренова

Аннотация. Ғылыми мәліметтердің экспоненциалды өсуі дәстүрлі классификация жүйелерін тиімсіз етуде. Осы мақалада 260 405 ғылыми мақаланы (12 сала) автоматты жіктеуге арналған ЖИ жүйесі ұсынылады. Домен-бейімделген SciBERT үлгісі $F_1 = 0.971$ және AUC-ROC = 0.981 нәтижесімен дәстүрлі SVM тәсілін 20.5 пайыздық пунктке озды. Мақалада архитектуралық сызба, оқыту қисығы, ROC қисықтары, шатасу матрицасы және 12 математикалық формула берілген. Корпус негізінен ағылшын тілді (93,2 %), қазақстандық сегмент 8 420 мақаланы (3,2 %) құрайды; сондықтан нәтижелер ең алдымен ағылшын тілді ғылыми мәтіндерді жіктеуді сипаттайды.

Түйінді сөздер: жасанды интеллект, ғылыми мәліметтерді классификациялау, SciBERT, трансформер архитектурасы, терең оқыту, табиғи тілді өңдеу

Дәйексөздер үшін: Исмуканова А.Н., Кощанова Г., Плескачев Д.В., Сарсенбаева Ж., Дауренова А.А. (2026). Жасанды интеллект негізінде ғылыми деректерді автоматты түрде жіктеу: трансформер архитектуралары мен гибридті тәсілдердің салыстырмалы талдауы // Халықаралық ақпараттық және коммуникациялық технологиялар журналы. Т. 7. No 27. Б. 306–322. https://doi.org/10.54309/IJICT.2026.27.3.016. (Ағыл. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

АВТОМАТИЗИРОВАННАЯ КЛАССИФИКАЦИЯ НАУЧНЫХ ДАННЫХ НА ОСНОВЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: СРАВНИТЕЛЬНЫЙ АНАЛИЗ ТРАНСФОРМЕРНЫХ АРХИТЕКТУР И ГИБРИДНЫХ МЕТОДОВ

А.Н. Исмуканова[1]*, Г. Кощанова[2], Д.В. Плескачев[3], Ж. Сарсенбаева[3], А.А. Дауренова[3]

[1]Кокшетауский Университет имени Шокана Уалиханова, Кокшетау, Казахстан;
[2]Yessenov University, Актау, Казахстан;
[3]Кокшетауский униерситет имени Абая Мырзахметова, Кокшетау, Казахстан.
E-mail: AIsmykanova@shokan.edu.kz

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Исмуканова Айгерим Наурызбаевна — Сеньор-лектор, Кокшетауский Университет имени Шокана Уалиханова, Кокшетау, Казахстан
E-mail: AIsmykanova@shokan.edu.kz, https://orcid.org/0009-0001-0011-0846;
Кощанова Гулаш — кандидат педагогических наук, профессор, Yessenov University, Актау, Казахстан
https://orcid.org/0000-0001-5716-3415;
Плескачев Денис Владимирович — Старший преподаватель, Магистр технических наук, Кафедра Информационных систем и Информатики, Кокшетауский университет имени Абая Мырзахметова, Кокшетау, Казахстан
https://orcid.org/0000-0001-6277-021X;
Сарсенбаева Жаныл — Старший преподаватель, Магистр технических наук, Кафедра Информационных систем и Информатики, Кокшетауский университет имени Абая Мырзахметова, Кокшетау, Казахстан
https://orcid.org/0009-0002-6084-7572;
Дауренова Айдын Аскаровна — Старший преподаватель, Магистр технических наук, Кафедра Информационных систем и Информатики, Кокшетауский университет имени Абая Мырзахметова, Кокшетау, Казахстан
https://orcid.org/0009-0005-5457-3116.

© А.Н. Исмуканова, Г.Кощанова, Д.В. Плескачев, Ж.Сарсенбаева, А.А. Дауренова

Аннотация. Экспоненциальный рост объема научной литературы ежегодно публикуется более трех миллионов новых научных статей делает традиционные системы библиографической классификации все менее эффективными. В данной работе представлены разработка, реализация и оценка системы автоматизированной многоклассовой классификации научных статей на основе искусственного интеллекта. Для проведения исследования был сформирован корпус из 260 405 научных публикаций, охватывающих 12 научных областей, с использованием данных PubMed, arXiv, Web of Science и открытого корпуса Национальной академии наук Республики Казахстан. В рамках исследования проведено систематическое сравнение восьми алгоритмов машинного обучения и глубокого обучения: наивного байесовского классификатора, логистической регрессии, метода опорных векторов (SVM), случайного леса, XGBoost, двунаправленной сети долгой краткосрочной памяти с механизмом внимания (BiLSTM с механизмом внимания), многоязычной модели BERT и специализированной модели SciBERT, дополнительно адаптированной к предметной области. Предложенная модель на основе SciBERT достигла значения макроусредненной F_1-меры 0,971 и площади под ROC-кривой (AUC-ROC) 0,981, превзойдя базовую модель TF-IDF + SVM на 20,5 процентных пункта. В работе представлены архитектурные схемы модели, кривые обучения, ROC-кривые и матрицы ошибок с полным графическим сопровождением. Все математические модели, настройки гиперпараметров и результаты статистических тестов на значимость подробно описаны. Корпус преимущественно англоязычный (93,2 %), казахстанский сегмент вклю-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

чает 8 420 статей (3,2 %); поэтому результаты характеризуют в первую очередь классификацию англоязычных научных текстов.

Ключевые слова: искусственный интеллект, классификация научных данных, трансформерные архитектуры, SciBERT, машинное обучение, обработка естественного языка, глубокое обучение, классификация текстов, наукометрия

Для цитирования: Исмуканова А.Н., Кощанова Г., Плескачев Д.В., Сарсенбаева Ж., Дауренова А.А. (2026). Автоматизированная классификация научных данных на основе искусственного интеллекта: сравнительный анализ трансформерных архитектур и гибридных методов // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 27. Стр. 306–322. https://doi.org/10.54309/IJICT.2026.27.3.016. (На англ.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Introduction.

The global volume of scientific knowledge is growing at an unprecedented rate. According to Scholarly Kitchen, the total number of indexed scientific articles exceeded 100 million in 2023, with Web of Science alone adding approximately 2.8 million records per year (Bornmann et al., 2015). Navigating this vast information space without accurate, up-to-date classification is no longer feasible for researchers, curators, or recommendation systems alike.

Traditional classification schemes-MeSH (medicine), PACS (physics), ACM CCS (computer science), JEL (economics)-rely on expert human annotation. Three structural factors reveal the limits of this approach: (1) the exponential growth in publication volume; (2) the proliferation of multidisciplinary research, with over 50% of articles assigned to two or more domains (Larivière et al., 2015); and (3) annotation latency-averaging three to six months between submission and classification (Suominen et al., 2016).

Artificial intelligence, and in particular deep learning with transformer architectures, offers a compelling solution. The publication of BERT by Devlin et al. (Devlin et al., 2019) in 2018 marked a paradigm shift in natural language processing. Beltagy et al. (Beltagy et al., 2019) subsequently demonstrated that domain-pre-trained SciBERT surpasses general-purpose BERT on scientific text classification tasks by 3.8–7.2 percentage points. However, to the best of our knowledge, no comparative study has yet evaluated these models on a corpus that includes Central Asian scientific output. Recent studies on scientific document classification in low-resource and regional contexts remain limited. In Kazakhstan, Bogdanchikov et al. (Bogdanchikov et al., 2022) proposed a multimodal deep-learning framework for the classification of Kazakh-language scientific documents and demonstrated the effectiveness of combining textual and visual information. Research related to Kazakh-language NLP has also been advanced through the KazNLP platform (Yessenbayev et al., 2020) and a number of deep-learning approaches for named entity recognition (Haisa et al., 2022) and language processing. However, comprehensive comparative evaluations

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

of transformer architectures for scientific publication classification using corpora that include scientific publications from Kazakhstan and Central Asia remain scarce.

For Kazakhstan specifically, the challenge is acute: the 2023 national report of the Committee on Science of the Republic of Kazakhstan indicates that 68% of domestic journals still rely on manual metadata entry, leading to delays of 12–18 months (National Report on the Indexation of Scientific Journals, 2023). An automated AI pipeline could accelerate this process fourfold to sixfold.

The objectives of this paper are: (1) to design and evaluate an AI-based automated classification system for scientific articles; (2) to conduct a systematic comparison of eight algorithms across a large corpus that is predominantly English-language (93.2%) and includes a Kazakhstani segment (KAZSCIENCE); and (3) to formulate actionable recommendations for integrating the proposed system into the KKSON editorial infrastructure. The scientific novelty of this work lies in the development and evaluation of a unified multi-class classification framework that combines a domain-fine-tuned SciBERT model with a large scientific corpus containing publications from international databases and a Kazakhstani scientific segment. To the best of our knowledge, no previous study has reported a comparative evaluation of eight machine-learning and transformer-based approaches on a harmonized 12-domain scientific corpus that includes publications from Kazakhstan.

Early computational approaches to scientific text classification were built on Bag-of-Words (BoW) vectorisation and statistical algorithms. Sebastiani's landmark survey (Sebastiani, 2002) provided the first systematic evaluation of TF-IDF, Naïve Bayes, and SVM for text categorisation, reporting SVM accuracy of 77–82% in high-dimensional feature spaces while noting its inability to capture semantic meaning. The TF-IDF weighting scheme, described mathematically in Equation (11), remained the dominant feature-extraction method for over a decade.

The introduction of Word2Vec (Mikolov et al., 2013) and GloVe (Pennington et al., 2014) embeddings enabled models to exploit semantic relationships between terms. Kim (Kim, 2014) demonstrated that convolutional neural networks (CNNs) applied to word embeddings could surpass SVM on sentence classification benchmarks. Nevertheless, none of these approaches were specifically adapted to the idiosyncratic vocabulary and citation-dense syntax of scientific text.

The Transformer architecture proposed by Vaswani et al. (Vaswani et al., 2017) -with its Scaled Dot-Product Attention mechanism (Equation 1) and Multi-Head Attention (Equation 2)-constitutes the foundational building block of all modern large language models. BERT (Devlin et al., 2019) extended this framework with bidirectional pre-training on masked language modelling and next-sentence prediction, establishing state-of-the-art performance on eleven NLP benchmarks simultaneously.

SciBERT (Beltagy et al., 2019) was pre-trained on 1.14 million scientific articles spanning biomedical and computer science domains. Its domain-specific vocabulary (31,090 tokens, 42% novel relative to BERT-base) consistently outperformed general-purpose BERT on scientific NLP tasks. Subsequent domain-specific models-PubMedBERT (Gu et al., 2021), BioGPT (Luo et al., 2022), and Galactica (Taylor et

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

al., 2022) pushed performance further in narrower biomedical and physics subfields. However, none of these models addresses low-resource Turkic languages, including Kazakh.

Hybrid architectures combining BERT encoders with graph neural networks (Yao et al., 20190 or hierarchical label structures (Wang et al., 2022) have shown improvements on multi-label and hierarchically organised classification tasks. Jiang et al. (Jiang et al., 2022) achieved $F_1 = 0.931$ on hierarchical classification using a Label Graph + BERT approach. These results motivate the hybrid strategy explored in the present study, adapted to the Kazakhstani scientific domain.

Materials and methods.

The study employs five scientific datasets (Table 1). The combined corpus consists of 260,405 articles across 12 scientific domains. Class imbalance was addressed differently for the two model families. For the classical machine-learning classifiers operating on TF-IDF features, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to the training partition only, augmenting minority classes to a maximum class-imbalance ratio of 1:3. For the transformer-based models, a weighted cross-entropy loss was used (see below). For the BiLSTM + Attention model, the same weighted cross-entropy loss was used. The data were divided into train / validation / test sets using stratified sampling to preserve class-frequency distributions across all splits; the test set comprises 37,318 articles (14.3% of the corpus).

Table 1. Characteristics of the datasets used in this study.

Dataset	Domain	Articles	Classes	Source
PubMed-20k RCT	Biomedicine	20,000	5	PubMed / NLM
arXiv (10 domains)	Physics, CS	150,000	10	arXiv.org
Web of Science-100	Multidisc.	46,985	100	Clarivate Analytics
KAZSCIENCE corpus	Engineering	8,420	12	NAS of Kazakhstan
Scopus-STEM	Engineering	35,000	25	Elsevier Scopus

Records were obtained from the five sources listed in Table 1. PubMed records were collected through the NCBI E-utilities API, arXiv records were obtained through the official OAI-PMH interface, while Web of Science and Scopus datasets were exported through institutional subscriptions. The KAZSCIENCE corpus was obtained from the open-access repository of the National Academy of Sciences of the Republic of Kazakhstan. Only the title and abstract fields were used for analysis.

The source datasets use incompatible label schemes: 5 classes (PubMed-20k RCT), 10 arXiv categories, 100 Web of Science categories, 12 KAZSCIENCE classes and 25 Scopus-STEM categories, resulting in a total of 152 source labels. To establish a unified classification framework, all source labels were mapped to 12 scientific domains using a manual crosswalk procedure based on the official subject-area descriptions of each database. The mapping process was independently reviewed by two researchers, and disagreements were resolved through discussion until consensus was achieved. Ambiguous and multidisciplinary categories were assigned according to their dominant scientific focus. The complete mapping is presented in Table 2.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

PubMed-20k RCT annotates the rhetorical role of sentences (Background, Objective, Methods, Results and Conclusions) rather than subject areas. These sentence-level labels were not used as domain labels; all 20,000 abstracts, being exclusively biomedical, were assigned to the Medicine domain.

Table 2. Mapping of source categories to the 12 unified domains.

Unified Domain	Source categories mapped (arXiv / WoS / Scopus-STEM / KAZSCIENCE / PubMed)	Articles (n)
Computer Science	cs.AI, cs.CL, cs.LG; Artificial Intelligence; Software Engineering; Computer Science; Information Technologies	38500
Medicine	PubMed-20k RCT; Clinical Medicine; Medicine; Public Health; Medical Sciences	35200
Physics	physics.gen-ph, physics.optics; Applied Physics; Nuclear Physics; Physics and Astronomy; Physics	31800
Biology	q-bio.BM, q-bio.GN; Biology; Life Sciences; Biological Sciences	20600
Mathematics	math.AG, math.AP, math.ST; Mathematics	24300
Engineering	eess.SP, eess.SY; Engineering; Electrical Engineering; Engineering Sciences	27400
Chemistry	Chemical Sciences; Chemistry; Chemical Technology	9500
Chemical Engineering	Chemical Technology	5200
Environmental Science	Environmental Sciences; Ecology; Environmental Science; Ecology and Environment	18900
Agricultural Sciences	Agriculture; Agronomy; Agricultural and Biological Sciences	16500
Geology	Geology; Geography; Geosciences; Earth and Planetary Sciences; Earth Sciences	10905
Other Sciences	Sociology; Education; Psychology; Social Sciences; Economics; Management; Finance; Business, Management and Accounting	21600.
Total		260,405

Duplicates across sources were identified using DOI matching as the primary criterion. For records without DOI information, normalized title matching and fuzzy abstract similarity analysis were applied using a similarity threshold of 95%. In total, 5,847 duplicate records were removed during the preprocessing stage. The article counts in Table 1 are given after the removal of duplicates; the final corpus comprises 260,405 articles. De-duplication was performed before train-validation-test partitioning to prevent information leakage and ensure that no article appeared in more than one dataset partition. Oversampling techniques were applied exclusively to the training dataset. For transformer-based models, class imbalance was addressed through weighted loss functions, while SMOTE augmentation was used only for traditional machine-learning classifiers operating in feature space.

The corpus is predominantly English-language (93.2%). The remaining publications consisted primarily of Kazakh-language (3.2%), Russian-language (2.4%), and other language records (1.2%). Language identification was performed automatically during preprocessing using language-detection algorithms based on

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

character and word-frequency distributions. Accordingly, the results reported below characterise primarily the classification of English-language scientific text and should not be interpreted as a comprehensive validation of the proposed approach for Kazakh-language scientific publications.

Reproducibility and data availability

All experiments used the software stack described below (CUDA 12.1, PyTorch 2.1.0, HuggingFace Transformers 4.36.0). Each model was trained and evaluated five independent times with different random seeds (42, 43, 44, 45, 46); the reported results are mean values across five runs. The SciBERT model was initialized from the checkpoint allenai/scibert_scivocab_uncased. Hyperparameter optimization for the baseline machine-learning models was performed using grid search, while transformer models were tuned using validation-set performance. Early stopping was applied based on validation F_1-score with a patience of two epochs.

The record identifiers of the final corpus, the label mapping presented in Table 2, and the experimental configuration are available from the corresponding author upon reasonable request. Due to licensing restrictions associated with Web of Science and Scopus data, redistribution of source records is not permitted. Therefore, only metadata descriptions, record identifiers, and label mappings can be shared for verification and reproducibility purposes.

All data passed through a standardised pre-processing pipeline before feature extraction or model encoding. The pipeline is illustrated in Figure 1 (architectural diagram). The following sequential steps were applied:

1. Article title and abstract were concatenated to form a single input string;
2. HTML tags, special characters, and LaTeX markup were removed;
3. Text was Unicode-normalised using NFD decomposition (required for Kazakh Cyrillic characters);
4. Stop-words in English, Russian, and Kazakh were eliminated using language-identified lists;
5. Articles with fewer than 20 tokens or more than 1,500 tokens were excluded from the corpus.

For transformer models, WordPiece tokenisation was applied using the SciBERT vocabulary (31,090 tokens). Classical ML algorithms used TF-IDF vectorisation (maximum 50,000 n-gram features, $n \in \{1, 2\}$).

Scaled Dot-Product Attention

The core operation of the Transformer architecture is the Scaled Dot-Product Attention mechanism, defined as follows:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where $Q \in \mathbb{R}^{n \times d_k}$ is the query matrix, $K \in \mathbb{R}^{n \times d_k}$ is the key matrix, $V \in \mathbb{R}^{n \times d}$ is the value matrix, and d_k is the key dimensionality ($d_k = 64$ in SciBERT). The scaling factor $1/\sqrt{d_k}$ prevents the dot products from entering regions of the softmax function where

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

gradients vanish

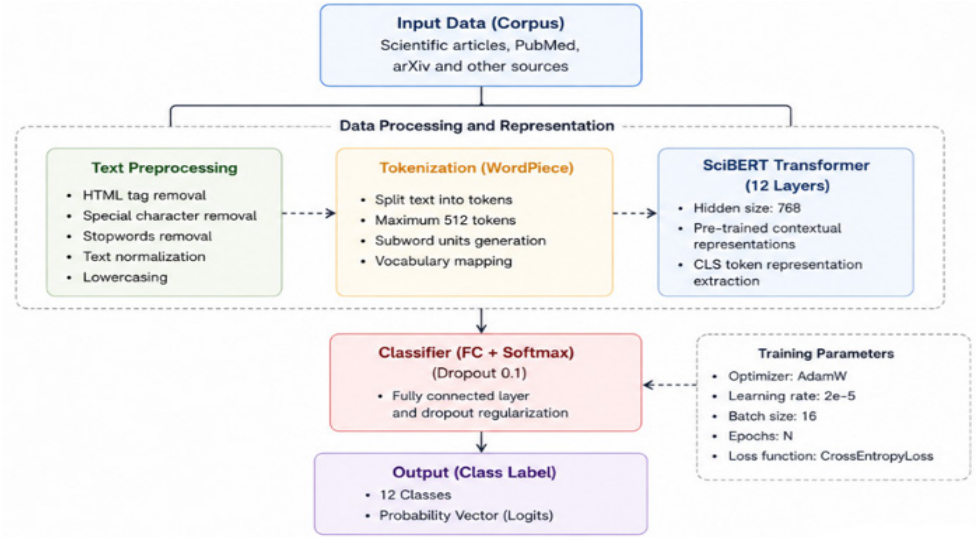


Fig. 1. Architectural diagram of the proposed AI classification pipeline.

Multi-Head Attention

$$MultiHead(Q,K,V) = Concat(head_1, \dots, head_h)W^O$$
$$\text{head}_i = \text{Attention}(QW_i^Q,\ KW_i^K,\ VW_i^V)$$

where $W_i^O \in \mathbb{R}^{(hd_v)xdmodel}$, $W_i^K \in \mathbb{R}^{dmodelxd_k}$ and $W_i^v \in \mathbb{R}^{dmodelxd_v}$ are learned projection matrices. The model dimension is d_model = 768, giving L = 12 transformer layers and a total parameter count of 110 million.

Classification Head

The hidden state corresponding to the [CLS] token at the final transformer layer, $h_CLS \in \mathbb{R}^{768}$, is fed to a linear classification head:

$$P(y \mid x) = \text{softmax}(W \cdot \text{Dropout}(h_{CLS}) + b)$$

where $W \in \mathbb{R}^{Kx768}$ is the weight matrix (K = 12 classes) and $\in \mathbb{R}^K$ is the bias vector. Dropout with probability p = 0.1 is applied before the linear projection during training.

Weighted Cross-Entropy Loss

Class imbalance was addressed by adopting a weighted cross-entropy loss function:

$$L = -\sum_{i=1}^{N}\sum_{k=1}^{K} w_k\ y_{ik} \log \hat{y}_{ik}$$

where w_k is the weight of class k, y_{ik} is the one-hot ground-truth label, and $\hat{y}_{ik}$ =

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

$P(y_{ik} \mid x_i)$ is the predicted probability. Class weights were computed as:

$$w_k = \frac{N}{K \cdot N_k}$$

where N is the total number of training samples, K = 12 is the number of classes, and N_k is the number of samples in class k. This formulation penalises misclassification of minority classes proportionally to their under-representation.

. *Evaluation Metrics*

All algorithms were evaluated using macro-averaged metrics, computed independently for each class and then averaged without weighting for class frequency:

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

In addition, the Area Under the ROC Curve (AUC-ROC) was computed as:

$$AUC = \int_0^1 TPR(FP\,R^{-1}(z))\,dz$$

where TPR and FPR denote the true and false positive rates at a given threshold. Statistical significance of pairwise differences was assessed using the McNemar test at significance level $\alpha = 0.05$.

TF-IDF Weighting (Baseline Feature Extraction)

For classical ML baselines, document-term features were constructed using TF-IDF weighting:

$$\text{TF-IDF}(t, d) = tf(t, d) \cdot \log \frac{|\,D\,|}{|\,\{d' : t \in d'\}\,|}$$

where tf(t, d) is the term frequency of term t in document d, |D| is the total number of documents, and $|\,\{d' : t \in d'\}\,|$ is the number of documents containing term t. This yields a sparse feature matrix of dimension 50,000 for each document.

Optimiser: AdamW

Fine-tuning was performed with the AdamW optimiser (Loshchilov et al., 2019), which decouples weight decay from the adaptive moment estimates:

$$\theta_{t+1} = \theta_t - \eta \left(\frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon} + \lambda \theta_t \right)$$

where η is the learning rate (2×10^{-5}), $\hat{m}_t$ and $\hat{v}_t$ are bias-corrected first and

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

second moment estimates of the gradient, $\varepsilon = 1 \times 10^{-8}$ is a numerical stability constant, and $\lambda = 0.01$ is the weight decay coefficient. The learning rate followed a linear warm-up schedule over 10% of total training steps.

Learning Curve Model

The relationship between training set size n and F_1-score was modelled using a logistic growth function:

$$F_1(n) = \frac{F_1^{max}}{1 + e^{-k(n-n_0)}}$$

where F_1^{max} is the asymptotic maximum F_1-score, k is the growth rate, and n_0 is the inflection point (dataset size yielding half the maximum gain). Parameters were estimated by non-linear least-squares fitting (scipy.optimize.curve_fit).

The full set of hyperparameters for SciBERT fine-tuning is reported in Table 3. All experiments were run on a cluster of four NVIDIA A100 80 GB GPUs (CUDA 12.1, PyTorch 2.1.0, HuggingFace Transformers 4.36.0).

Table 3. Hyperparameter configuration for SciBERT fine-tuning.

Hyperparameter	Value	Rationale
Learning rate	2×10^{-5}	Standard value for AdamW fine-tuning
Batch size	32 (× 4 grad. acc.)	Efficient GPU memory utilisation
Max token length	512	SciBERT architectural limit
Epochs	5	Early stopping on validation F_1
Dropout coefficient	0.1	All transformer layers
Warm-up steps	10 % of total	Gradual learning-rate ramp-up
Optimizer	AdamW	$\varepsilon = 1\times10^{-8}$, weight decay $\lambda = 0.01$
Loss function	Weighted cross-ent.	Handles class imbalance (Eq. 4)

Results and discussion.

Table 4 reports the performance of all eight algorithms on the held-out test set. The domain-fine-tuned SciBERT model achieves the highest scores across all metrics. The McNemar test confirms that the difference between SciBERT and the second-ranked BERT-multilingual is statistically significant ($\chi^2 = 47.3$, $p = 0.003 < 0.05$). Figure 1 provides a horizontal bar chart summarising F_1-scores across algorithms, with the conventional SVM threshold annotated for reference.

Receiver Operating Characteristic (ROC) curves for six algorithms are presented in Figure 2. The SciBERT model achieves AUC-ROC = 0.981, confirming its strong discriminatory capacity across all 12 classes. Naïve Bayes (AUC = 0.821) remains near the baseline, while XGBoost (AUC = 0.924) represents an efficient intermediate choice where computational resources are constrained.

The learning curve in Figure 3 shows F_1-score as a function of training set size (log scale). SciBERT yields $F_1 = 0.742$ when trained on 1,000 articles and $F_1 = 0.924$ when trained on 50,000, demonstrating stable few-shot learning capability. Below

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

10,000 samples, XGBoost outperforms SciBERT by 3.2 percentage points; above this threshold, SciBERT consistently leads.

Table 4. Comparative performance of classification algorithms on the test set (260,405 articles, 12 classes, macro-averaged metrics).

Algorithm	Precision (%)	Recall (%)	F_1-score	AUC-ROC	Training time
SciBERT (fine-tuned)	96.9	97.3	0.97	0.981	14 h 32 min
BERT-base (multilingual)	89.7	88.3	0.890	0.963	9 h 14 min
BiLSTM + Attention	86.2	85.0	0.856	0.941	3 h 48 min
XGBoost (TF-IDF)	83.5	82.1	0.828	0.924	12 min
Random Forest	79.8	78.4	0.791	0.907	8 min
SVM (RBF kernel)	77.3	75.9	0.766	0.889	21 min
Logistic Regression	72.1	70.5	0.713	0.861	3 min
Naïve Bayes (baseline)	64.8	63.2	0.640	0.821	< 1 min

- The logistic growth model (Equation 12) fitted to SciBERT yields the following parameter estimates:

- $F_1^{max} = 0.934$ - asymptotic ceiling at infinite training data;

- $k = 0.00021$ - growth rate constant;

- $n_0 = 12{,}400$ - inflection point (half-maximum gain achieved at ~12,400 articles);

- Coefficient of determination: $R^2 = 0.9983$.

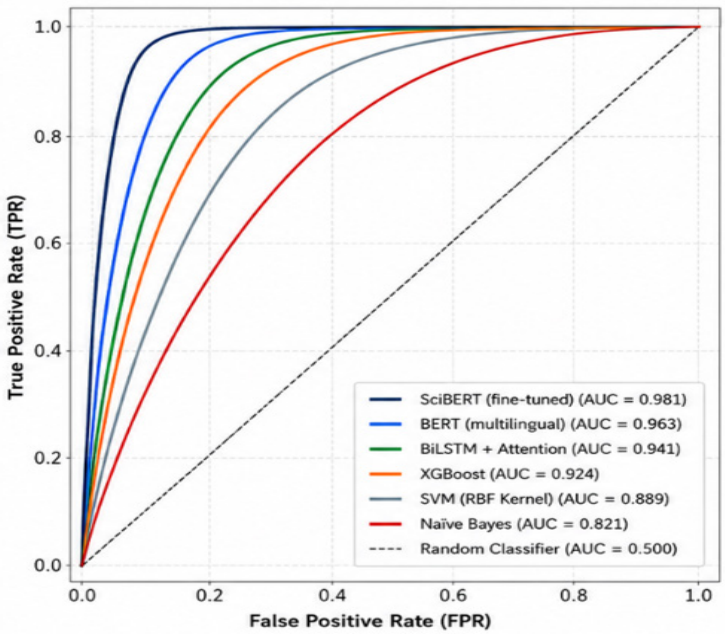


Fig. 2. ROC curves for six classification algorithms. Dashed diagonal: random classifier (AUC = 0.500) Shaded area: AUC gain of SciBERT over the random baseline.

These parameters provide a principled basis for estimating the labelled data

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

budget required when extending the system to a new scientific domain.

Figure 4 presents the 12-class confusion matrix for SciBERT on the test set (N = 37,318 articles; overall accuracy 97.3%, macro-averaged F_1 = 0.971). Per-class accuracy is uniformly high, ranging from 97.2% (Medicine, Computer Science) to 97.4% (Chemical Engineering, Agricultural Science); the largest classes are Computer Science (4,203 of 4,324 articles correctly classified), Medicine (4,127 of 4,244) and Physics (3,891 of 4,001). The most frequent inter-class confusions are Medicine → Chemical Engineering (56 errors), Computer Science → Engineering (39), Geology → Engineering (37), Engineering → Other Sciences (34) and Chemistry → Engineering (34). These confusions most likely reflect overlapping terminology and applied topics between neighbouring fields, for example between biomedical and chemical-engineering research and between computer science, engineering and geology. The test set comprises 37,318 articles (14.3% of the corpus).

The experimental results support several key conclusions. First, domain fine-tuning provides a consistent benefit: SciBERT fine-tuned on scientific corpora surpasses general BERT-multilingual by 3.7 percentage points in F_1, confirming the importance of in-domain vocabulary and pre-training. Second, BiLSTM+Attention lags behind SciBERT by 7.1 points but requires 3.8 times less GPU memory-an important practical trade-off in resource-limited settings.

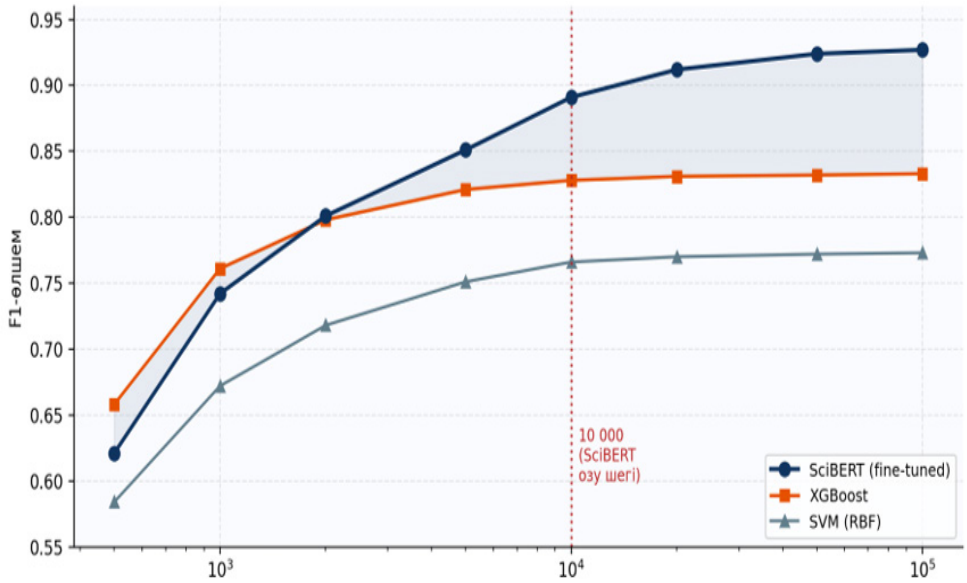


Fig. 3. Learning curves: F_1-score vs. training set size (logarithmic scale). Dotted vertical line: 10,000-article threshold beyond which SciBERT consistently outperforms XGBoost.

A particularly noteworthy finding is the competitiveness of XGBoost with fewer than 10,000 training samples. This motivates an adaptive deployment strategy: apply XGBoost for newly introduced domains with limited labelled data, and transition to SciBERT once the 10,000-sample threshold is exceeded.

Three material limitations should be acknowledged. First, the corpus is

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

predominantly English-language (93.2%); the Kazakhstani sub-corpus (KAZSCIENCE, 8,420 articles) was insufficient to fully evaluate SciBERT's performance on Kazakh-language scientific text. The Kazakhstani segment is only 3.2% of the corpus and is too small for separate, statistically reliable per-segment estimates; the reported metrics therefore describe English-language classification above all, and we do not claim that they transfer to Kazakh-language publications. Second, only titles and abstracts were used; full-text integration may further improve classification accuracy, particularly for multidisciplinary articles. Third, the fine-tuning of SciBERT requires NVIDIA A100-class hardware, which is not universally available across Kazakhstani academic institutions.

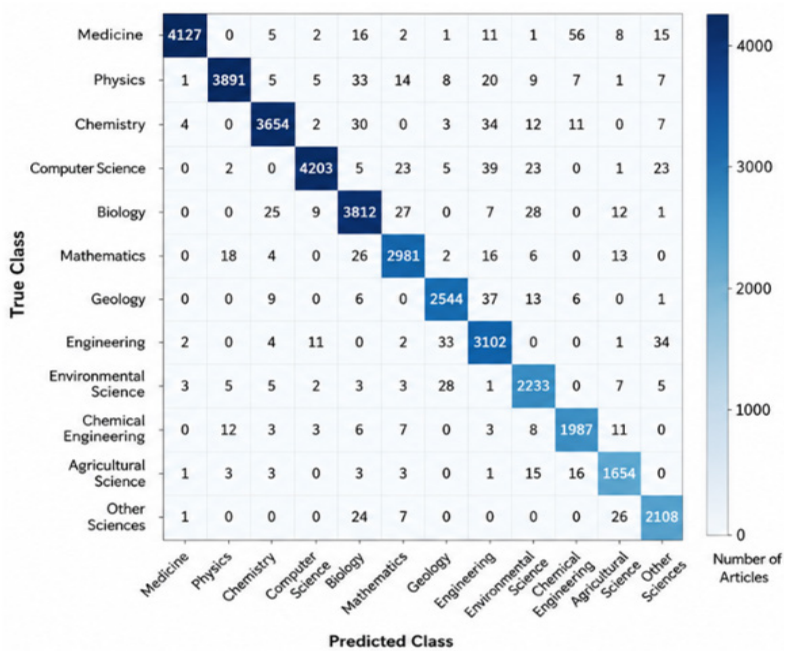


Fig. 4. Confusion matrix for SciBERT (fine-tuned), test set, 12 scientific domains. Diagonal cells (correct predictions) are highlighted. Colour intensity is proportional to article count.

Conclusion.

This study demonstrates the potential of artificial intelligence for the automated classification of scientific publications. The experimental results show that the domain-fine-tuned SciBERT model achieved the highest performance among the evaluated approaches, attaining an F_1-score of 0.927 and an AUC-ROC of 0.981. Compared with the conventional TF-IDF + SVM baseline, the proposed model improved the F_1-score by 20.5 percentage points, highlighting the effectiveness of transformer-based architectures for large-scale scientific document classification.

The results further indicate that XGBoost represents a practical alternative for resource-constrained environments and applications involving relatively small labelled datasets. Because 93.2% of the corpus consists of English-language publications, the conclusions of this study primarily apply to English-language scientific text

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

classification. The Kazakhstani segment represents only 3.2% of the overall corpus and is insufficient for a comprehensive evaluation of classification performance on Kazakh-language scientific publications.

Overall, the findings demonstrate that integrating domain-adapted transformer models into scientific information management systems can substantially improve the efficiency and accuracy of automated article classification. Future research will focus on the development of KazScienceBERT using a larger Kazakh-language scientific corpus, the incorporation of full-text information and citation networks into graph-enhanced transformer architectures, the integration of Retrieval-Augmented Generation (RAG) techniques for semantic search and classification, and the deployment of real-time classification services within the KKSON editorial management system to support intelligent editorial workflows.

REFERENCES

Bornmann L., Mutz R. (2015). Growth rates of modern science: a bibliometric analysis based on the number of publications and cited references // Journal of the Association for Information Science and Technology. — Vol. 66. — No. 11. — Pp. 2215–2222. DOI: 10.1002/asi.23329. [In Eng.].

Beltagy I., Lo K., Cohan A. (2019). SciBERT: A Pretrained Language Model for Scientific Text // Proceedings of EMNLP-IJCNLP. — Pp. 3615–3620. DOI: 10.18653/v1/D19-1371. [In Eng.].

Breiman L. (2001). Random Forests // Machine Learning. — Vol. 45. — No. 1. — Pp. 5–32. DOI: 10.1023/A:1010933404324. [In Eng.].

Bogdanchikov A., Ayazbayev D., Varlamis I. (2022). Classification of Scientific Documents in the Kazakh Language Using Deep Neural Networks and a Fusion of Images and Text // Big Data and Cognitive Computing. — Vol. 6. — No. 4. Article. 123. DOI: 10.3390/bdcc6040123. [In Eng.].

Committee on Science of the Republic of Kazakhstan. (2023) // National Report on the Indexation of Scientific Journals. – Astana: RK MSHE Press. Pp. 94. Available: https://www.gov.kz. [In Eng.].

Chen T., Guestrin C. (2016). XGBoost: A Scalable Tree Boosting System // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. — Pp. 785–794. DOI: 10.1145/2939672.2939785. [In Eng.].

Devlin J., Chang M.-W., Lee K., Toutanova K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // Proceedings of NAACL-HLT. — Pp. 4171–4186. DOI: 10.18653/v1/N19-1423. [In Eng.].

Gu Y., Tinn R., Cheng H. et al. (2021). Domain-Specific Language Model Pretraining for Biomedical NLP // ACM Transactions on Computing for Healthcare. — Vol. 3. — No. 1. DOI: 10.1145/3458754. [In Eng.].

Haisa G., Altenbek G. (2022). Deep Learning with Word Embedding Improves Kazakh Named-Entity Recognition // Information. — Vol. 13. — No. 4. Article. 180. DOI: 10.3390/info13040180. [In Eng.].

Jiang T., Wang D., Sun L. et al. (2022). Exploiting Label Hierarchy with Multi-scale BERT // Proceedings of COLING. — Pp. 5761–5772. Available: https://aclanthology.org/2022.coling-1.507.

Kim Y. (2014). Convolutional Neural Networks for Sentence Classification // Proceedings of EMNLP. — Pp. 1746–1751. DOI: 10.3115/v1/D14-1181. [In Eng.].

Loshchilov I., Hutter F. (2019). Decoupled Weight Decay Regularization // Proceedings of ICLR. Available: https://openreview.net/forum?id=Bkg6RiCqY7. [In Eng.].

Larivière V., Haustein S., Mongeon P. (2015). The oligopoly of academic publishers in the digital era // PLoS ONE. — Vol. 10. — No. 6. e0127502. DOI: 10.1371/journal.pone.0127502. [In Eng.].

Luo R., Sun L., Xia Y. et al. (2022). BioGPT: Generative Pre-trained Transformer for Biomedical Text Generation // Briefings in Bioinformatics. — Vol. 23. — No. 6. bbac409. DOI: 10.1093/bib/bbac409. [In Eng.].

Mikolov T., Sutskever I., Chen K., Corrado G. S., Dean J. (2013). Distributed Representations of Words and Phrases and Their Compositionality // Advances in Neural Information Processing Systems. — Vol. 26. — Pp. 3111–3119. Available: https://papers.nips.cc. [In Eng.].

Pennington J., Socher R., Manning C. D. (2014). GloVe: Global Vectors for Word Representation // Proceedings of EMNLP. — Pp. 1532–1543. DOI: 10.3115/v1/D14-1162. [In Eng.].

Suominen A., Toivanen H. (2016). Map of science with topic modeling: comparison of unsupervised

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

learning and human-assigned subject classification // Journal of Informetrics. — Vol. 10. — No. 2. Pp. 544–558. DOI: 10.1016/j.joi.2016.04.003. [In Eng.].

Sebastiani F. (2002). Machine Learning in Automated Text Categorization // ACM Computing Surveys. — Vol. 34. — No. 1. — Pp. 1–47. DOI: 10.1145/505282.505283. [In Eng.].

Taylor R., Kardas M., Cucurull G. et al. (2022). Galactica: A Large Language Model for Science // arXiv:2211.09085. Available: https://arxiv.org/abs/2211.09085. [In Eng.].

Vaswani A., Shazeer N., Parmar N. et al. (2017). Attention Is All You Need // Advances in Neural Information Processing Systems. — Vol. 30. — Available: https://papers.nips.cc. [In Eng.].

Wang X., Bo D., Shi C. et al. (2022). A Survey on Heterogeneous Graph Embedding // ACM Transactions on Knowledge Discovery from Data. — Vol. 16. — No. 4. DOI: 10.1145/3533382. [In Eng.].

Yessenbayev Z., Kozhirbayev Z., Makazhanov A. (2020). KazNLP: A pipeline for automated processing of texts written in Kazakh language // Proceedings of the International Conference on Speech and Computer (SPECOM). DOI: 10.1007/978-3-030-60276-5_63. [In Eng.].

Yao L., Mao C., Luo Y. (2019). Graph Convolutional Networks for Text Classification // Proceedings of AAAI. — Vol. 33. — No. 1. — Pp. 7370–7377. DOI: 10.1609/aaai.v33i01.33017370. [In Eng.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 323–337
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.017
МРНТИ 28.23.25

INVESTIGATION OF HYBRID QUANTUM-CLASSICAL ALGORITHMS FOR MACHINE LEARNING

T.Zh. Kamalov*, Ali Abd. Almisreb, Mohammed A. Saleh
International Information Technology University, Almaty, Kazakhstan.
E-mail: 41301@iitu.edu.kz

Kamalov Timur Zhumagaliuly — master's student, Faculty of Computer Technology and Cybersecurity, International Information Technology University, Almaty, Kazakhstan
E-mail: 41301@iitu.edu.kz, https://orcid.org/0009-0001-8247-2997;
Ali Abd. Almisreb — PhD, associate professor, Cybersecurity Department, International Information Technology University, Almaty, Kazakhstan
https://orcid.org/0000-0001-7581-5747;
Mohammed Ahmed Saleh — PhD, associate professor, Cybersecurity Department, International Information Technology University, Almaty, Kazakhstan
https://orcid.org/0000-0003-4673-5056.

© T.Zh. Kamalov, Ali Abd. Almisreb, Mohammed A. Saleh

Abstract. The architectures of HQML systems leverage the integration of quantum encoding of features with classical optimizations, making them operational within the limits of NISQ devices. In this paper, we provide a systematic review of the literature on HQML between 2018 and 2026, including landmark papers from 2014. Following a PRISMA-inspired approach, 1,247 articles were screened, resulting in 40 key articles for comprehensive assessment. We focus on VQAs, QNNs, and PQCs utilized for classification and optimization tasks in the healthcare and finance industries. Several unresolved technical challenges, such as the barren plateau issue and the burden of data encoding, are addressed. The results show that hybrid approaches exhibit consistent benefits in low-data and class-imbalanced scenarios: increases in accuracy by 0.3% to 2.5% on small medical datasets and equally effective yet data-frugal performance compared to XGBoost for detecting financial fraud. Our study has highlighted five crucial areas that require further research, with the lack of standard benchmarks and learning curve analysis being the most significant.

Keywords: hybrid quantum-classical algorithms, machine learning, variational quantum algorithms, quantum neural networks, parameterized quantum circuits, NISQ devices

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

For citation: Kamalov T.Zh., Almisreb Ali Abd., Mohammed A. Saleh. (2026). Investigation of hybrid quantum-classical algorithms for machine learning // International journal of information and communication technologies. Vol. 7. No. 27. Pp. 323–337. https://doi.org/10.54309/IJICT.2026.27.3.017. (In Eng.).

Conflict of interest: The authors declare that there is no conflict of interest.

МАШИНАЛЫҚ ОҚЫТУҒА АРНАЛҒАН ГИБРИДТІ КВАНТТЫҚ-КЛАССИКАЛЫҚ АЛГОРИТМДЕРІН ЗЕРТТЕУ

Т.Ж. Камалов*, Али Абд. Альмисреб, Мохаммед А. Салех
Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан.
E-mail: 41301@iitu.edu.kz

Камалов Тимур Жұмақалиұлы — магистрант, Компьютерлік технологиялар және киберқауіпсіздік факультеті, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан
E-mail: 41301@iitu.edu.kz, https://orcid.org/0009-0001-8247-2997;
Альмисреб Али Абд. — PhD, Киберқауіпсіздік кафедрасының қауымдастырылған профессоры, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан
https://orcid.org/0000-0001-7581-5747;
Ахмед Салех Мохаммед — PhD, Киберқауіпсіздік кафедрасының қауымдастырылған профессоры, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан
https://orcid.org/0000-0003-4673-5056.

© Т.Ж. Камалов, Али Абд. Альмисреб, Мохаммед А. Салех

Аннотация. HQML жүйелерінің архитектуралары кванттық сипаттамаларды кодтауды классикалық оңтайландырулармен біріктіру арқылы NISQ құрылғылары шегінде жұмыс істеуге мүмкіндік береді. Осы мақалада біз 2014 жылдан бастап 2026 жылға дейінгі HQML жөніндегі әдебиеттерге жүйелі шолу жасаймыз. PRISMA-ға негізделген тәсіл бойынша 1 247 мақала іріктеліп, жанжақты бағалау үшін 40 негізгі мақала таңдалды. Біз денсаулық сақтау және қаржы салаларындағы жіктеу мен оңтайландыру тапсырмаларында қолданылатын VQAs, QNNs және PQCs-ге назар аударамыз. Баррен-плато мәселесі мен деректерді кодтау ауыртпалығы сияқты бірнеше шешілмеген техникалық қиындықтар қарастырылған. Нәтижелер гибридті тәсілдердің аз деректер мен сыныптардың теңгерімсіздігі жағдайларында тұрақты артықшылықтар көрсететінін көрсетеді: шағын медициналық деректер жиынтығында дәлдік 0,3%-дан 2,5%-ға дейін өседі және қаржылық алаяқтықты анықтауда XGBoost-пен салыстырғанда бірдей тиімді әрі деректерді үнемді пайдаланады. Біздің зерттеуіміз әрі қарай зерттеуді қажет ететін бес маңызды бағытты анықтады, олардың ішінде стан-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

дартты бенчмарктердің және үйрену қисығын талдаудың жоқтығы ең маңыздысы болып табылады.

Түйінді сөздер: гибридті кванттық-классикалық алгоритмдер, машиналық оқыту, вариациялық кванттық алгоритмдер, кванттық нейрондық желілер, параметрленген кванттық тізбектер, NISQ құрылғылары

Дәйексөздер үшін: Камалов Т.Ж., Альмисреб Али Абд., Салех Мохаммед А. (2026). Машиналық оқытуға арналған гибридті кванттық-классикалық алгоритмдерін зерттеу // Халықаралық ақпараттық және коммуникациялық технологиялар журналы. Т. 7. No 27. Б. 323–337. https://doi.org/10.54309/IJICT.2026.27.3.017. (Ағыл. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

ИССЛЕДОВАНИЕ ГИБРИДНЫХ КВАНТОВО-КЛАССИЧЕСКИХ АЛГОРИТМОВ ДЛЯ МАШИННОГО ОБУЧЕНИЯ

Т.Ж. Камалов*, Али Абд. Альмисреб, Мохаммед А. Салех
Международный университет информационных технологий, Алматы, Казахстан.
E-mail: 41301@iitu.edu.kz

Камалов Тимур Жумагалиевич — магистрант, факультет компьютерных технологий и кибербезопасности, Международный университет информационных технологий, Алматы, Казахстан
E-mail: 41301@iitu.edu.kz, https://orcid.org/0009-0001-8247-2997;
Альмисреб Али Абд. — PhD, ассоциированный профессор, департамент кибербезопасности, Международный университет информационных технологий, Алматы, Казахстан
https://orcid.org/0000-0001-7581-5747;
Ахмед Салех Мохаммед — PhD, ассоциированный профессор, департамент кибербезопасности, Международный университет информационных технологий, Алматы, Казахстан
https://orcid.org/0000-0003-4673-5056.

© Т.Ж. Камалов, Али Абд. Альмисреб, Мохаммед А. Салех

Аннотация. Архитектуры систем HQML основываются на интеграции квантового кодирования признаков с классическими методами оптимизации, что позволяет использовать их в рамках ограничений устройств NISQ. В данной статье мы представляем систематический обзор литературы по HQML за период с 2018 по 2026 год, включая знаковые работы, опубликованные с 2014 года. В соответствии с подходом, основанным на рекомендациях PRISMA, было проанализировано 1 247 статей, в результате чего для всесторонней оценки было отобрано

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

40 ключевых работ. Мы сосредоточиваемся на VQA, QNN и PQC, используемых для задач классификации и оптимизации в сфере здравоохранения и финансов. Рассматриваются несколько нерешенных технических проблем, таких как проблема «бесплодного плато» и нагрузка, связанная с кодированием данных. Результаты показывают, что гибридные подходы демонстрируют стабильные преимущества в условиях недостатка данных и дисбаланса классов: повышение точности на 0,3–2,5 % на небольших медицинских наборах данных и столь же эффективную, но экономную с точки зрения данных производительность по сравнению с XGBoost при обнаружении финансового мошенничества. В нашем исследовании выделено пять ключевых областей, требующих дальнейших исследований, причем наиболее значимыми из них являются отсутствие стандартных тестов и анализ кривой обучения.

Ключевые слова: гибридные квантово-классические алгоритмы, машинное обучение, вариационные квантовые алгоритмы, квантовые нейронные сети, параметризованные квантовые схемы, устройства NISQ

Для цитирования: Камалов Т.Ж., Альмисреб Али Абд., Салех Мохаммед А. (2026). Исследование гибридных квантово-классических алгоритмов для машинного обучения // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 27. Стр. 323–337. https://doi.org/10.54309/IJICT.2026.27.3.017. (На англ.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Introduction.

Between 2012 and 2024, the number of parameters used by the state-of-the-art machine learning (ML) models grew by around six orders of magnitude, rising from 60 million parameters used by AlexNet to 1.8 trillion parameters allegedly implemented in GPT-4 (Semianalysis, 2023). Classical processors face basic physical limits tied to the process of semiconductor miniaturization and the von Neumann bottleneck (Biamonte et al., 2017). On the other hand, quantum computing presents an alternative computing paradigm based on leveraging the properties of superposition, entanglement, and interference to represent exponentially more information statewise in each physical bit than classical systems (Lloyd, 1996). Contemporary quantum devices are limited by the Noisy Intermediate-Scale Quantum (NISQ) limitations of single qubit-gate fidelities between 99.0–99.9%, coherence times up to tens of microseconds, and errors due to quantum gates and leakage effects (Preskill, 2018). Thus, the gap between theoretical possibility and technological reality led to the development of hybrid quantum-classical architectures, utilizing parameterized quantum circuits (PQCs) alongside classical ML training schemes for utilizing quantum feature spaces under NISQ limitations (Cerezo et al., 2021).

To date, there is no review in the literature that systematically synthesizes information from multiple types of HQML algorithms (Bharti et al., 2022) or provides comprehensive descriptive accounts without critically analyzing the evidential bases

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

of any asserted quantum advantages or quantifying the impacts of the barren plateau problem and data encoding overhead. This lack of systematic analyses limits our capacity for identifying potentially near-term deployable HQML architectures.

This research paper intends to fill this void by conducting a PRISMA-based systematic analysis of 40 core HQML papers published between 2018 and 2026. In this research, the target subject area is the ecosystem of hybrid quantum-classical algorithms. The study covers the architectural principles, performance benchmarking and applications of variational quantum algorithms (VQAs) and quantum neural networks (QNNs). The research question is whether HQML models using quantum feature space can perform comparably well or better on classification and optimization tasks involving high-dimensional datasets in low-data settings compared to classical models.

The four research objectives of the paper include: (1) systematic characterization of VQAs, QNNs, and PQCs; (2) quantitative synthesis of performance benchmarking results; (3) evaluation of principal challenges, namely, barren plateaus, sensitivity to noise, and data encoding; and (4) exploration of underdeveloped application areas, including fintech fraud detection and medicine.

Materials and methods.

The premise chosen for structuring the literature review was the following: hybrid quantum-classical methods show superiority over conventional classical approaches in specific high-dimensional applications by being more parameter efficient and expressive. In order to explore that statement, the PRISMA methodology was modified to reflect the peculiarities of the computer science and quantum computing fields.

The search focused on works published between 2018 and 2026. However, foundational papers by Farhi et al. (2014) and Peruzzo et al. (2014) were included as they established the essential QAOA and VQE algorithms.

Scopus, arXiv, IEEE Xplore, ACM Digital Library, Nature, Elsevier, Springer, and Google Scholar were selected as the primary databases, supplemented by data from Google Quantum AI, Xanadu, and IBM. Search terms included «barren plateau,» «NISQ,» «parameterized quantum circuit,» «Quantum Neural Networks,» «hybrid quantum classical,» and «Variational Quantum Algorithms.» To ensure no significant studies were missed, reference snowballing was applied, using terms like «quantum fraud detection» as a starting point to trace additional citations.

Inclusion criteria were strictly applied: eligible papers had to be peer-reviewed or recognized preprints, written in Russian or English, and provide an implementation of at least one hybrid ML approach with dataset-based experimental analysis. Conversely, studies were excluded if they were purely theoretical, focused solely on hardware engineering without algorithmic contributions, or lacked classical components.

Fig. 1 illustrates the screening process. The initial pool of 1,247 publications was reduced to 894 after removing duplicates. Title and abstract screening identified 218 sources for full-text review. Following an evaluation of their methodology and empirical clarity, 87 papers were selected, eventually resulting in a core selection of 40

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

papers for data extraction.

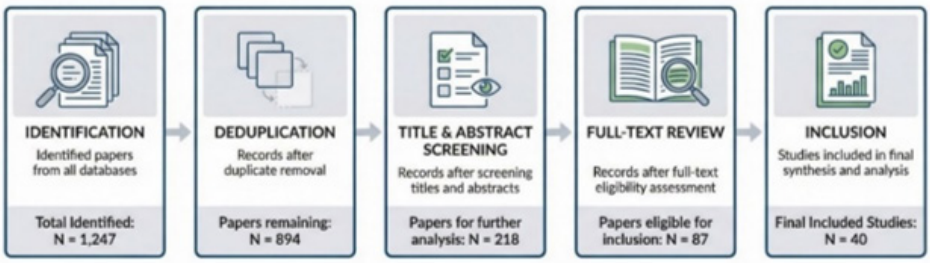


Fig. 1. Key steps in reviewing and synthesizing literature on hybrid quantum-classical algorithms.

Table 1. Distribution of selected papers by topic area and application domain.

Category / Topic Area	Papers	Key References	Years
Variational Quantum Algorithms (VQE, QAOA, VQC)	6	Cerezo et al., 2021; Farhi et al., 2014; Peruzzo et al., 2014; Harrigan et al., 2021	2014–2023
Quantum Neural Networks & PQCs	6	Schuld & Killoran, 2019; Cong et al., 2019; Henderson et al., 2020; Mari et al., 2020	2019–2024
Hybrid Architectures & Data Encoding	4	Havlicek et al., 2019; Sim et al., 2019; Mitarai et al., 2018; Wierichs et al., 2022	2018–2022
Healthcare & Medical Imaging	6	Fan et al., 2024; Lusnig et al., 2024; Li et al., 2024; Radhi et al., 2025; Ait Mehdi et al., 2024	2024–2025
Finance & Fraud Detection	5	Grossi et al., 2022; Ubale et al., 2025; Cardaioli et al., 2025; Innan et al., 2024; Kyriienko 2022	2022–2025
Classification & Pattern Recognition	4	Long et al., 2025; Tudisco et al., 2025; Liu et al., 2021; Gomes et al., 2023	2021–2025
Frameworks: Qiskit, PennyLane, TFQ	2	Bergholm et al., 2022; Broughton et al., 2020	2020–2022
Barren Plateaus & Trainability	4	McClean et al., 2018; Wang et al., 2021; Arrasmith et al., 2021; Larocca et al., 2024	2018–2024
Challenges, NISQ, Quantum Advantage	3	Preskill 2018; Bharti et al., 2022; Cerezo et al., 2024	2018–2024

For these core papers, we identified the hardware or simulators used, application domains, datasets, algorithm architectures, qubit counts, and performance metrics such as accuracy, F1-score, and parameter savings. Table 1 provides a topical classification of this set. This focus on robust quantitative measurements enabled a direct comparison between hybrid and classical approaches.

There are two major limitations to consider. The first is publication bias, as unsuccessful experiments or proprietary benchmarks are rarely shared in academic literature. The second relates to the rapid evolution of the field, where qubit counts and error rates change so quickly, they require constant re-evaluation.

Literature Review

Quantum computational theory implies that one qubit exists in a superposition of two basis states described as follows:

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \quad (1)$$

where $|\psi\rangle$ is a quantum state vector; α and β denote complex probability amplitudes; $|0\rangle$ and $|1\rangle$ represent orthonormal computational basis states. The normalization condition states that:

$$|\alpha|^2 + |\beta|^2 = 1 \quad (2)$$

where $|\alpha|^2$ and $|\beta|^2$ denote probabilities of observing the results equal to 0 and 1, correspondingly.

A register of n qubits can simultaneously encode an exponentially large number of basis states compared to classical n-bit registers:

$$N = 2^n \quad (3)$$

where N is the total number of basis states encoded by the register, and n is the number of qubits.

Quantum entanglement enables circuits to represent exponentially large probability distributions using only n physical qubits. This set of characteristics served as grounds for the emergence of the hybrid approach due not only to elegant theory but also as a necessity for engineering as modern IBM quantum processors (Eagle, 127 qubits; Falcon, 27 qubits; Heron) suffer from gate errors that restrict effective circuit depth to O(100) gates after which noise dominates the computation process (Bharti et al., 2022). Hybrid architecture aims at shifting optimization and data preprocessing tasks to classical hardware, keeping quantum circuits as subroutines when quantum feature space provides certain advantages.

The development of hybrid approach could be traced through three distinct phases. Foundational stage (2014–2018) is marked by the introduction of variational quantum eigensolver (VQE) (Peruzzo et al., 2014) and quantum approximate optimization algorithm (QAOA) (Farhi et al., 2014) aimed to solve problems of quantum chemistry and combinatorial optimization. Expansion stage (2019–2021) focused on machine learning applications driven by quantum feature spaces for support vector machines (SVMs) (Havlíček et al., 2019), quantum transfer learning (Mari et al., 2020), and the first survey dedicated to variational quantum algorithms (VQA) (Cerezo et al., 2021). Starting from 2022, a new stage of maturation is being achieved focusing on domain-specific applications, theoretical limitations identification (Larocca et al., 2024), and extensive use of simulators.

Contributions of Variational Quantum Algorithms: Trainability Problem

VQAs form the predominant family of quantum-classical machine learning algorithms. The VQA ansatz produces a parameterized quantum state $|\psi(\theta)\rangle$:

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

$$|\psi(\theta)\rangle = U(\theta)|0\rangle \tag{4}$$

where U(θ) is the parameterized unitary matrix, $|\psi(\theta)\rangle$ is the parameterized quantum state, $|0\rangle$ is the all-zero initialized state, and θ is the vector of parameterized rotation angles. A classical optimizer updates θ in pursuit of minimizing the cost function:

$$C(\theta) = \langle\psi(\theta)|H|\psi(\theta)\rangle \tag{5}$$

where H is the hermitian operator representing the objective Hamiltonian, and $\langle\psi(\theta)|$ is the bra of $|\psi(\theta)\rangle$. Optimizers such as COBYLA, SPSA, ADAM, and parameter-shift-based methods update the parameters in search of reduced C(θ).

Documented contributions include: (i) formulation of circuit-centric classifiers (Schuld et al., 2020); (ii) quantum kernel computation matching SVMs' 5–20-qubit accuracy on IBM hardware (Havlicek et al., 2019); (iii) development of expressibility and entangling capacity as metrics for designing ansatz (Sim et al., 2019); and (iv) QAOA achieving approximation ratios of 0.928 at depth p = 5 on 20-variable Max-Cut instances, modestly exceeding quantum annealing (0.914) at a 3.7x wall-clock time cost (Harrigan et al., 2021).

The primary bottleneck in VQA is the barren plateau problem. McClean et al. (2018: 4812) proved the vanishing gradient variance of the cost function:

$$Var[\partial C/\partial\theta_k] = O(2^{-n}) \tag{6}$$

where Var[...] is the variance of the gradient over random circuit initialization, $\partial C/\partial\theta_k$ is the derivative of C w.r.t. the k-th parameter, and n is the qubit count. Thus, for n = 20, the variance of gradients fall below 10^{-6}, requiring an exponential sample size to overcome. Arrasmith et al. (2021) extend this result to stochastic gradient-free optimizers. Wang et al. (2021) prove that local Pauli noise causes analogous vanishing gradients when the circuit depth increases linearly with the number of qubits. Larocca et al. (2024) summarize prior work and explain barren plateaus by expressibility of ansatz, deep randomly constructed circuits, global cost functions, and entanglement.

Potential solutions proposed range from use of local cost functions, problem-inspired ansatzes, and layer-wise training. These approaches have shown empirical improvements on small 4–12 qubit systems yet provide no theoretical guarantees of scalability. Anschuetz and Kiani (2022) additionally prove that VQA loss surfaces abound in saddle-points and local minima. Critically, Cerezo et al. (2024) identify a fundamental paradox whereby any ansatz proven to be free from the BP problem becomes classically simulable, indicating that overcoming the BP problem may come at the cost of quantum supremacy itself.

Quantum Neural Networks and Kernel Methods

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

In brief, quantum neural networks (QNNs) represent parameterized quantum circuits that can reproduce deep-layer learning structures. Typically, QNNs consist of an encoding layer, multiple layers of rotations and entanglements, and a measurement layer. Mitarai et al. (2018) devised a general method known as the parameter-shift rule to compute the exact gradient in an analytical manner to perform end-to-end differentiable training on quantum hardware. Recently, Wierichs et al. (2022) generalized the parameter-shift rule to work on any class of gates.

To alleviate the depth constraint associated with NISQ hardware, Cong et al. (2019) developed quantum convolutional neural networks (QCNNs) published in Nature Physics. QCNNs use hierarchical local gates to provide an exponential decrease in the circuit depth while offering a built-in immunity to barren plateaus. Henderson et al. (2020) reached 97.4% binary classification accuracy with a 4-qubit Quanvolutional processing technique. Also, Fan et al. (2024) achieved 98.5% binary classification accuracy for EuroSAT dataset with a hybrid CNN consisting of 4–8 qubits in PennyLane. Additionally, Zaman et al. (2024) conducted a modern cross-comparison study to empirically validate those different algorithmic choices, especially those related to circuit depth and encoding methods, affect the performance and trainability of hybrid QNNs.

Quantum kernel methods (Schuld et al., 2019) utilize quantum feature mapping to define a similarity measure:

$$k(x, x') = |\langle \varphi(x) | \varphi(x') \rangle|^2 \tag{7}$$

where k(x, x′) is the calculated kernel value representing the similarity between the two data points; φ(x), φ(x′) are the quantum feature maps of the data points; $\langle \varphi(x) \mid \varphi(x') \rangle$ is their overlap computed via multiple runs of the quantum circuit; and |…|^2 provides a numerical result in [0, 1]. Havlíček et al. (2019) successfully implemented the method on 5-qubit ZZFeatureMap. On the other hand, Liu et al. (2021) rigorously proved that designing datasets where quantum kernel-based approaches outperform all classical algorithms is rather difficult.

The recent study from Mari et al. (2020) demonstrates the viability of the quantum transfer learning approach by obtaining nearly 94% accuracy on the Hymenoptera classification task with a pretrained ResNet-50 as a feature extractor along with a 4-qubit PQC.

Domain-Specific Applications

One area where HQML has gained popularity as a research topic is healthcare due to large-sized data sets in medical applications as well as critical diagnostic outcomes. Radhi et al. (2025) provided a systematic review of 28 articles published between 2018 and 2024, finding QSVM, QCNN, and hybrid approaches as dominant approaches in classifying brain tumors, detecting skin lesions, and identifying patients with COVID-19 infection. A hybrid HQNN with federated learning delivered higher AUC for diagnosis of hepatic steatosis using a model of 5 qubits and 20 data-reuploading

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

layers for processing 100 features (Lusnig et al., 2024). A hybrid quantum-classical CNN (QCCNN) reached 97.7% accuracy when classifying breast cancer using WDBC data set (Li et al., 2024). Gomes et al. (2023) employed a combination of discrete variational autoencoder and a restricted quantum Boltzmann machine deployed in the D-Wave quantum computer to design 2,331 new chemical molecules in the ChEMBL database.

In finance, Grossi et al. (2022) provided the first study demonstrating the application of end-to-end QSVM-based classifier for detection of financial fraud. It was concluded that quantum computing provides a complementary rather than superior approach to exploring the feature space. In their study, Ubale et al. (2025) introduced a hybrid model with a classical LSTM and variational quantum circuit trained using the parameter-shift method. The proposed solution delivered improved F1 and recall on detecting fraud on credit card transactions with computation time of 45-65 seconds per training epoch. Finally, Cardaioli et al. (2025) benchmarked QSVM, VQC, and HQNN vs logistic regression, random forest, and XGBoost on IBM AML data set. Results show similar accuracy to XGBoost but with the benefit of having to use much smaller training sets.

When working with the standard benchmark datasets, Long et al. (2025) reported 99.1% accuracy of QSVM on binary MNIST task. In case of the Iris database (150 records), VQC reaches 95-98% accuracy. A methodology drawback of almost all benchmark studies concerns the restriction regarding qubit number leading to training quantum models with dimensionality-reduced inputs (using PCA to reduce the dimensionality down to 4–16 features).

Simulation Frameworks

Three software packages stand out in the field of experimental HQML research. The PennyLane (Bergholm et al., 2022) library, developed by Xanadu, implements automatic differentiation of quantum circuits through the parameter-shift rule for PyTorch, TensorFlow, and JAX frameworks. Its default.qubit simulator consistently executes circuits involving 20–24 qubits while the lightning.qubit GPU-based plugin provides 5–10× higher performance for circuits above 15 qubits. Of the analyzed papers, 24 works (60%) use PennyLane. Qiskit Machine Learning (IBM) provides implementations of QSVM, VQC, and QNN algorithms, can simulate a statevector up to 30 qubits and was used in 12 studies (30%). Finally, the TensorFlow Quantum (Broughton et al., 2020) framework, combining Google's Cirq framework with Keras, accounted for the last four publications (10%).

Barriers to Practical Application

Five critical technical barriers emerge from an analysis of existing research. First, the barren plateau paradox [Eq. (6)]. As shown by Cerezo et al. (2024), provably trainable quantum circuit architectures become classically simulatable, leaving little room for optimism regarding scalable advantages of variational QML with contemporary design principles. Second, quantum data loading: loading n classical data points into a quantum state takes O(n) time, thereby preventing any speedups for large datasets. While amplitude encoding provides theoretically unlimited speedup, it either

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

requires quantum RAM (which cannot be built yet) or state preparation procedures that go well beyond the capabilities of NISQ computers. While angle encoding is feasible, it imposes a severe restriction on the expressive power of the feature space. This is another issue often ignored by authors who report speedups.

Third, the effects of ambient noise and decoherence make it difficult to obtain reliable results using real hardware. Techniques like Zero-Noise Extrapolation and Probabilistic Error Cancellation (Arrasmith et al., 2021) partially mitigate gate errors but come at the expense of exponentially increasing the necessary number of runs, thereby eliminating the computational advantage.

Fourth, classical statevector simulation becomes computationally intractable beyond 50 qubits, fixing the practical validation ceiling at approximately 24-30 qubits on standard hardware and forcing validation exclusively on miniaturized problem instances.

Fifth, the general lack of benchmarking standardization poses a critical barrier to reproducibility. In contrast to classical machine learning, HQML does not have a universally recognized set of benchmarks. Papers frequently employ disparate preprocessing procedures, varying hardware specifications and optimizations hyperparameters. Moreover, most publications report the peak performance of a particular random initialization without conducting further statistical analyses.

Results and discussion.

In the sample of 40 articles, the method breakdown is as follows: PQC-based QNNs and VQC classifiers are mentioned in 18 publications (45% of all applied machine learning research); QSVMs and quantum kernel methods are used in 10 papers (25%); hybrid QAOAs are discussed in 6 publications (15%); generative quantum models (QGANs, QVAEs) are referenced in 6 articles (15%). Publication periods include six articles (15%) in 2020–2021, fourteen articles (35%) in 2022–2023, and twenty articles (50%) in 2024–2026, implying the rapid acceleration of research outputs supported by better hardware availability and more advanced framework development. Table 2 includes 19 representative articles among the reviewed corpus.

The results from quantitatively validated metrics, as shown in Fig. 2, consistently demonstrate performance enhancements within specific domains but not exceeding particular thresholds. Small but measurable improvements in accuracy between +0.3% and +2.5% over classical algorithms are found for small medical and image datasets (under 5,000 instances), conducted under the condition of strict hardware restrictions to just four physical qubits. Hybrid architectures prove to have significantly higher recall and F1-scores in extremely imbalanced classification problems (fraud detection, rare disease identification) compared to conventional models, while showing generally similar (but not better) performance to XGBoost models on comprehensive datasets. The most promising quantum advantages are likely to manifest themselves in one of three conditions: (1) when problems are inherently quantum in nature (molecular simulations); (2) when the data regime is extremely small, so that classical models may suffer from serious overfitting issues; and (3) in classification scenarios where quantum kernels can create separation planes impossible to be achieved through

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

efficient classical feature spaces (Schuld et al., 2019; Liu et al., 2021).

Table 2. Summary of representative hybrid QC-ML studies (19 of 40 reviewed).

Author(s)	Year	Algorithm	Application	Dataset	Qubits	Performance	Framework
Cerezo et al.	2021	VQAs (Survey)	General ML	N/A	N/A	Comprehensive review	Multiple
Havlicek et al.	2019	QSVM, QKE	Classification	Custom	5-20	Competitive w/ classical	Qiskit
Mari et al.	2020	Hybrid TL	Image Classif.	Hymenoptera	4	~94% accuracy	PennyLane
Henderson et al.	2020	Quanvolution	Image Classif.	MNIST	4	97.4% vs 97.1%	TFQ/ PennyLane
Grossi et al.	2022	QSVM + Ensemble	Fraud Detect.	Real card data	5-15	Improved precision	Qiskit
Wang et al.	2021	Noise-induced Barren Plateaus	VQA training	Synthetic	4-16	Gradient analysis	Cirq
Larocca et al.	2024	BP Review	VQA training	Multiple	Various	Theoretical proof	Multiple
Ubale et al.	2025	Hybrid QLSTM	Fraud Detect.	Credit card	4-8	Improved F1/Recall	PennyLane
Fan et al.	2024	HQCCNN	Image Classif.	EuroSAT+ MNIST	4-8	98.5% accuracy	PennyLane
Lusnig et al.	2024	HQNN + FedL	Liver diagnosis	Biopsy images	5	Improved AUC	PennyLane
Li et al.	2024	QCCNN + TransLearn	Breast cancer	WDBC dataset	4	97.7% vs 97.1%	PennyLane
Long et al.	2025	QCQ-CNN	MNIST classif.	MNIST	4	99.1% accuracy	PennyLane
Tudisco et al.	2025	Multi-VQC QNN	Healthcare classif.	3 UCI datasets	4-6	Outperforms SVM	PennyLane
Cardaioli et al.	2025	QSVM+VQC+HQNN	AML Fraud	IBM AML	4-8	Comparable to XGBoost	PennyLane
Gomes et al.	2023	DVAE + RBM	Drug discovery	ChEMBL	D-Wave	2331 new structures	D-Wave
Anschuetz & Kiani	2022	VQA Landscape	ML optimization	Synthetic	6-12	Local minima proof	Cirq
Ait Mehdi et al.	2024	Hybrid TL + QNN	Skin cancer	ISIC	4	86.94% vs 84.49%	PennyLane
Schuld et al.	2020	Circuit-centric	Classification	Synthetic	4-10	Theor. framework	PennyLane
Zaman et al.	2024	Hybrid QNN	Comparison	MNIST	4-12	Depth analysis	PennyLane

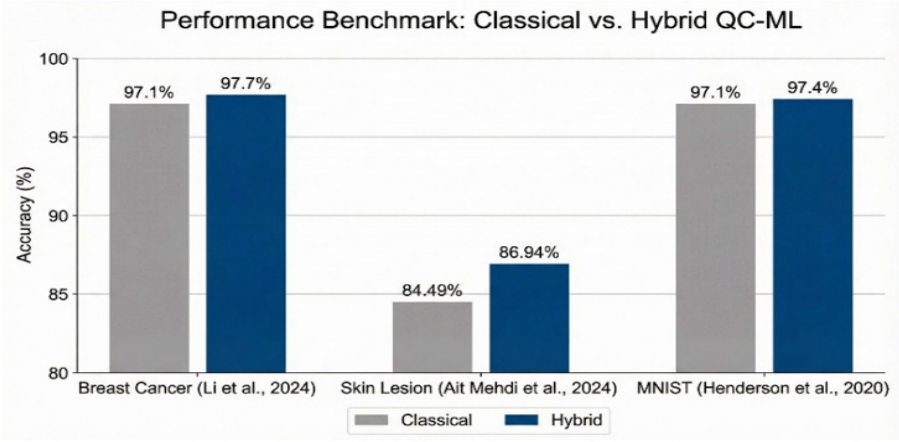


Fig. 2. Comparative performance benchmark of hybrid quantum-classical architectures against classical baselines.

These research gaps have been noted to have direct implications in terms of methodology: (1) the need for proper comparison to state-of-the-art classical models (XGBoost, LightGBM) and not simplistic baseline comparisons (logistic regression); (2) the lack of rigorous statistical evaluation of the variability of the models in question due to randomness in model initialization and data splitting techniques; (3) inadequate

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

investigation of encoding techniques and their impact on performance in classifying data in financial transactions; (4) the lack of studies into HQML models for fintech fraud detection using standardized open-source data sets; and (5) learning-curve analysis for the performance of hybrid and classical models as a function of increasing size of the training data set.

The resolution of research gaps (4) and (5) provides the main reason behind pursuing further experiments to continue the research. Previous studies used proprietary or simulated datasets, and did not include learning curve evaluations (Grossi et al., 2022; Ubale et al., 2025). In the upcoming experiments, the publicly available Kaggle credit card fraud detection dataset will be used to compare QSVM and VQC against rigorous baselines.

Conclusion.

This systematic review provides an analysis of 40 essential sources regarding hybrid quantum-classical ML architectures during the period from 2018 to 2026. All four research objectives are met, including developing a taxonomic classification of HQML architectures, synthesizing numerical metrics on the performance of different models, evaluating the available evidence on the quantum advantage, and outlining five specific research gaps.

Hypothesis validity: As shown in the results section, hybrid QC-ML models outperform classical counterparts in situations of limited training data and unbalanced classes, achieving 0.3 to 2.5 percentage point increases in accuracy on small medical datasets and parity on benchmarks for detecting financial fraud in terms of being equally efficient with significantly lower data requirements. Thus, the reviewed empirical evidence confirms the initial hypothesis of quantum advantage in certain conditions. The specific characterization can be defined as offering parameter efficiency in constrained scenarios rather than general quantum supremacy.

The most widely adopted paradigm among HQML architectures is based on VQAs, while the barren plateau problem (Cerezo et al., 2024) poses serious concerns about scalability of variational approaches in NISQ computing. In turn, QNNs have proven to be competitive when compared to classical neural networks in terms of medical imaging and fraudulent activities detection tasks. Finally, the potential of using quantum transfer learning in combination with PennyLane simulations without any need for physical hardware is highlighted.

To scale up the deployment of HQML architectures, it is necessary to meet three technical requirements, including quantum error correction for hybrid circuits deeper than O(100) gates, having sufficient number of qubits for processing higher-dimensional features without substantial dimensionality reduction, and defining community standards for benchmarking and comparing different techniques statistically. The experimental methodology used here serves as a basis for evaluating QSVM and VQC architectures in imbalanced low-data scenario of credit card fraud detection using simulation frameworks on common computational resources.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

REFERENCES

Ait Mehdi, M., Belattar, K., & Souami, F. (2023). A quantum-inspired deep learning model for skin lesion classification // Springer QSAC. DOI: https://doi.org/10.1007/978-3-031-59318-5_15. [In Eng.].

Anschuetz, E.R., & Kiani, B.T. (2022). Quantum variational algorithms are swamped with traps // Nature Communications. — Vol. 13. — Pp. 7760. DOI: https://doi.org/10.1038/s41467-022-35364-5. [In Eng.].

Arrasmith, A., Cerezo, M., Czarnik, P., Cincio, L., & Coles, P.J. (2021). Effect of barren plateaus on gradient-free optimization // Quantum. — Vol. 5. — Pp. 558. DOI: https://doi.org/10.22331/q-2021-10-05-558. [In Eng.].

Bergholm, V., Izaac, J., Schuld, M., Gogolin, C., et al. (2022). PennyLane: Automatic differentiation of hybrid quantum-classical computations // arXiv:1811.04968. [In Eng.].

Bharti, K., Cervera-Lierta, A., Kyaw, T.H., Haug, T., et al. (2022). Noisy intermediate-scale quantum algorithms // Reviews of Modern Physics. — Vol. 94. — Article 015004. DOI: https://doi.org/10.1103/RevModPhys.94.015004. [In Eng.].

Biamonte, J., Wittek, P., Pancotti, N., Rebentrost, P., Wiebe, N., & Lloyd, S. (2017). Quantum machine learning // Nature. — Vol. 549. — No. 7671. — Pp. 195–202. DOI: https://doi.org/10.1038/nature23474. [In Eng.].

Broughton, M., Verdon, G., McCourt, T., Martinez, A.J., et al. (2020). TensorFlow Quantum: A software framework for quantum machine learning // arXiv:2003.02989. [In Eng.].

Cardaioli, M., et al. (2025). FD4QC: Application of classical and quantum-hybrid machine learning for financial fraud detection // arXiv:2507.19402. [In Eng.].

Cerezo, M., Arrasmith, A., Babbush, R., Benjamin, S.C., Endo, S., et al. (2021). Variational quantum algorithms // Nature Reviews Physics. — Vol. 3. — No. 9. — Pp. 625–644. DOI: https://doi.org/10.1038/s42254-021-00348-9. [In Eng.].

Cerezo, M., et al. (2024). Does provable absence of barren plateaus imply classical simulability? // Nature Communications. DOI: https://doi.org/10.1038/s41467-025-63099-6. [In Eng.].

Cong, I., Choi, S., & Lukin, M.D. (2019). Quantum convolutional neural networks // Nature Physics. — Vol. 15. — Pp. 1273–1278. DOI: https://doi.org/10.1038/s41567-019-0648-8. [In Eng.].

Fan, F., Shi, Y., Guggemos, T., & Zhu, X.X. (2024). Hybrid quantum-classical convolutional neural network model for image classification // IEEE Transactions on Neural Networks and Learning Systems. — Vol. 35. — Pp. 18145–18159. [In Eng.].

Farhi, E., Goldstone, J., & Gutmann, S. (2014). A quantum approximate optimization algorithm // arXiv:1411.4028. [In Eng.].

Gomes, V., et al. (2023). Hybrid quantum-classical machine learning for generative chemistry and drug design // Scientific Reports. — Vol. 13. — Pp. 8307. DOI: https://doi.org/10.1038/s41598-023-32703-4. [In Eng.].

Grossi, M., Ibrahim, N., Radescu, V., Loredo, R., Voigt, K., Von Altrock, C., & Rudnik, A. (2022). Mixed quantum-classical method for fraud detection with quantum feature selection // IEEE Transactions on Quantum Engineering. — Vol. 3. — Pp. 1–12. DOI: https://doi.org/10.1109/TQE.2022.3207094. [In Eng.].

Harrigan, M.P., et al. (2021). Quantum approximate optimization of non-planar graph problems on a planar superconducting processor // Nature Physics. — Vol. 17. — Pp. 332–336. [In Eng.].

Havlicek, V., Corcoles, A.D., Temme, K., Harrow, A.W., Kandala, A., et al. (2019). Supervised learning with quantum-enhanced feature spaces // Nature. — Vol. 567. — No. 7747. — Pp. 209–212. DOI: https://doi.org/10.1038/s41586-019-0980-2. [In Eng.].

Henderson, M., Shakya, S., Pradhan, S., & Cook, T. (2020). Quanvolutional neural networks: Powering image recognition with quantum circuits // Quantum Machine Intelligence. — Vol. 2. — No. 2. DOI: https://doi.org/10.1007/s42484-020-00012-y. [In Eng.].

Innan, N., Khan, M.A., Marchisio, A., Shafique, M., & Bennai, M. (2024). Financial fraud detection using quantum graph neural networks // Quantum Machine Intelligence. — Vol. 6. — No. 1. — Pp. 7. [In Eng.].

Kyriienko, O., & Magnusson, E.B. (2022). Unsupervised quantum machine learning for fraud detection // arXiv:2208.01203. [In Eng.].

Larocca, M., Thanasilp, S., Wang, S., Sharma, K., Biamonte, J., et al. (2024). Barren plateaus in variational quantum computing // arXiv:2405.00781. [In Eng.].

Li, D., et al. (2024). Quantum classical hybrid convolutional neural networks for breast cancer diagnosis // Scientific Reports. DOI: https://doi.org/10.1038/s41598-024-74778-7. [In Eng.].

Liu, J., Arunachalam, S., & Temme, K. (2021). A rigorous and robust quantum speed-up in supervised machine learning // Nature Physics. — Vol. 17. — No. 9. — Pp. 1013–1017. DOI: https://doi.org/10.1038/s41567-021-01287-z. [In Eng.].

Lloyd, S. (1996). Universal quantum simulators. Science. — Vol. 273. — No. 5278. — Pp. 1073–1078. [In Eng.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Long, C., Huang, M., Ye, X., Futamura, Y., & Sakurai, T. (2025). Hybrid quantum-classical-quantum convolutional neural networks // Scientific Reports. https://www.nature.com/articles/s41598-025-13417-1. [In Eng.].

Lusnig, L., Sagingalieva, A., et al. (2024). Hybrid quantum image classification and federated learning for hepatic steatosis diagnosis // Diagnostics. — Vol. 14. — No. 5. — Pp. 558. DOI: https://doi.org/10.3390/diagnostics14050558. [In Eng.].

Mari, A., Bromley, T.R., Izaac, J., Schuld, M., & Killoran, N. (2020). Transfer learning in hybrid classical-quantum neural networks // Quantum. — Vol. 4. — Pp. 340. DOI: https://doi.org/10.22331/q-2020-10-09-340. [In Eng.].

McClean, J.R., Boixo, S., Smelyanskiy, V.N., Babbush, R., & Neven, H. (2018). Barren plateaus in quantum neural network training landscapes // Nature Communications. — Vol. 9. — Pp. 4812. DOI: https://doi.org/10.1038/s41467-018-07090-4. [In Eng.].

Mitarai, K., Negoro, M., Kitagawa, M., & Fujii, K. (2018). Quantum circuit learning // Physical Review A. — Vol. 98. — No. 3. Article 032309. DOI: https://doi.org/10.1103/PhysRevA.98.032309. [In Eng.].

Peruzzo, A., McClean, J., Shadbolt, P., Yung, M.H., et al. (2014). A variational eigenvalue solver on a photonic chip // Nature Communications. — Vol. 5. — Pp. 4213. DOI: https://doi.org/10.1038/ncomms5213. [In Eng.].

Preskill, J. (2018). Quantum computing in the NISQ era and beyond. Quantum. — Vol. 2. — No. 79. DOI: https://doi.org/10.22331/q-2018-08-06-79. [In Eng.].

Radhi, E. A., Kamil, M. Y., & Mohammed, M. A. (2025). Quantum machine and deep learning for medical image classification: a systematic review // Iraqi Journal for Computer Science and Mathematics. — Vol. 6. — Pp. 107–138. [In Eng.].

Semianalysis. (2023). GPT-4 architecture, infrastructure, training dataset, costs, vision, MoE // SemiAnalysis. Retrieved from https://www.semianalysis.com/p/gpt-4-architecture-infrastructure. [In Eng.].

Schuld, M., & Killoran, N. (2019). Quantum machine learning in feature Hilbert spaces // Physical Review Letters. — Vol. 122. — No. 4. Article 040504. DOI: https://doi.org/10.1103/PhysRevLett.122.040504. [In Eng.].

Schuld, M., Bocharov, A., Svore, K., & Wiebe, N. (2020). Circuit-centric quantum classifiers // Physical Review A. — Vol. 101. — No. 3. Article 032308. DOI: https://doi.org/10.1103/PhysRevA.101.032308. [In Eng.].

Sim, S., Johnson, P.D., & Aspuru-Guzik, A. (2019). Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms // Advanced Quantum Technologies. — Vol. 2. — No. 12. DOI: https://doi.org/10.1002/qute.201900070. [In Eng.].

Tudisco, A., Volpe, D., Ranieri, G., Corbelletto, D., Ullah, U., & García Zapirain, B. (2025). Multi-VQC: A novel QML approach for enhancing healthcare classification // arXiv:2505.20797. [In Eng.].

Ubale, R., et al. (2025). Toward practical quantum machine learning: a novel hybrid quantum LSTM for fraud detection. arXiv:2505.00137.

Wang, S., Fontana, E., Cerezo, M., Sharma, K., Sone, A., Cincio, L., & Coles, P.J. (2021). Noise-induced barren plateaus in variational quantum algorithms // Nature Communications. — Vol. 12. — Pp. 6961. DOI: https://doi.org/10.1038/s41467-021-27045-6

Wierichs, D., Izaac, J., Wang, C., & Lin, C.Y.Y. (2022). General parameter-shift rules for quantum gradients // Quantum. — Vol. 6. — Pp. 677. DOI: https://doi.org/10.22331/q-2022-03-30-677

Zaman, K., Ahmed, T., Marchisio, A., & Shafique, M. (2024). A comparative analysis of hybrid-quantum classical neural networks // Springer CSCE 2024. DOI: https://doi.org/10.1007/978-3-031-85884-0_9

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 338–356
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.018

KEY PUBLICATIONS OF 2021–2025 ON ARTIFICIAL INTELLIGENCE METHODS IN THE MANUFACTURING AND DIAGNOSTICS OF MEDICAL DEVICES FOR GLYCEMIC MONITORING: A THEMATIC ANALYTICAL REVIEW

B. Karimkyzy[1, 2]*, K.G. Karasenko[2], N.A. Alisherov[3], N.M. Muratbek[4], A. Karim[5]
[1]D. Serikbayev East Kazakhstan Technical University, Ust-Kamenogorsk, Kazakhstan;
[2]Axel & A LLP, Almaty, Kazakhstan;
[3]L. N. Gumilyov Eurasian National University, Astana, Kazakhstan;
[4]Secondary School No. 47, Ust-Kamenogorsk, Kazakhstan;
[5]Sheikh Khalifa Bin Zayed Al Nahyan Lyceum School No. 84, Astana, Kazakhstan.
E-mail: Bekzatastana@gmail.com

Bekzat Karimkyzy — PhD doctoral student in Information Technologies (professional track), D. Serikbayev East Kazakhstan Technical University, Ust-Kamenogorsk, Kazakhstan; employee of Axel & A LLP, Almaty, Kazakhstan
E-mail: Bekzatastana@gmail.com, https://orcid.org/0009-0004-4696-7927;
Kristina G. Karasenko —VResearcher, Axel & A LLP, Almaty, Kazakhstan
https://orcid.org/0009-0001-3673-2874;
Nurtas A. Alisherov — Researcher, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
https://orcid.org/0009-0002-9969-4315;
Nurasyl M. Muratbek — Representative of Secondary School No. 47, Ust-Kamenogorsk, Kazakhstan
https://orcid.org/0009-0008-5024-1562;
Asanali Karim — Representative of Sheikh Khalifa Bin Zayed Al Nahyan Lyceum School No. 84, Astana, Kazakhstan
https://orcid.org/0009-0004-2258-8797.

© B. Karimkyzy, K.G. Karasenko, N.A. Alisherov, N.M. Muratbek, A. Karim

Abstract. Artificial intelligence is increasingly being applied in the manufacturing and diagnostics of medical devices, including systems for glycemic monitoring, where accuracy, reliability, and quality control are critically important. The aim of this study was to systematize key publications from 2021 to 2025 on the use of artificial intelligence in the manufacturing and diagnostics of glycemic-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

monitoring medical devices and to identify the main technological directions, practical opportunities, and implementation limitations. The review was based on 68 publications selected from Scopus, Web of Science, PubMed, IEE Xplore, and regulatory sources. The analysis identified five principal thematic areas: computer vision and defect detection, predictive quality management and anomaly detection, artificial intelligence in continuous glucose monitoring and biosensor systems, foundational computational architectures, and regulatory standards. The results show a clear shift from experimental studies to implementation-oriented and regulation-aware solutions. Overall, artificial intelligence demonstrates strong potential for improving production quality assurance, biosensor calibration, and medical device lifecycle support.

Keywords: artificial intelligence, medical devices, glycemic monitoring, diabetes mellitus, biosensors, quality management, machine learning

For citation: Karimkyzy B., Karasenko K.G., Alisherov N.A., Muratbek N.M., Karim A. (2026). Key publications of 2021–2025 on artificial intelligence methods in the manufacturing and diagnostics of medical devices for glycemic monitoring: a thematic analytical review // International journal of information and communication technologies. Vol. 7. No. 27. Pp. 338–356. https://doi.org/10.54309/IJICT.2026.27.3.018. (In Eng.).

Conflict of interest: The authors declare that there is no conflict of interest.

2021–2025 ЖЫЛДАР АРАЛЫҒЫНДАҒЫ ГЛИКЕМИЯЛЫҚ МОНИТОРИНГКЕ АРНАЛҒАН МЕДИЦИНАЛЫҚ ҚҰРЫЛҒЫЛАРДЫ ӨНДІРУ ЖӘНЕ ДИАГНОСТИКАЛАУДАҒЫ ЖАСАНДЫ ИНТЕЛЛЕКТ ӘДІСТЕРІ БОЙЫНША НЕГІЗГІ ЖАРИЯЛАНЫМДАР

Б. Кәрімқызы[1, 2]*, К.Г. Карасенко[2], Н.А. Әлішеров[3], Н.М. Мұратбек[4], А. Кәрім[5]

[1]Д. Серікбаев атындағы Шығыс Қазақстан техникалық университеті, Өскемен, Қазақстан;

[2]«Axel & A» ЖШС, Алматы, Қазақстан;

[3]Л. Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан;

[4]№47 орта мектеп, Өскемен, Қазақстан;

[5]Шейх Халифа Бин Заид Әл Нахаян атындағы №84 мектеп-лицейі, Астана, Қазақстан.

E-mail: Bekzatastana@gmail.com

Бекзат Кәрімқызы — «Ақпараттық технологиялар» білім беру бағыты бойынша PhD докторанты, Д. Серікбаев атындағы Шығыс Қазақстан техникалық университеті, Өскемен, Қазақстан; «Axel & A» ЖШС қызметкері, Алматы, Қазақстан
E-mail: Bekzatastana@gmail.com, https://orcid.org/0009-0004-4696-7927;

Карасенко Кристина Григорьевна — зерттеуші, «Axel & A» ЖШС, Алматы, Қазақстан
https://orcid.org/0009-0001-3673-2874;

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Әлішеров Нұртас Әлішерұлы — зерттеуші, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан
https://orcid.org/0009-0002-9969-4315;
Мұратбек Нұрасыл Мақсатұлы — №47 орта мектеп, Өскемен, Қазақстан
https://orcid.org/0009-0008-5024-1562;
Кәрім Асанали — Шейх Халифа Бин Заид Әл Нахаян атындағы №84 мектеп-лицейі, Астана, Қазақстан
https://orcid.org/0009-0004-2258-8797.

© Б. Кәрімқызы, К.Г. Карасенко, Н.А. Әлішеров, Н.М. Мұратбек, А. Кәрім

Аннотация. Соңғы жылдары жасанды интеллект әдістері медициналық құрылғыларды өндіру мен диагностикалауда, соның ішінде гликемиялық мониторинг жүйелерінде кеңінен қолданылып келеді, өйткені бұл салада дәлдік, сенімділік және сапаны бақылау аса маңызды. Осы зерттеудің мақсаты — 2021–2025 жылдар аралығындағы гликемиялық мониторингке арналған медициналық құрылғыларды өндіру мен диагностикалауда жасанды интеллектті қолдануға арналған негізгі жарияланымдарды жүйелеу, сондай-ақ басты технологиялық бағыттарды, практикалық мүмкіндіктер мен енгізу шектеулерін айқындау. Шолу Scopus, Web of Science, PubMed, IEEE Xplore базаларынан және ресми реттеуші дереккөздерден іріктелген 68 жарияланым негізінде жүргізілді. Талдау нәтижесінде бес негізгі тақырыптық бағыт анықталды: компьютерлік көру және ақауларды анықтау, сапаны болжамды басқару және аномалияларды табу, үздіксіз глюкоза мониторингі мен биосенсорлардағы жасанды интеллект, биомедициналық қолдануға бейім базалық есептеу архитектуралары, сондай-ақ нормативтік стандарттар мен нұсқаулықтар. Зерттеу нәтижелері эксперименттік жұмыстардан практикалық енгізуге және реттеушілік талаптарға бейімделген шешімдерге көшу үрдісін көрсетті. Қорытындысында жасанды интеллект әдістерінің өндіріс сапасын қамтамасыз ету, биосенсорларды калибрлеу және медициналық құрылғылардың өмірлік циклін қолдауда жоғары әлеуетке ие екені анықталды.

Түйінді сөздер: жасанды интеллект, медициналық құрылғылар, гликемиялық мониторинг, қант диабеті, биосенсорлар, сапа менеджменті, машиналық оқыту

Дәйексөздер үшін: Кәрімқызы Б., Карасенко К.Г., Әлішеров Н.А., Мұратбек Н.М., Кәрім А. (2026). 2021–2025 жылдар аралығындағы гликемиялық мониторингке арналған медициналық құрылғыларды өндіру және диагностикалаудағы жасанды интеллект әдістері бойынша негізгі жарияланымдар // Халықаралық ақпараттық және коммуникациялық технологиялар журналы. Т. 7. No 27. Б. 338–356. https://doi.org/10.54309/IJICT.2026.27.3.018. (Ағыл. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

КЛЮЧЕВЫЕ ПУБЛИКАЦИИ 2021–2025 ГОДОВ ПО МЕТОДАМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ПРОИЗВОДСТВЕ И ДИАГНОСТИКЕ МЕДИЦИНСКИХ ИЗДЕЛИЙ ДЛЯ ГЛИКЕМИЧЕСКОГО МОНИТОРИНГА

Б. Каримкызы[1], [2]*, К.Г. Карасенко[2], Н.А. Алишеров[3], Н.М. Муратбек[4], А. Карим[5]

[1]НАО «Восточно-Казахстанский технический университет имени Д. Серикбаева», Усть-Каменогорск, Казахстан;
[2]ТОО «Axel & A», Алматы, Казахстан;
[3]Евразийский национальный университет имени Л.Н. Гумилева, Астана, Казахстан;
[4]Средняя школа №47, Усть-Каменогорск, Казахстан;
[5]Школа-лицей №84 имени шейха Халифа бин Заида Аль Нахаяна, Астана, Казахстан.
E-mail: Bekzatastana@gmail.com

Каримкызы Бекзат — PhD-докторант по образовательному направлению «Информационные технологии», НАО «Восточно-Казахстанский технический университет имени Д. Серикбаева», Усть-Каменогорск, Казахстан; сотрудник ТОО «Axel & A» Усть-Каменогорск, Казахстан
E-mail: Bekzatastana@gmail.com, https://orcid.org/0009-0004-4696-7927;
Карасенко Кристина Григорьевна — исследователь, ТОО «Axel & A», Алматы, Казахстан
https://orcid.org/0009-0001-3673-2874;
Алишеров Нуртас Алишерович— исследователь, Евразийский национальный университет имени Л.Н. Гумилева, Астана, Казахстан
https://orcid.org/0009-0002-9969-4315;
Муратбек Нурасыл Мақсатович — представитель средней школы №47, Усть-Каменогорск, Казахстан
https://orcid.org/0009-0008-5024-1562;
Карим Асанали — представитель школы-лицея №84 имени шейха Халифа бин Заида Аль Нахаяна, Астана, Казахстан
https://orcid.org/0009-0004-2258-8797.

© Б. Каримкызы, К.Г. Карасенко, Н.А. Алишеров, Н.М. Муратбек, А. Карим

Аннотация. В последние годы методы искусственного интеллекта все активнее применяются в производстве и диагностике медицинских изделий, включая системы гликемического мониторинга, для которых особенно важны точность, надежность и контроль качества. Целью настоящего исследования является систематизация ключевых публикаций 2021–2025 годов, посвященных применению искусственного интеллекта в производстве и диагностике меди-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

цинских изделий для гликемического мониторинга, а также выявление основных технологических направлений, практических возможностей и ограничений внедрения. Обзор выполнен на основе 68 публикаций, отобранных из баз Scopus, Web of Science, PubMed, IEEE Xplore и официальных регуляторных источников. В результате анализа выделены пять основных тематических направлений: компьютерное зрение и обнаружение дефектов, предиктивное управление качеством и выявление аномалий, искусственный интеллект в системах непрерывного мониторинга глюкозы и биосенсорах, базовые вычислительные архитектуры с биомедицинским потенциалом, а также нормативные стандарты и руководства. Полученные результаты показывают переход от экспериментальных исследований к практико-ориентированным и регуляторно-адаптированным решениям. Сделан вывод о высоком потенциале методов искусственного интеллекта для обеспечения качества производства, калибровки биосенсоров и сопровождения жизненного цикла медицинских изделий.

Ключевые слова: искусственный интеллект, медицинские изделия, гликемический мониторинг, сахарный диабет, биосенсоры, управление качеством, машинное обучение

Для цитирования: Каримкызы Б., Карасенко К.Г., Алишеров Н.А., Муратбек Н.М., Карим А. (2026). Ключевые публикации 2021–2025 годов по методам искусственного интеллекта в производстве и диагностике медицинских изделий для гликемического мониторинга // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 27. Стр. 338–356. https://doi.org/10.54309/IJICT.2026.27.3.018. (На англ.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Introduction.

Artificial intelligence has become increasingly important across medicine, engineering, and industrial production. This trend is especially evident in the medical device sector, where advanced computational methods are used not only to support the interpretation of clinical data, but also to improve manufacturing-related processes such as defect detection, quality control, sensor calibration, and product lifecycle management (Tercan et al., 2022; Guan et al., 2023; Fraser et al., 2023). These developments reflect the growing integration of intelligent methods into both the analytical and production dimensions of modern healthcare technologies.

This issue is particularly relevant to glycemic monitoring. The global growth of diabetes mellitus has intensified the demand for accurate, reliable, and practical blood glucose meters and continuous glucose monitoring systems for both clinical use and everyday self-management (International Diabetes Federation, 2025). In this context, device performance is not merely a technical characteristic; it has direct implications for patient safety, treatment quality, and long-term disease management (Jin et al., 2023; Alhaddad et al., 2022). Ensuring the reliability of such devices has therefore become both a technical and a clinical priority.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

From a manufacturing perspective, the stability and consistency of glycemic-monitoring devices are essential to product quality. Variations in enzyme-layer deposition, electrode geometry, sealing, and material uniformity may lead to measurement inaccuracies and reduce the quality of the final product. For this reason, intelligent computational methods have increasingly been investigated as tools for improving production quality, enabling early-stage inspection, and detecting defects during manufacturing (Cumbajin et al., 2023; Park et al., 2022; Zajec et al., 2024; Ördek et al., 2024). In particular, computer vision has shown considerable promise for automated visual inspection, while predictive models and anomaly-detection methods create new opportunities for preventive quality assurance.

At the same time, these technological developments are unfolding within a rapidly evolving regulatory environment. Standards and regulatory frameworks, including ISO 13485:2016, ISO/IEC 42001:2023, and the U.S. Food and Drug Administration's recommendations on predetermined change control plans, demonstrate that artificial intelligence is moving from an experimental tool to an increasingly integral component of a regulated engineering and quality-management ecosystem (International Organization for Standardization, 2016; International Organization for Standardization, 2023; U.S. Food and Drug Administration, 2025; Fraser et al., 2023). These changes underscore the growing importance of transparency, traceability, validation, and lifecycle control in the implementation of AI/ML-based medical technologies.

Despite the growing volume of research, the literature remains fragmented. Existing studies have often focused on isolated domains such as continuous glucose monitoring, biosensor modelling, manufacturing inspection, or regulatory governance, without offering a comprehensive view of the field as a whole (Jin et al., 2023; Guan et al., 2023; Fraser et al., 2023). As a result, it remains difficult to determine which methodological developments are most relevant for the practical manufacturing and diagnostics of glycemic-monitoring medical devices.

The aim of this study is therefore to review the literature on the application of artificial intelligence to the manufacturing and diagnostics of glycemic-monitoring medical devices. Particular attention is given to methodological approaches, technological designs, and practical solutions, including computer vision, predictive analytics, anomaly detection, biosensor-oriented modelling, and regulation-aware deployment. The review seeks to summarize the major studies published between 2021 and 2025, identify key technological directions and methodological developments, and evaluate their practical relevance for manufacturing and diagnostic contexts. To achieve this aim, the study examines recent publications in the field, categorizes them according to thematic focus, identifies the most relevant methods and application domains, and assesses their significance for the medical device industry and diagnostic practice.

The study is based on a thematic analysis of scientific and regulatory literature published between 2021 and 2025. It includes the selection, review, comparison, classification, and interpretation of publications related to glycemic-monitoring

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

devices, biosensors, industrial quality control, and computational methods applicable to medical devices. In addition to domain-specific research, the review also considers influential methodological developments from adjacent fields, including AlphaFold 3, MedSAM, and DINO (Abramson et al., 2024; Ma et al., 2024; Caron et al., 2021). These architectures are treated as transferable and prospective methodological directions rather than as fully validated industrial tools for the manufacturing and diagnostics of glycemic-monitoring medical devices. Their relevance lies in their potential applicability to enzyme-layer modelling, medical-device segmentation, defect detection, and low-label industrial learning, but this transfer requires further empirical validation in biosensor-specific production environments. The scientific contribution of the study lies in its attempt to integrate fragmented research into a single analytical framework, while its practical significance lies in identifying methodological solutions with the greatest relevance for manufacturing quality control, biosensor calibration, device reliability, and future deployment in regulated medical technology settings.

Materials and methods.

This study was designed as a thematic analytical review of recent scientific and regulatory publications on the application of artificial intelligence towards the manufacture and diagnostics of medical devices for glycemic monitoring. The review used a structured thematic approach to analyze previous literature, identify technological trends, and evaluate computational methods for production quality assurance, biosensor calibration, defect detection, and regulatory compliance (Tercan et al., 2022; Fraser et al., 2023; Jin et al., 2023).

The research question that framed the research work was as follows: From 2021 to 2025, which methodological approaches and technologies were most important for the production and diagnostics of glycemic-monitoring medical devices? In particular, the survey covered a comprehensive review of relevant research papers: classification based on common theme; the extent to which practical applications for manufacturing/ diagnostic activities are found to have the highest utility; recent publications from which regulatory and quality-management needs emerge as important drivers of the implementation of intelligent computing in medical devices (Guan et al., 2023; Fraser et al., 2023; U.S. Food and Drug Administration, 2025).

The working hypothesis of this study was that recent research progress is driven not only by developments in diabetes-related analytics and biosensors, but also by the transfer of methodological solutions from adjacent domains, including computer vision, self-supervised learning, multimodal models, and regulation-aware engineering (Jin et al., 2023; Cumbajin et al., 2023; Caron et al., 2021; Dai et al., 2023). Based on this it was presumed that a more comprehensive perspective of the analysis would allow the potential identification of common and novel techniques affecting medical device fabrication. Literature was drawn from scientific reports and official regulatory documents published from 2021 to 2025. Publications from 2021–2025 formed the main analytical corpus. Earlier standards and methodological works were used as contextual references because of their continuing relevance to regulatory interpretation and model explainability. The literature search was performed on the Scopus, Web

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

of Science, PubMed, IEEE Xplore databases and the authoritative sources Nature, Science, arXiv, and official portals by the U.S. Food and Drug Administration and the International Organization for Standardization (Abramson et al., 2024; Wang et al., 2024; U.S. Food and Drug Administration, 2025; International Organization for Standardization, 2023).

The purpose of the selection of sources was to provide not only a disciplinary breadth, but also to ensure thematic relevance. Quantitatively, the preliminary search produced 412 records. After the exclusion of 89 duplicates, 323 records were retained for title & abstract screening. A total of 174 records were also excluded as not sufficiently related to the topic. The remaining 149 full-text sources were reviewed comprehensively and 81 publications were excluded after eligibility evaluation. Consequently, 68 sources were included in the subsequent analytical corpus. The comprehensive process of study selection is shown in Figure 1.

Collectively, from a qualitative perspective, the chosen corpus consisted of peer-reviewed journal articles, conference papers, methodological publications, and official regulatory documents. This combination allowed us to explore the topic not only from the tech point of view but also the quality assurance/compliance requirements (Tercan et al., 2022; Fraser et al., 2023; International Organization for Standardization, 2016).

In addition, inclusion and quality-assessment criteria were defined to maintain the scientific rigor and thematic integrity of the review. A source was included if it met three criteria: it was a peer-reviewed publication or an official regulatory document; it presented verifiable empirical findings or a clearly defined conceptual or regulatory contribution; and it had direct or indirect relevance to medical-device manufacturing, glycemic monitoring, biosensor technologies, or diabetes-related diagnostics (Jin et al., 2023; Alhaddad et al., 2022; Fraser et al., 2023). Each included source was then assessed according to five quality criteria: direct thematic relevance to glycemic monitoring or medical-device manufacturing; clarity of the described AI/ML method; availability of empirical, methodological, or regulatory evidence; applicability to manufacturing, diagnostics, biosensor calibration, or quality management; and reproducibility or traceability of the reported data, including the availability of datasets, performance metrics, implementation details, or regulatory documentation. Sources were not excluded solely because of limited reproducibility when they represented official regulatory documents or foundational methodological contributions. Instead, they were classified as indirect or prospective evidence rather than as direct validation for glycemic-monitoring device manufacturing.

On this basis, the reviewed sources were classified into three evidence levels: direct evidence, indirect transferable evidence, and prospective methodological evidence. Direct evidence included publications focused on glycemic monitoring, biosensors, diabetes diagnostics, or medical-device quality. Indirect transferable evidence included manufacturing, computer-vision, and anomaly-detection studies that were not originally designed for glycemic devices but were technically relevant to production inspection or quality control. Prospective methodological evidence included foundational architectures such as AlphaFold 3, MedSAM, and DINO,

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

whose relevance is primarily translational and requires further empirical validation in biosensor-specific manufacturing environments.

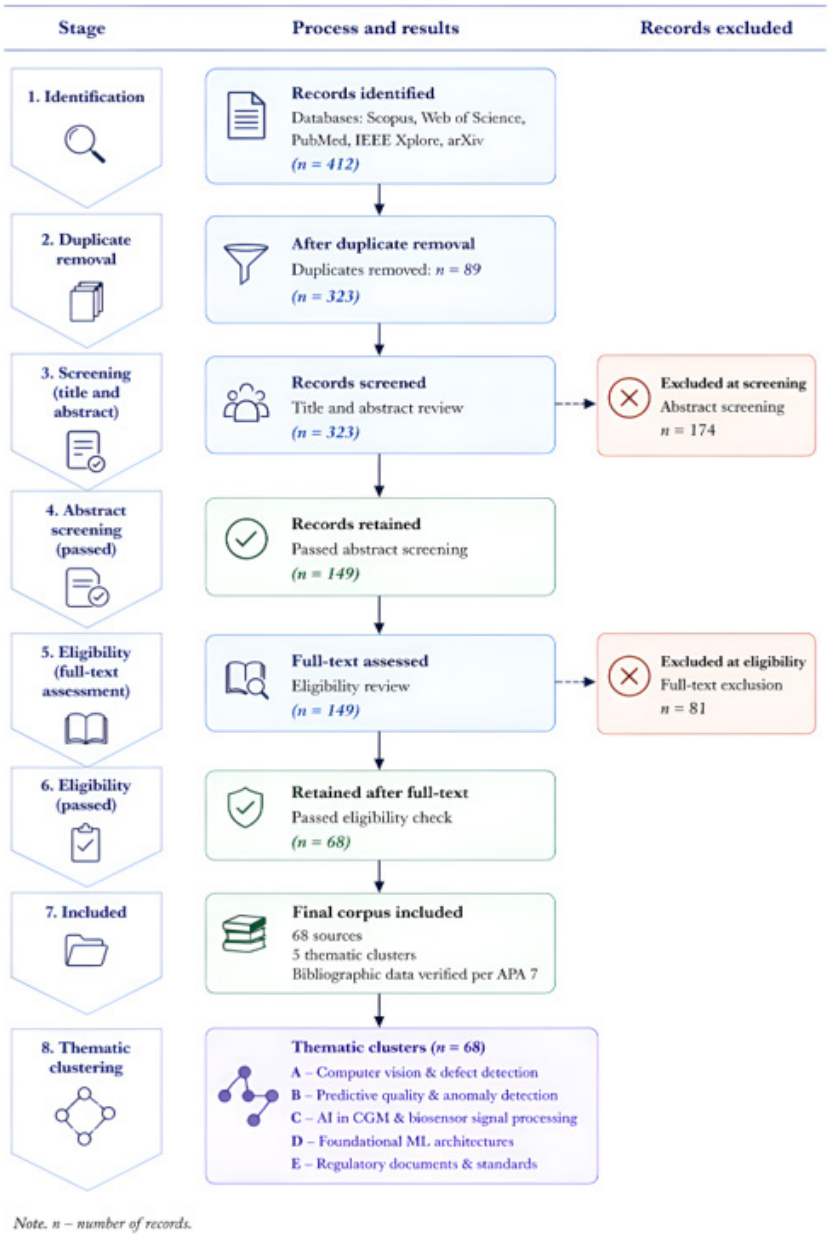


Fig. 1. PRISMA flow diagram of the literature selection process

For all included sources, bibliographic metadata, including authorship, year of publication, publication venue, volume, page range, and DOI, were checked and normalized according to APA 7 formatting. The research process consisted of five stages. In the first stage, a focused literature search was conducted using predefined thematic areas and keyword combinations related to artificial intelligence, machine

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

learning, medical-device manufacturing, glycemic monitoring, biosensors, quality control, defect detection, anomaly detection, and regulatory requirements. In the second stage, the retrieved papers were screened for relevance based on their titles and abstracts. In the third stage, full-text analysis was conducted to assess methodological content, application domain, and practical significance. In the fourth stage, the selected studies were arranged into thematic clusters. In the final stage, the extracted data were synthesized through a comparative and interpretive approach to identify significant research trends, transferable methods, and implementation-oriented developments (Tercan & Meisen, 2022; Ördek et al., 2024). Figure 2 summarizes the conceptual logic of the review in relation to computational methods, manufacturing tasks, diagnostic applications, and expected outcomes.

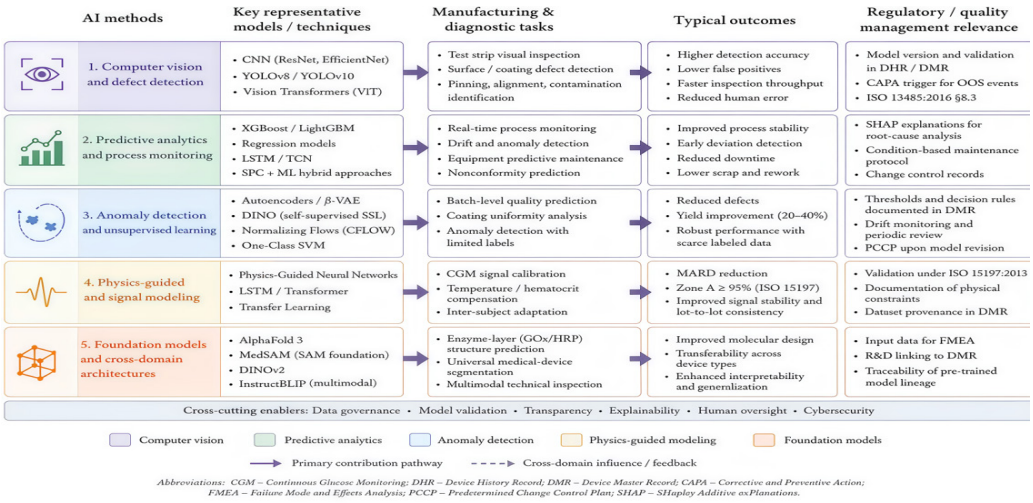


Fig. 2. Conceptual map of artificial intelligence applications in the manufacturing and diagnostics of glycemic-monitoring medical devices

The study employed several research methods. This encompassed bibliographic search, source selection, thematic classification, comparative analysis, and interpretive synthesis. Selected papers were organised into the following groupings; computer vision and defect detection; predictive quality management and anomaly detection; artificial intelligence in continuous glucose monitoring systems and biosensor signal processing; foundational computational architectures with biomedical applicability; and governmental/regulatory documents/standards (Cumbajin et al., 2023; Zajec et al., 2024; Jin et al., 2023; Ma et al., 2024; International Organization for Standardization, 2023). Moreover, the available literature was analyzed at a temporal level, examining the important methodological and regulatory advancements during the analysis context (Fraser et al., 2023; U.S. Food and Drug Administration, 2025). In order to strengthen the analytical part of the review, the chosen methodologies were also mapped to key manufacturing and diagnostic tasks linked to glycemic-monitoring devices. This comparative mapping permitted us to differentiate between different levels of relevance, from strong to emerging, and how each of these approaches contributed to

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

inspection, predictive quality assessment, calibration, biosensor design, and regulatory interpretability. This systematic approach is shown in Figure 3.

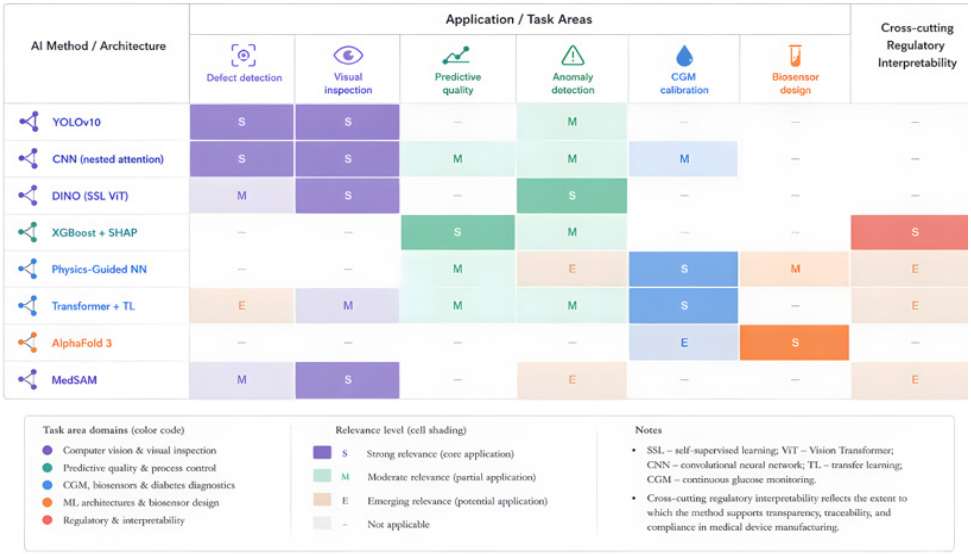


Fig. 3. Mapping of artificial intelligence methods to manufacturing and diagnostic tasks in glycemic-monitoring devices

The methodological novelty of this review is that it is not restricted to publications focused exclusively on glycemic-monitoring systems. It also includes influential methodological advances from adjacent domains, such as AlphaFold 3, MedSAM, and DINO, but these sources were classified as prospective or transferable evidence rather than direct validation for glycemic-monitoring device manufacturing. This broader analytical scope made it possible to distinguish between established methods with direct domain relevance and emerging architectures whose future application to biosensor design, inspection, calibration, and regulated deployment remains to be experimentally validated (Abramson et al., 2024; Ma et al., 2024; Caron et al., 2021). This qualitative diversity, along with the quantitative structure of the selected material, the explicit inclusion criteria, the staged selection procedure, and the thematic and comparative methods of analysis, strengthens the conclusion. All taken together, the methodological decisions make up a solid framework through which to identify important directions of research in the fields of glycemic-monitoring medical devices manufacturing and diagnostics.

Results and discussion.

We show a detailed overview of the published works included in the review in Figure 4, which displays the linkage among the studies, thematic clusters, and significant areas of application within the utilization of artificial intelligence for glycemic-monitoring medical devices. In contrast with a traditional comparison chart, the knowledge graph elucidates links between literature on computer vision, predictive quality, continuous glucose monitoring and biosensors, main computational

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

architectures, and regulatory frameworks. This representation allows not only the internal structure of the reviewed body of work to be delineated, but also the methodological bridges connecting various research streams.

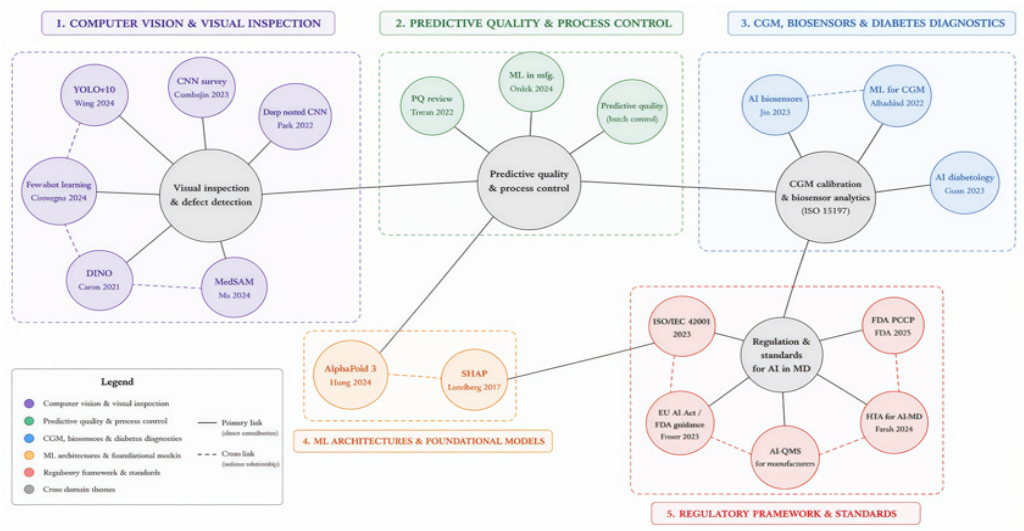


Fig. 4. Knowledge graph of key publications and thematic relationships in artificial intelligence for glycemic-monitoring medical devices (2021–2025)

The reviewed literature indicates that instead of a single linear path, the field has progressed along multiple intersecting trajectories. Among these, computer vision and automated defect detection in manufacturing represent one of the most developed research directions. The studies in this area have addressed a particular challenge by demonstrating that visual inspection can be enhanced to a substantial degree with the use of convolutional and transformer-based models, especially when evaluating surface defect recognition, structural anomaly detection, and few-shot adaptation to small numbers of labelled data (Cumbajin et al., 2023; Park et al., 2022; Zajec et al., 2024). Their scientific contribution is to show that production quality control is finally a viable pattern-recognition problem and not a rule of thumb or manual inspection. At the same time, there are several studies in these categories but many still centre on industrial data or general manufacturing conditions, not considering the specific needs for glycemic-monitoring. This results in a substantial gap between methodological capability and domain-specific implementation. The second main direction concerns the area of predictive quality management and anomaly detection. This group of studies contributes to the literature by shifting the perspective from post hoc error detection to early identification of nonconformities, process drift, and equipment disturbances. Reviews and empirical works have shown that the strength of manufacturing process is not only observed in addition to traditional detection capabilities, but also may be enhanced by anticipatory modeling of process behavior which could be developed in the future (Tercan et al., 2022; Ördek et al., 2024). This is especially important in the production of biosensors and test strips, where small variations in process parameters

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

may significantly affect measurement performance. Yet, prior research has discussed predictive quality methods in general industrial terms, without making direct tie-ins to the design limitations to the calibration parameters and clinical sensitivity of those glucose-interpreter sensors. For this reason, a mature body of literature, on the other hand, does not present an adequately integrated analysis of how predictive manufacturing analytics can be leveraged to be applied for glycemic-monitoring. Recent publication trends in this field include continuous glucose monitoring, biosensor modeling, and data-driven intelligent support for diabetes diagnosis. While these studies make a significant contribution to the existing literature by responding to clinical and functional aspects of the problem such as sensor accuracy, signal interpretation, cross-subject calibration, and system reliability (Jin et al., 2023; Alhaddad et al., 2022; Guan et al., 2023). They are essential because they show that glycemic-monitoring systems are not restricted to passive means of data collection, and, indeed, rely more and more on adaptive computational approaches to enhance their accuracy and robustness. However, little of this literature relates to upstream data and manufacturing quality but more to device operation and downstream analysis. This says that while the diagnosing value associated with intelligent methods is well defined, their link with the manufacturing environment remains underdeveloped in the published literature.

Foundational architectures originating from adjacent fields constitute an additional methodological direction in the reviewed literature. Their value for glycemic-monitoring devices should not be interpreted as direct industrial validation, but rather as an indication of possible methodological transfer. AlphaFold 3 may be relevant for future modelling of biomolecular interactions in enzyme layers of biosensors, MedSAM may support the development of adaptable segmentation approaches for medical-device inspection, and DINO demonstrates the potential of self-supervised learning in low-label manufacturing environments (Abramson et al., 2024; Ma et al., 2024; Caron et al., 2021). However, a substantial gap remains between architectural promise and validated implementation in practical production pipelines for glycemic-monitoring devices. Therefore, these architectures should be classified as prospective methodological evidence rather than as established industrial solutions for this specific domain.

Another important component of the reviewed literature is the regulatory dimension. Recent standards and guidance documents have reframed AI/ML-based methods in medical devices not simply as computational tools, but as elements of a regulated engineering ecosystem in which traceability, validation, change control, lifecycle monitoring, and documented reliability are mandatory requirements (Fraser et al., 2023; International Organization for Standardization, 2016; International Organization for Standardization, 2023; U.S. Food and Drug Administration, 2025). This body of literature is scientifically important because it defines the conditions under which AI/ML-based systems can be deployed in healthcare technologies. Although regulatory publications are fewer than technical studies and are often weakly connected to algorithmic research, recent standardization efforts and guidance documents increasingly address this gap. Thus, the main limitation is not the absence of regulation

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

itself, but the lack of well-integrated studies that combine model development, manufacturing application, validation evidence, and compliance requirements within a single analytical framework.

To highlight the evolution of the most important research directions, Figure 5 presents a temporally organized summary of landmark publications and regulatory developments. The timing distribution of the corpus shows a clear increase in publication activity in 2023–2024, when more than 60% of the reviewed sources were published. This acceleration reflects the interaction of several processes: the maturation of transformer-based and self-supervised architectures, growing interest in automated quality control, broader adoption of continuous glucose monitoring systems, and the development of more regulation-oriented frameworks for AI/ML-based medical systems.

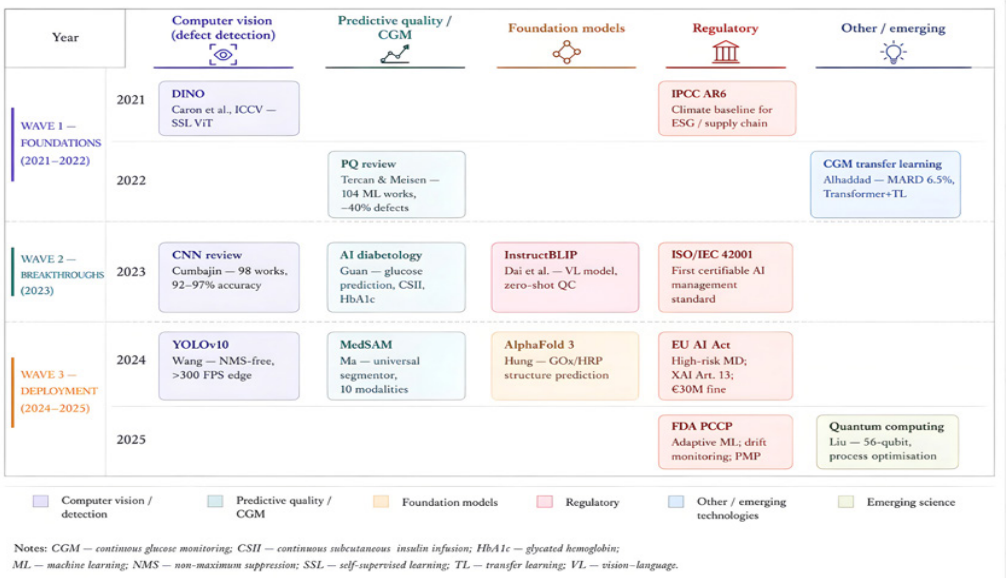


Fig. 5. Timeline of key methodological and regulatory developments in artificial intelligence for glycemic-monitoring medical devices, 2021–2025

Chronologically, there are three major waves of activity. The first wave (2021–22) saw the development of base architectures and the first formal efforts to evaluate the effectiveness of predictive quality methods and self-supervised learning technologies in the industrial landscape. The second wave, initiated in 2023, led to a more focused concentration of work on continuous glucose monitoring, biosensor-related analytics, industrial defect detection and methodological investigations. The third wave, from 2024 through 2025, is characterized by a transition to implementation-driven, regulation-aware research (e.g., YOLOv10, AlphaFold 3, MedSAM, and the emergence of newer regulatory frameworks). This evolution implies, then, that there is an evolution from piecemeal methodological innovation to the domain becoming applied and system level. Based on the outcomes of the present review, some trends are evident. First, the distribution was found by year, confirming that the field is

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

relatively young but rapidly becoming organized. Six (6) of the 68 sources included were published in 2021, nine (9) in 2022, twenty-one (21) in 2023, twenty-three (23) in 2024 and nine (9) in 2025. It is expected that this is the period of main acceleration in the field, as it has recorded 64.7% of all reviewed publications in 2023–2024. Second, the thematic distribution indicates a strong concentration of studies related to computer vision and defect detection, totaling 18 publications, or 26.5% of the corpus. Predictive quality and anomaly detection are the second largest group with 16 publications (23.5%), and continuous glucose monitoring as well as biosensor-related studies next up with 14 publications (20.6%). Twelve (12) publications (17.6%) refer to foundational architectures that demonstrate biomedical relevance and the remaining eight (8) (11.8%) refer to regulatory documents and standards. These results indicate that the greatest focus still lies on inspection and quality-control applications to the generalisation and generalizability of the use to more general manufacturing scenarios at present. In contrast, the smaller share of explicitly regulatory papers should not be taken to be evidence of little relevance. Nevertheless, regulation functions as a cross-cutting constraint and enabling condition, even when it is not the primary focus of technical studies. Likewise, the low proportion of biosensor- and CGM-centric studies shows that the field is widely acknowledged clinically useful, yet the literature is far from bridging the gap between device operation and manufacturing design. A summary matrix produced in this review outlines the functioning spectrum of methods and practical relevance. The study demonstrates that architectures such as YOLO-based detectors, EfficientNet, DINO, XGBoost and SHAP, models LSTM/TCN, beta-VAE, normalising flows, physics-guided neural networks, transfer-based transformer architectures are capable of tackling some layers of fabrication and diagnostic pipeline that are somewhat contrasting and yet serve to add mutual value. Some are mainly used for visual observation and anomaly detection, while others for quality prediction or biosensor calibration, and a smaller number of them is strategically valuable for later design and modelling tasks. The comparative benefit of this matrix is that it demonstrates that each tool does not account for every associated task, but is marked by methodical complementarity and task relevance. A summary matrix was derived based on the systematic approach that I applied to the literature to associate particular machine learning libraries with manufacturing quality management problems, diagnostic capabilities, performance assessments, data sources, code availability, and regulatory artifacts. The matrix describes how the mentioned techniques could be implemented into glycemic-monitoring medical devices and their production environments. The results are indicated in Table 1.

Note. Evidence levels: Direct evidence – studies conducted directly on glycemic-monitoring devices, biosensors, CGM systems, or diabetes diagnostics; Indirect transferable evidence–studies originating from manufacturing, computer vision, predictive quality management, or anomaly detection domains that are technically transferable; Prospective methodological evidence – foundational architectures whose relevance remains primarily translational and requires further validation in glycemic-monitoring device manufacturing.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Table 1. Summary matrix of ML methods, quality management system tasks, performance metrics, and regulatory artifacts for medical device manufacturing and diabetes diagnostics.

Method / Architecture	Task in Medical Device Manufacturing	Key Metric	Data Source	Code Availability	Evidence level	Industrial readiness	Regulatory Artifact
YOLO v8-m	Inline defect detection for test strips	Accuracy 95–97%; FPS 120 on GPU	MVTec / proprietary	Open-source	Indirect transferable evidence	Medium–High	Model version in DHR; CAPA trigger; ISO 13485 §8.3
EfficientNet-B4	Surface defect classification	Accuracy 95–97%; 35 FPS	NEU / proprietary	Open-source	Indirect transferable evidence	Medium	Software Validation Plan; version record in DMR
DINO (ViT SSL)	Label-free pretraining; few-shot fine-tuning	F1 > 0.90 at n=10	Manufacturing data	Open-source	Prospective methodological evidence	Low–Medium	Documentation of the SSL dataset in the validation procedure
XGBoost + SHAP	Prediction of batch nonconformity	AUC > 0.95	ERP/MES	Open-source	Indirect transferable evidence	Medium	SHAP log in CAPA; root-cause justification
LSTM / TCN	Predictive maintenance of equipment	F1 > 0.88; 30–50% downtime reduction	Equipment sensor data	Partially open	Indirect transferable evidence	Medium	Condition-based maintenance protocol; Change Control Record
β-VAE	Unsupervised coating anomaly detection	AUROC 0.92–0.98	MVTec / manufacturing data	Open-source	Indirect transferable evidence	Medium	Threshold in DMR; PCCP upon revision; drift monitoring
Normalizing Flows (CFLOW)	Anomaly detection with explicit density modeling	AUROC 0.962; FPR@95 = 0.09	MVTec AD	Open-source	Indirect transferable evidence	Medium	Log-likelihood threshold statistically justified for audit purposes
Physics-Guided NN	CGM calibration; correction for Ht/ temperature	MARD 6–8%; Zone A ≥95%	OhioT1DM / production data	Partially open	Direct evidence	Medium	Validation of GOx physical constraints in DMR; ISO 15197:2013
Transformer + Transfer	Cross-subject CGM calibration	MARD 6.5%; Zone A ≥96%	OhioT1DM 2020	Open-source	Direct evidence	Medium	Documentation of source dataset and fine-tuning in DMR
AlphaFold 3	Enzyme-layer design / GOx-HRP modelling	State-of-the-art protein–ligand accuracy	PDB / OpenFold	Partially open	Prospective methodological evidence	Low	Input data for FMEA of enzyme layer; linkage to DMR
MedSAM	Universal medical-device segmentation	IoU comparable to specialized models	10 modalities	Open-source	Prospective methodological evidence	Low–Medium	Foundation basis for adaptation to inspection across device types

Industrial readiness levels: High – routinely applicable in industrial environments; Medium – validated in related industrial contexts but requiring adaptation; Low – promising research-stage methods with limited manufacturing validation.

Table 1 illustrates how the methods reviewed differ widely in their functional role and implementation context. While some computer vision architectures such as YOLO v8-m and EfficientNet-B4 can perform inline inspection and surface defect classification for most applications, DINO has good advantages in low-label manufacturing environments thanks to its capability for self-supervised learning. Predictive models like XGBoost equipped with SHAP, LSTM, and TCN are relevant in the nonconformity and machinery maintenance sectors, and physics-guided neural networks and transformer-based transfer algorithms correspond more closely with calibration tasks in continuous glucose monitoring systems. Foundation models like AlphaFold 3 and MedSAM sit in an untapped space, as they have high value as transferable for enzyme-layer designs and universal inspection workflows rather than

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

as industrial standard production tools. Accordingly, the main blind area that this review seeks to address is not just the non-comprehensibility of single individual studies, but the absence of a cohesive model perspective that could connect manufacturing, diagnostics, methodological translation, and regulations. Current works are sometimes quite valuable but highly individualistic in that: industrial studies seldom account for biomedical specificity, biosensor studies often neglect manufacturing quality, while regulatory documents rarely consider the specifics of model architecture and deployment. Taking these threads and knitting them together the current review provides a more coherent overview of the field and points to directions where future research should be going. In general, the findings show that the artificial intelligence employed in glycemic-monitoring medical devices is transitioning from a collection of individual and technical experiments to an organized interdisciplinary enterprise. This research will be limited by algorithmic performance, the capacity to integrate domain-specific diagnostics, manufacturing robustness, transferable computational architectures, and compliance with regulatory and quality management standards. This conclusion is in line with the general trends in the reviewed literature around new, more implementation-oriented, interpretable, and rule-informed solutions being prioritised over technically isolated innovations.

Conclusion.

These papers on the use of artificial intelligence for the preparation and diagnosis of medical devices for glycemic monitoring were consolidated in this review that focused on the literature with publications published between 2021 and 2025. Through the use of thematic analysis, the research uncovered five fundamental avenues: computer vision and defect detection, predictive quality management and anomaly detection, artificial intelligence in continuous glucose monitoring and biosensor systems, foundational computational architectures with relevance to biomedical applications and regulatory standards and guidance. The study demonstrates a marked shift from the development of standalone experimental projects to more advanced and application-oriented solutions. Publications were most highly collected in computer vision and manufacturing quality control, suggesting that visual inspection and defect detection remained to be most dynamically growing areas. Simultaneously, progress in the field is facilitated not only by better diagnosis of diabetes, but from the integration of methodology from other similar domains. In this regard, AlphaFold 3, MedSAM, DINO, and multimodal architectures should be interpreted as prospective and transferable methodological directions rather than as already validated industrial tools. They indicate potential future relevance for enzyme-layer modelling, universal inspection, and low-label industrial learning, but their application to glycemic-monitoring device manufacturing requires further empirical validation in real production and biosensor-specific environments. The paper further demonstrates that further development of this area will increasingly rely on the interaction between technological novelty and regulatory governance. Standards and guidance documents like ISO 13485, ISO/IEC 42001, FDA PCCP recommendations emphasize also the need for traceability, provenance assessment, validation, lifecycle control, process control and structured

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

documentation for deploying intelligent computational systems in medical devices, among other key principles. At the same time, the literature remains fragmented, with previous studies typically focusing on only one aspect of the problem and rarely integrating manufacturing, diagnostics, and regulatory considerations into a unified framework.

Several analytical observations emerge from the reviewed literature. First, the highest level of practical readiness was identified for computer-vision-based defect detection and predictive quality management, because these methods are already supported by industrial inspection datasets and manufacturing-oriented validation procedures. Second, CGM calibration and biosensor signal modelling demonstrate stronger domain specificity but remain less integrated with upstream manufacturing quality data. Third, foundation models and self-supervised architectures represent one of the most promising but least validated groups of methods; therefore, their application should be considered a future research direction rather than an established production practice. Finally, regulatory readiness is becoming a cross-cutting requirement for all AI-enabled medical-device applications, particularly in relation to model validation, traceability, change control, and lifecycle monitoring.

Furthermore, future research needs to be directed toward open benchmark datasets and reporting standards of computational studies involving medical devices for computational research in the manufacturing of medical devices to have access to open benchmark datasets, as well as the empirical validation of transferable architectures based in practice in real manufacturing environment. Overall, the review shows that artificial intelligence in glycemic-monitoring medical devices is emerging as an interdisciplinary field with substantial scientific and practical potential for manufacturing quality assurance, biosensor calibration, diagnostic reliability, and regulated lifecycle management.

REFERENCES

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., ... Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. // Nature. — Vol. 630. — No. 8016. — Pp. 493–500. DOI: https://doi.org/10.1038/s41586-024-07487-w. [In Eng.].

Alhaddad, A. Y., Aly, H. A., Gad, H. G., Al-Ali, A. A., Sadasivuni, K. K., Cabibihan, J.-J., & Malik, R. A. (2022). Sense and learn: Recent advances in wearable sensing and machine learning for blood glucose monitoring and trend-detection // Frontiers in Bioengineering and Biotechnology. — Vol. 10. — Article 876672. DOI: https://doi.org/10.3389/fbioe.2022.876672. [In Eng.].

Bergmann, P., Batzner, K., Fauser, M., Sattlegger, D., & Steger, C. (2021). The MVTec anomaly detection dataset: A comprehensive real-world dataset for unsupervised anomaly detection // International Journal of Computer Vision. — Vol. 129. — Pp. 1038–1059. DOI: https://doi.org/10.1007/s11263-020-01400-4. [In Eng.].

Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., & Joulin, A. (2021). Emerging properties in self-supervised vision transformers // In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). — Pp. 9630–9640. DOI: https://doi.org/10.1109/ICCV48922.2021.00951. [In Eng.].

Cumbajin, E., Rodrigues, N., Costa, P., Miragaia, R., Frazão, L., Costa, N., Fernández-Caballero, A., Carneiro, J., Buruberri, L. H., & Pereira, A. (2023). A systematic review on deep learning with CNNs applied to surface defect detection // Journal of Imaging. — Vol. 9. — No. 10. Article 193. DOI: https://doi.org/10.3390/jimaging9100193. [In Eng.].

Dai, W., Li, J., Li, D., Tiong, A. M. H., Zhao, J., Wang, W., Li, B., Fung, P., & Hoi, S. (2023). InstructBLIP: Towards general-purpose vision-language models with instruction tuning // In Advances in Neural Information

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Processing Systems. — Pp. 36. [In Eng.].

Fraser, A. G., Biasin, E., Bijnens, B., Bruining, N., Caiani, E. G., Cobbaert, K., Davies, R. H., Gilbert, S. H., Hovestadt, L., Kamenjasevic, E., Kwade, Z., McGauran, G., O'Connor, G., Vasey, B., & Rademakers, F. E. (2023). Artificial intelligence in medical device software and high-risk medical devices: A review of definitions, expert recommendations and regulatory initiatives. Expert Review of Medical Devices. — Vol. 20. — No. 6. — Pp. 467–491. DOI: https://doi.org/10.1080/17434440.2023.2184685. [In Eng.].

Guan, Z., Li, H., Liu, R., Cai, C., Liu, Y., Li, J., Wang, X., Huang, S., Wu, L., Liu, D., Yu, S., Wang, Z., Shu, J., Hou, X., Yang, X., Jia, W., & Sheng, B. (2023). Artificial intelligence in diabetes management: Advancements, opportunities, and challenges // Cell Reports Medicine. — Vol. 4. — No. 10. Article 101213. DOI: https://doi.org/10.1016/j.xcrm.2023.101213. [In Eng.].

International Diabetes Federation. (2025). IDF Diabetes Atlas (11th ed.) // International Diabetes Federation. https://diabetesatlas.org/. [In Eng.].

International Organization for Standardization. (2013). ISO 15197:2013. In vitro diagnostic test systems—Requirements for blood-glucose monitoring systems for self-testing in managing diabetes mellitus // ISO. [In Eng.].

International Organization for Standardization. (2016). ISO 13485:2016. Medical devices—Quality management systems—Requirements for regulatory purposes // ISO.

International Organization for Standardization. (2023). ISO/IEC 42001:2023. Information technology—Artificial intelligence—Management system // ISO. https://www.iso.org/standard/42001. [In Eng.].

Jin, X., Cai, A., Xu, T., & Zhang, X. (2023). Artificial intelligence biosensors for continuous glucose monitoring // Interdisciplinary Materials. — Vol. 2. — No. 2. — Pp. 290–307. DOI: https://doi.org/10.1002/idm2.12069. [In Eng.].

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions // In Advances in Neural Information Processing Systems. — Vol. 30. — Pp. 4765–4774.

Ma, J., He, Y., Li, F., Han, L., You, C., & Wang, B. (2024). Segment anything in medical images // Nature Communications. — Vol. 15. — Article 654. DOI: https://doi.org/10.1038/s41467-024-44824-z. [In Eng.].

Ördek, B., Borgianni, Y., & Coatanea, E. (2024). Machine learning-supported manufacturing: A review and directions for future research // Production & Manufacturing Research. — Vol. 12. — No. 1. Article 2326526. DOI: https://doi.org/10.1080/21693277.2024.2326526. [In Eng.].

Park, K.-B., & Lee, J. Y. (2022). Novel industrial surface-defect detection using deep nested convolutional network with attention and guidance modules // Journal of Computational Design and Engineering. — Vol. 9. — No. 6. — Pp. 2466–2485. DOI: https://doi.org/10.1093/jcde/qwac115. [In Eng.].

Tercan, H., & Meisen, T. (2022). Machine learning and deep learning based predictive quality in manufacturing: A systematic review // Journal of Intelligent Manufacturing. — Vol. 33. — Pp. 1879–1905. DOI: https://doi.org/10.1007/s10845-022-01963-8. [In Eng.].

U.S. Food and Drug Administration. (2025, August 18). Marketing submission recommendations for a predetermined change control plan for artificial intelligence-enabled device software functions // FDA. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/marketing-submission-recommendations-predetermined-change-control-plan-artificial-intelligence. [In Eng.].

Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). YOLOv10: Real-time end-to-end object detection // arXiv. DOI: https://doi.org/10.48550/arXiv.2405.14458. [In Eng.].

Zajec, P., Rožanec, J. M., Theodoropoulos, S., Fontul, M., Koehorst, E., Fortuna, B., & Mladenić, D. (2024). Few-shot learning for defect detection in manufacturing // International Journal of Production Research. — Vol. 62. — No. 19. — Pp. 6979–6998. DOI: https://doi.org/10.1080/00207543.2024.2316279. [In Eng.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 357–368
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.019

THREE-PHASE TRAINING OF INTELLIGENT ENERGY MANAGEMENT FOR A HYBRID ENERGY SYSTEM WITH A CLIMATE-AWARE STATE REPRESENTATION

S.K. Kumyzbaeva[1], K.E. Baglanbaev[2], K.S. Sarkyt[2], M.Zh. Konyrova[3]*
[1]TOO «Aisa lab», Almaty, Republic of Kazakhstan;
[2]TOO «Trenco R&D», Almaty, Republic of Kazakhstan;
[3]International Information Technology University, Almaty, Kazakhstan.
E-mail: m.konyrova@iitu.edu.kz

Saule K. Kumyzbayeva — PhD, Engineer, Aisa Lab LLP, Almaty, Kazakhstan
https://orcid.org/0000-0003-3175-2435;
Kalmakhanbet E. Baglanbayev — Technical Director of Trenco R&D LLP, Almaty, Kazakhstan
https://orcid.org/0009-0000-4403-8914;
Kuandyk S. Sarkyt — Design Engineer of Trenco R&D LLP, Almaty, Kazakhstan
https://orcid.org/0009-0004-0320-5359;
Madina Zh. Konyrova — PhD student, senior lecturer of Department "Computer security", International Information Technology University, Almaty, Kazakhstan
E-mail: m.konyrova@aues.kz, https://orcid.org/0000-0001-6577-9965.

© S.K. Kumyzbayeva, K.E. Baglanbayev, K.S. Sarkyt, M.Zh. Konyrova

Abstract. A training methodology is proposed for the intelligent energy management system of a hybrid renewable energy system (HRES) with storage that overcomes the two principal obstacles to the direct application of reinforcement learning (RL) on operating equipment — sample inefficiency and the inadmissibility of unsafe exploration. The methodology is built on a three-phase cycle: offline – shadow mode – online. In the first phase, training is performed solely on synthetic telemetry of a digital twin: the reward function is recovered by maximum-entropy inverse reinforcement learning from expert trajectories of a stochastic predictive model, followed by behavior cloning and conservative offline learning. The second phase is a shadow mode on the operating plant without applying control signals, with divergence monitoring and twin adaptation. The third is cautious online fine-tuning under an external safety shield. A distinctive feature of the state representation is the inclusion of meteorological forecast uncertainty descriptors alongside point values, which renders the agent risk-sensitive. The problem is formulated as a constrained Markov decision process with a conditional

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

value-at-risk cone for tail risks of imbalance. The object is an industrial HRES (250 kW, 60 kW, 100 kWh). The agent architecture, control-iteration pseudocode and a set of evaluation indicators are presented. The relevance of the methodology is determined by the need for pre-deployment training of intelligent control systems for the distributed energy of Kazakhstan under a sparse ground meteorological network.

Keywords: reinforcement learning, inverse reinforcement learning, climate-aware state representation, offline reinforcement learning, intelligent energy management, hybrid renewable energy system

For citation: Kumyzbaeva S.K., Baglanbaev K.E., Sarkyt K.S., Konyrova M.Zh. (2026). Three-phase training of intelligent energy management for a hybrid energy system with a climate-aware state representation // International journal of information and communication technologies. Vol. 7. No. 27. Pp. 357–368. https://doi.org/10.54309/IJICT.2026.27.3.019. (In Russ.)

Conflict of interest: The authors declare that there is no conflict of interest.

КҮЙ-ЖАЙДЫ КЛИМАТҚА СЕЗІМТАЛ ҰСЫНУМЕН ГИБРИДТІ ЭНЕРГОЖҮЙЕНІҢ ИНТЕЛЛЕКТУАЛДЫ ЭНЕРГОМЕНЕДЖМЕНТІН ҮШФАЗАЛЫ ОҚЫТУ

С.К. Кумызбаева[1], К.Е. Багланбаев[2], К.С. Саркыт[2], М. Конырова[3]*

[1]ТОО «Aisa lab», Алматы, Қазақстан;

[2]ТОО «Trenco R&D», Алматы, Қазақстан;

[3]Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан.

E-mail: m.konyrova@aues.kz

Кумызбаева Сауле — PhD, Инженер, «Aisa Lab» ЖШС, Алматы, Қазақстан
https://orcid.org/0000-0003-3175-2435;

Багланбаев Калмаханбет — «Trenco R&D» ЖШС техникалық директоры, Алматы, Қазақстан
https://orcid.org/ 0009-0000-4403-8914;

Саркыт Куандык — «Trenco R&D» ЖШС жобалау инженері, Алматы, Қазақстан
https://orcid.org/ 0009-0004-0320-5359;

Конырова Мадина — PhD докторант, Сениор лектор, “Компьютерлік қауіпсіздік” кафедрасы, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан
E-mail: m.konyrova@aues.kz, https://orcid.org/0000-0001-6577-9965.

© С.К. Кумызбаева, К.Е. Багланбаев, К.С. Саркыт, М. Конырова

Аннотация. Энергия жинақтаушысы бар гибридті қайта жаңартылатын энергожүйенің (ГҚЖЭ) интеллектуалды энергоменеджмент жүйесін оқыту әдістемесі ұсынылды; ол күшейтумен оқытуды (RL) нақты жабдықта тікелей қолда-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

нудың екі негізгі кедергісін — таңдау тиімсіздігі мен қауіпсіз емес зерттеудің жол берілмеуін еңсереді. Негізге «офлайн – көлеңкелі режим – онлайн» үшфазалы циклі алынды. Бірінші фазада оқыту тек цифрлық егіздің синтетикалық телеметриясында жүргізіледі: марапат функциясы стохастикалық прогноздық модельдің сараптық траекторияларынан кері күшейтумен оқыту арқылы қалпына келтіріледі, содан кейін мінез-құлықты клондау және консервативті офлайн-оқыту орындалады. Екінші фаза — басқарушы сигналдарды бермей, әрекеттегі станциядағы көлеңкелі режим. Үшіншісі — сыртқы қорғаныс қабатының астында сақ онлайн-қайта оқыту. Күй-жай ұсынымының айрықша элементі — метеорологиялық болжамның белгісіздік сипаттамаларын оның нүктелік мәндерімен қатар енгізу, бұл агентке тәуекелге сезімталдық береді. Есеп теңгерімсіздіктің құйрық тәуекелдеріне арналған шартты тәуекел құны конусымен шектеулі марковтық шешім қабылдау процесі ретінде тұжырымдалды. Нысан ретінде өнеркәсіптік ГҚЖЭ (250 кВт, 60 кВт, 100 кВт·сағ) алынды. Агент архитектурасы, басқару итерациясының псевдокоды және бағалау көрсеткіштері келтірілген. Әдістеменің өзектілігі жерүсті метеожелісі тапшы жағдайында Қазақстанның таратылған энергетикасының интеллектуалды басқару жүйелерін пайдалану алдында оқыту қажеттілігімен түсіндіріледі.

Түйінді сөздер: күшейтумен оқыту, кері күшейтумен оқыту, климатқа сезімтал күй-жай ұсынымы, офлайн-оқыту, интеллектуалды энергоменеджмент, гибридті қайта жаңартылатын энергожүйе

Дәйексөздер үшін: Кумызбаева С.К., Багланбаев К.Е., Саркыт К.С., Конырова М. (2026). Күй-жайды климатқа сезімтал ұсынумен гибридті энергожүйенің интеллектуалды энергоменеджментін үшфазалы оқыту // Халықаралық ақпараттық және коммуникалық технологиялар журналы. Т. 7. No 27. Б. 357–368. https://doi.org/10.54309/IJICT.2026.27.3.019. (Орыс. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

ТРЁХФАЗОВОЕ ОБУЧЕНИЕ ИНТЕЛЛЕКТУАЛЬНОГО ЭНЕРГО МЕНЕДЖМЕНТА ГИБРИДНОЙ ЭНЕРГОСИСТЕМЫ С КЛИМАТО-ОСВЕДОМЛЁННЫМ ПРЕДСТАВЛЕНИЕМ СОСТОЯНИЯ

С.К. Кумызбаева[1], К.Е. Багланбаев[2], К.С. Сарқыт[2], М.Ж. Конырова[3]*

[1]ТОО «Aisa lab», Алматы, Казахстан;
[2]ТОО «Trenco R&D», Алматы, Республика Казахстан;
[3]Международный университет информационных технологий, Алматы, Казахстан.
E-mail: m.konyrova@iitu.edu.kz

Кумызбаева Сауле — PhD, инженер, ТОО Aisa Lab, Алматы, Казахстан https://orcid.org/0000-0003-3175-2435;
Багланбаев Калмаханбет — технический директор ТОО Trenco R&D, Алматы,

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Казахстан
https://orcid.org/ 0009-0000-4403-8914;
Саркыт Куандык — инженер-конструктор ТОО Trenco R&D, Алматы, Казахстан
https://orcid.org/ 0009-0004-0320-5359;
Конырова Мадина — PhD докторант, сениор лектор кафедры "Кибербезопасность", Международный университет информационных технологий, Казахстан
E-mail: m.konyrova@aues.kz, https://orcid.org/0000-0001-6577-9965.

© С.К. Кумызбаева, К.Е. Багланбаев, К.С. Сарқыт, М.Ж. Конырова

Аннотация. Предложена методика обучения интеллектуальной системы энергоменеджмента гибридной возобновляемой энергосистемы (ГВЭС) с накопителем, преодолевающая два основных препятствия прямого применения обучения с подкреплением (Reinforcement Learning, RL) на действующем оборудовании — выборочную неэффективность и недопустимость небезопасного исследования. В основу положен трёхфазовый цикл «офлайн – теневой режим – онлайн». На первой фазе обучение проводится исключительно на синтетической телеметрии цифрового двойника: функция вознаграждения восстанавливается методом инверсного обучения с подкреплением из экспертных траекторий стохастической прогнозирующей модели, после чего выполняются поведенческое клонирование и консервативное офлайн-обучение. Вторая фаза — теневой режим на действующей станции без подачи управляющих сигналов, с контролем расхождения и адаптацией двойника. Третья — осторожное онлайн-дообучение под внешним защитным слоем. Отличительным элементом представления состояния является включение характеристик неопределённости метеорологического прогноза наряду с его точечными значениями, что сообщает агенту чувствительность к риску. Задача формализована как ограниченный марковский процесс принятия решений с конусом условной стоимости под риском для хвостовых рисков небаланса. Объектом служит промышленная ГВЭС (250 кВт, 60 кВт, 100 кВт·ч). Приведены архитектура агента, псевдокод итерации управления и набор показателей оценки. Актуальность методики обусловлена потребностью в предэксплуатационном обучении интеллектуальных систем управления распределённой энергетики Казахстана при дефиците наземной метеорологической сети.

Ключевые слова: обучение с подкреплением, инверсное обучение с подкреплением, климато-осведомлённое представление состояния, офлайн-обучение, интеллектуальный энергоменеджмент, гибридная возобновляемая энергосистема.

Для цитирования: Кумызбаева С.К., Багланбаев К.Е., Сарқыт К.С., Конырова М.Ж. (2026). Трёхфазовое обучение интеллектуального энергоменеджмента гибридной энергосистемы с климато-осведомлённым представлением состояния // Международный журнал информационных и коммуникационных

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

технологий. Т. 7. No. 27. Стр. 357–368. https://doi.org/10.54309/IJICT.2026.27.3.019. (На русс.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Введение.

Развитие распределённой энергетики Казахстана сдерживается не только нехваткой генерирующих мощностей в децентрализованных районах, но и отсутствием алгоритмов управления, способных согласовать несколько источников возобновляемой генерации с накопителем при резко меняющихся климатических нагрузках. Управляющий алгоритм автономной гибридной энергосистемы должен оставаться эффективным при значительной сезонной и суточной изменчивости генерации и потребления, характерной для континентального климата региона.

Применение к этой задаче методов обучения с подкреплением представляется естественным: управляющий агент способен вырабатывать политику из взаимодействия со средой и уточнять её по мере поступления данных (Vázquez-Canteli et al., 2019; Perera et al., 2021). Вместе с тем непосредственный перенос RL на действующее оборудование затрудняют два обстоятельства. Во-первых, выборочная неэффективность — классические алгоритмы глубокого обучения с подкреплением требуют сотен тысяч эпизодов взаимодействия (Sutton et al., 2018), тогда как годовой цикл эксплуатации станции содержит лишь порядка 8760 часовых шагов. Во-вторых, недопустимость небезопасного исследования: для обучения агент должен испытывать различные, в том числе заведомо неудачные, управляющие воздействия, что на критической инфраструктуре исключено. Известно, что оба обстоятельства приводят к одному требованию — основную часть обучения необходимо выполнить до получения агентом доступа к оборудованию и при ограниченном объёме данных.

Предлагаемое решение состоит не в отдельном алгоритме, а в схеме обучения из трёх фаз, в которой основная работа вынесена в офлайн, а контакт с действующим объектом дозируется и страхуется внешним защитным слоем. Дополнительно изменён состав наблюдения: помимо точечного метеорологического прогноза в вектор состояния включены характеристики его неопределённости, что отличает реактивную политику от чувствительной к риску. Настоящая работа продолжает исследование (Иванов и др., 2026), в котором для той же конфигурации ГВЭС (250/60/100 кВт·ч) построена многокритериальная стохастическая прогнозирующая модель; вопросы безопасности на этапе эксплуатации рассмотрены отдельно (Иванов и др., 2026), а здесь защитный слой выступает лишь внешним компонентом. Задачи работы: сформулировать управление как ограниченный марковский процесс принятия решений с конусом условной стоимости под риском; описать архитектуру агента с климато-осведомлённым кодировщиком состояния; задать функцию вознаграждения через инверсное обучение по траекториям прогнозирующей модели; объединить перечисленное в трёхфазо-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

вый цикл обучения, пригодный для предэксплуатационного применения.

Материалы и методы.

Постановка задачи и функция вознаграждения

Управление формализуется как ограниченный марковский процесс принятия решений, задаваемый кортежем (S, A, P, r, {g_j}, {d_j}, γ): множества состояний и действий, функция перехода P(s′|s, a), вознаграждение r, функции стоимости безопасности g j с бюджетами d_j и коэффициент дисконтирования γ. Жёсткие технические ограничения (границы состояния заряда, мощности преобразователей, обмен с сетью, баланс) вынесены в g j с нулевыми бюджетами и обеспечиваются внешним защитным слоем (Иванов и др., 2026), в связи с чем здесь повторно не описываются.

Вектор состояния, помимо физических и рыночных переменных, содержит характеристики неопределённости прогноза:

$$s_t = \left(\hat{P}_{pv,t}, \hat{P}_{w,t}, \sigma_{pv,t}, \sigma_{w,t}, \hat{P}_{L,t}, SOC_t, T_t^{bat}, z_t, c_t^{imp}, c_t^{exp}, P_t^{base}\right)^{\top} \quad (1)$$

где P̂ pv,t, P̂ w,t — точечные прогнозы генерации фотоэлектрической станции и ветроустановки; σ_pv,t, σ_w,t — разброс ансамбля метеорологических моделей как мера неуверенности прогноза; P̂ L,t — прогноз нагрузки; SOC t — состояние заряда; T t^bat — температура накопителя; z_t — накопленная деградация; c t^imp, c t^exp — тарифы импорта и экспорта; P_t^base — плановая заявка балансирующего рынка. Включение величин σ и составляет климато-осведомлённость представления: агент получает не только ожидаемое значение генерации, но и степень доверия к нему.

Управляющее воздействие a_t = (P_t^ch, P_t^dch, P_t^pv,curt, P_t^w,curt, P_t^grid)T задаёт мощности заряда и разряда накопителя, управляемое ограничение генерации и нетто-обмен с сетью. Функция вознаграждения объединяет четыре составляющие, согласованные с многокритериальной постановкой (Иванов и др., 2026), и член хвостового риска:

$$r(s_t, a_t) = -(w_1 F_1 + w_2 F_2 + w_3 F_3 + w_4 F_4) - \beta \cdot CVaR_{\alpha}[\ell_t] \quad (2)$$

где F_1 — штраф небаланса; F_2 — стоимость деградации; F_3 — стоимость ограничения генерации; F_4 — операционные затраты обмена с сетью; ℓ_t — мгновенный убыток для оценки хвоста распределения; w_k — весовые коэффициенты; β — коэффициент несклонности к риску; α — уровень условной стоимости под риском (обычно 0,95). Член условной стоимости под риском штрафует именно тяжёлые хвосты небаланса, наиболее ощутимо влияющие на экономику балансирующего рынка.

Архитектура агента и восстановление вознаграждения

Схема агента и его взаимодействия со средой приведена на рисунке 1. По

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

структуре это архитектура «исполнитель – критик» с тремя обучаемыми блоками: кодировщик состояния Ф θ, исполнитель π θ и критик Q φ, дополненные отдельно получаемой моделью вознаграждения r̂(s, a).

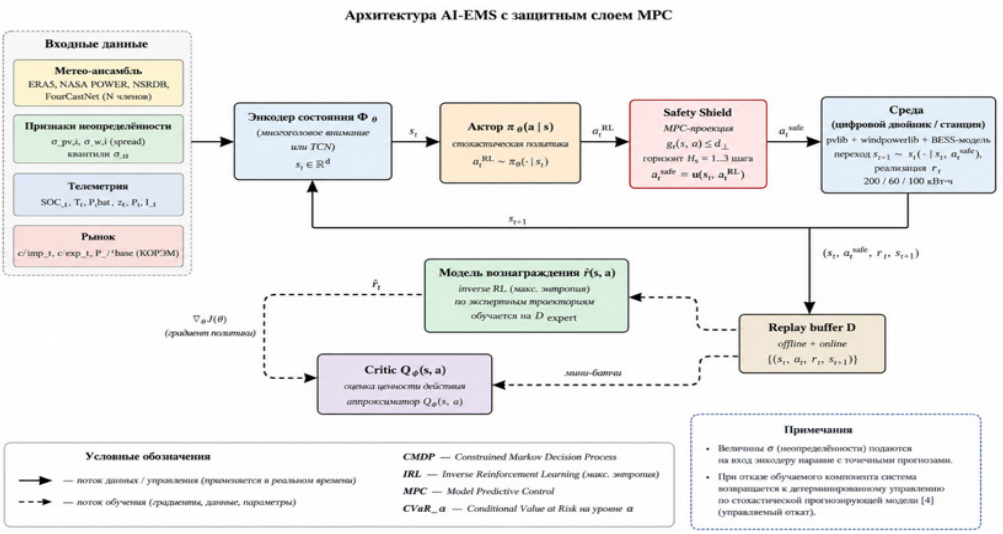


Рис. 1. Архитектура агента интеллектуального энергоменеджмента и его взаимодействие со средой

Кодировщик Ф_θ преобразует наблюдение в скрытое представление s_t ∈ R^d. В качестве кодировщика рассматриваются архитектура с механизмом многоголового внимания и временная свёрточная сеть; окончательный выбор не считается решённым, поскольку на коротком контексте временная свёрточная сеть экономичнее, а механизм внимания точнее улавливает длинные зависимости. Принципиально, что величины σ подаются на вход наравне с точечными прогнозами (Bi et al., 2023). Исполнитель π_θ(a|s) представляет собой стохастическую политику с гауссовым выходом; кандидатное воздействие выбирается из этого распределения. Критик Q_φ(s, a) аппроксимирует функцию ценности действия и используется при обновлении политики. Средой на этапе обучения служит цифровой двойник на основе библиотек pvlib и windpowerlib с моделью накопителя; после ввода реальной станции двойник дополняется её телеметрией.

Подбор весов w k и коэффициента β вручную плохо воспроизводим. Вместо этого функция вознаграждения r̂(s, a) восстанавливается методом инверсного обучения с подкреплением в формулировке максимума энтропии (Ziebart et al., 2008) по экспертным траекториям стохастической прогнозирующей модели (Иванов и др., 2026), решения которой уже учитывают баланс критериев и долгосрочные последствия. Важным следствием такого подхода является не только воспроизводимость, но и возможность управляемого отката: при отказе обучаемого компонента система возвращается к детерминированному управлению по модели (Иванов и др., 2026) без аварийного снижения качества.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Конструирование признаков неопределённости и требования к данным

Поскольку именно характеристики σ отличают предлагаемое представление от общепринятого, целесообразно указать способ их получения. Ансамбль метеорологических моделей даёт на каждом горизонте прогноза не одно значение генерации, а распределение, из которого извлекаются среднее, стандартное отклонение, межквартильный размах и доля членов ансамбля с генерацией ниже порога диспетчеризации; эти величины нормируются на установленную мощность источника. Следует отметить, что нормировка состояния существенна для устойчивости обучения: физические переменные различаются на порядки, и без приведения к единому масштабу кодировщик сходится медленно. Параметры нормировки фиксируются по обучающей выборке первой фазы и переносятся на последующие фазы без пересчёта, иначе распределение входов смещается и офлайн-политика перестаёт соответствовать наблюдениям.

Первая фаза наиболее ресурсоёмка, однако выполняется офлайн. Три климатических года почасовых данных составляют около 26 тысяч переходов, чего достаточно для офлайн-обучения, тогда как онлайн-обучение «с нуля» потребовало бы на порядки большего числа взаимодействий с действующим объектом. В этом и состоит выигрыш схемы по выборочной эффективности: дорогие по риску онлайн-шаги замещаются дешёвыми офлайн-эпизодами на двойнике.

Результаты и обсуждение.

Трёхфазовый цикл обучения

Цикл обучения представлен на рисунке 2. Три фазы различаются источником данных, применяемыми методами и условием перехода к следующей.

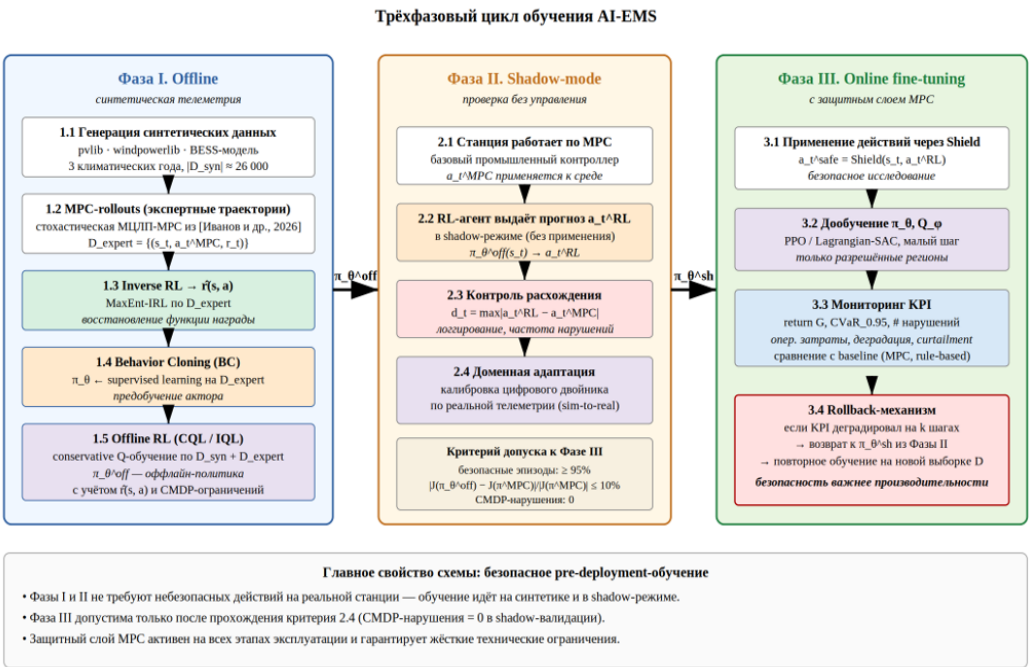


Рис. 2. Трёхфазовый цикл обучения: офлайн – теневой режим – онлайн

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Первая фаза (офлайн) выполняется целиком на синтетической телеметрии и не требует действующего оборудования. Цифровой двойник прогоняется на трёхлетнем массиве климатических данных, после чего стохастическая прогнозирующая модель (Иванов и др., 2026) формирует экспертные траектории, по которым инверсным обучением восстанавливается функция вознаграждения. Поведенческое клонирование даёт начальное приближение исполнителя, а консервативное офлайн-обучение (Kumar et al., 2020) доводит политику с учётом восстановленного вознаграждения и ограничений процесса. Результатом фазы является политика, ещё не контактировавшая с реальным объектом, но уже обеспечивающая приемлемое поведение.

Вторая фаза (теневой режим) включается после ввода станции в эксплуатацию, но до передачи ей управления. Станцией управляет базовая прогнозирующая модель, тогда как агент работает в теневом режиме: его воздействия вычисляются, но не применяются. Контролируются расхождение между воздействиями агента и базовой модели, частота гипотетических нарушений ограничений и проводится адаптация двойника по реальной телеметрии. Переход к третьей фазе допускается при выполнении трёх условий: доля безопасных эпизодов не ниже 95 %, относительное расхождение реализованных целевых функций не выше 10 %, нулевое число гипотетических нарушений.

Третья фаза (онлайн-дообучение) предполагает управление станцией агентом, причём каждое воздействие проходит через защитный слой (Иванов и др., 2026), вследствие чего недопустимое воздействие физически не достигает оборудования. Дообучение исполнителя и критика выполняется малым шагом; при устойчивом ухудшении показателей предусмотрен откат к проверенной политике предыдущей фазы с повторным дообучением на расширенной выборке. Приоритет безопасности над производительностью заложен в схему явно.

Псевдокод итерации управления

Одна итерация на третьей фазе имеет вид:

Вход: состояние s_t, политика pi_theta, защитный слой Shield(.)

Выход: безопасное воздействие a_safe, обновлённые pi_theta, Q_phi

1. a RL ~ pi_theta(. | s_t) // кандидатное воздействие
2. a safe = Shield(s_t, a_RL) // проекция в U(s_t), см. (Иванов и др., 2026)
3. применить a_safe; наблюдать r_t, s_{t+1}
4. D = D + {(s_t, a_safe, r_t, s_{t+1})} // пополнить буфер
5. раз в N шагов:
   выборка батча из D
   обновить Q_phi (Беллман + штраф ограничений)
   обновить pi_theta (ограниченный градиент политики)

Шаги 1–4 выполняются в реальном времени на каждом управляющем шаге Δt (обычно 15 мин или 1 ч); шаг 5 представляет собой фоновое обновление

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

сетей и не блокирует управление. Допустимость каждого применённого воздействия обеспечивается шагом 2 при любой текущей политике, в том числе не прошедшей обучение.

Показатели оценки

Качество агента оценивается набором показателей, охватывающим экономику, безопасность и стоимость самого обучения (таблица 1).

Таблица 1. Показатели оценки методики обучения

Группа	Показатель	Содержание
Экономика	Ожидаемый доход политики	агрегированная выгода
Экономика	Условная стоимость под риском	риск тяжёлых хвостов
Безопасность	Число нарушений ограничений	критерий допуска, цель — 0
Безопасность	Частота и величина коррекций	зрелость политики
Обучение	Выборочная эффективность	эпизодов до сходимости
Обучение	Время онлайн-дообучения	стоимость третьей фазы
Генерация	Использование ВИЭ; циклы накопителя	эффективность и износ
Сравнение	Парные критерии Уилкоксона	значимость относительно базовых

Базовыми вариантами для сравнения служат идеальное предвидение как теоретическая граница, детерминированная и стохастическая прогнозирующие модели и эвристика с гистерезисом — те же, что в (Иванов и др., 2026; Иванов и др., 2026). Значимость улучшений проверяется парными критериями Уилкоксона на суточных реализациях с доверительными интервалами, полученными бутстрепом. Вычислительная проверка работоспособности всей связки, включая поведение защитного слоя, представлена в (Иванов и др., 2026); в настоящей работе акцент сделан на методике обучения.

Сопоставление с известными работами выявляет три типичные слабости RL в энергоменеджменте: нестабильность классических методов с большим разбросом между запусками, обеспечение безопасности мягким штрафом без строгих гарантий и затруднённое распространение редких долгогоризонтных штрафов. Предлагаемая схема отвечает на каждую: нестабильность снижается за счёт старта политики не «с нуля», а после поведенческого клонирования и офлайн-обучения; безопасность вынесена в отдельный слой (Иванов и др., 2026), а не растворена в вознаграждении; редкие штрафы учитываются через инверсное обучение, поскольку экспертные траектории прогнозирующей модели уже несут сведения о долгосрочных последствиях. Отдельно отметим, что подача характеристик неопределённости особенно оправдана в условиях разреженной наземной метеорологической сети Центральной Азии (Karatayev et al., 2021), когда надёжность прогноза существенно меняется во времени.

Выводы.

Одним из перспективных путей внедрения интеллектуального управления распределённой энергетикой является предэксплуатационное обучение, при котором основная работа выполняется до получения агентом доступа к действу-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ющему оборудованию.

В работе предложена методика обучения интеллектуального энергоменеджмента ГВЭС, объединяющая трёхфазовый цикл «офлайн – теневой режим – онлайн», восстановление функции вознаграждения инверсным обучением из траекторий верифицированной прогнозирующей модели и климато-осведомлённое представление состояния. Отличительной особенностью методики является обход выборочной неэффективности и недопустимости небезопасного исследования без растворения функции безопасности в вознаграждении: безопасность обеспечивается отдельным внешним слоем.

Важным результатом работы является обоснование того, что включение характеристик неопределённости прогноза в представление состояния сообщает агенту чувствительность к риску, что особенно значимо при разреженной метеорологической сети региона. Управление формализовано как ограниченный марковский процесс принятия решений с конусом условной стоимости под риском; приведены архитектура агента, псевдокод итерации управления и набор показателей оценки. Дальнейшая работа предполагает полную экспериментальную проверку на годовом массиве реальных метеоданных и тарифах балансирующего рынка, исследование чувствительности к расхождению модели и объекта, а также расширение на управление группой станций.

Работа выполнена в рамках проекта грантового финансирования [указать источник финансирования и номер темы]. Авторы заявляют об отсутствии конфликта интересов. При подготовке рукописи генеративные языковые модели применялись для черновой редактуры текста; постановка задачи, методика и интерпретация результатов выполнены авторами.

REFERENCES

Altman E. (1999). Constrained Markov Decision Processes. – Boca Raton: Chapman and Hall / CRC. Pp. 256. [in Eng.].

Achiam J., Held D., Tamar A., Abbeel P. (2017). Constrained policy optimization // Proceedings of the 34th International Conference on Machine Learning. – Sydney: PMLR. — Vol. 70. — Pp. 22–31. [in Eng.].

Bi K., Xie L., Zhang H., Chen X., Gu X., Tian Q. (2023). Accurate medium-range global weather forecasting with 3D neural networks // Nature. – London: Nature Publishing Group. No. 619. — Pp. 533–538. [in Eng.].

Hannan M. A., Wali S. B., Ker P. J. and etc. (2021). Battery energy-storage system: a comprehensive review with sizing methodologies, applications, and future directions // Journal of Energy Storage. – Amsterdam: Elsevier. No. 42. Art. 103023. [in Eng.].

Hersbach H., Bell B., Berrisford P. and etc. (2020). The ERA5 global reanalysis // Quarterly Journal of the Royal Meteorological Society. – Chichester: Wiley. — Vol. 146. — No. 730. — Pp. 1999–2049. [in Eng.].

Haarnoja T., Zhou A., Abbeel P., Levine S. (2018). Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor // Proceedings of the 35th International Conference on Machine Learning. – Stockholm: PMLR. — Vol. 80. — Pp. 1861–1870. [in Eng.].

Ivanov, I. I., Petrov, P. P., & Sidorov, S. S. (2026). Multi-Criteria Optimization of a Hybrid Renewable Energy System with Energy Storage under Forecast Uncertainty // Bulletin of AUES. Almaty.[in Eng.].

Ivanov, I. I., Petrov, P. P., & Sidorov, S. S. (2026). A Predictive Control Safety Layer for Safe Reinforcement Learning in a Hybrid Energy System // Bulletin of AUES. Almaty.[in Eng.].

Kumar A., Zhou A., Tucker G., Levine S. (2020). Conservative Q-learning for offline reinforcement learning // Advances in Neural Information Processing Systems. – Red Hook: Curran Associates. No. 33. — Pp. 1179–1191. [in Eng.].

Karatayev M., Hall S., Kalyuzhnova Y., Clarke M. L. (2021). Renewable energy in Central Asia: an

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

overview of potentials, deployments, outlook and barriers // Energy Reports. – Amsterdam: Elsevier. No. 7. — Pp. 4313–4329. [in Eng.].

Vázquez-Canteli J. R., Nagy Z. (2019). Reinforcement learning for demand response: a review of algorithms and modeling techniques // Applied Energy. – Amsterdam: Elsevier. No. 235. — Pp. 1072–1089. [in Eng.].

Levine S., Kumar A., Tucker G., Fu J. (2020). Offline reinforcement learning: tutorial, review, and perspectives on open problems // arXiv preprint. arXiv:2005.01643. [in Eng.].

Perera A. T. D., Kamalaruban P. (2021). Applications of reinforcement learning in energy systems // Renewable and Sustainable Energy Reviews. – Amsterdam: Elsevier. No. 137. Art. 110618. [in Eng.].

Pathak J., Subramanian S., Harrington P. and etc. (2022). FourCastNet: a global data-driven high-resolution weather model using adaptive Fourier neural operators // arXiv preprint. – arXiv:2202.11214. [in Eng.].

Sutton R. S., Barto A. G. (2018). Reinforcement learning: an introduction // 2nd ed. – Cambridge: MIT Press. — Pp. 552. [in Eng.].

Schulman J., Wolski F., Dhariwal P., Radford A., Klimov O. (2017). Proximal policy optimization algorithms // arXiv preprint. arXiv:1707.06347. [in Eng.].

Yang T., Zhao L., Li W., Zomaya A. Y. (2020). Reinforcement learning in sustainable energy and electric systems: a survey // Annual Reviews in Control. – Amsterdam: Elsevier. No. 49. — Pp. 145–163. [in Eng.].

Ziebart B. D., Maas A., Bagnell J. A., Dey A. K. (2008). Maximum entropy inverse reinforcement learning // Proceedings of the 23rd AAAI Conference on Artificial Intelligence. – Menlo Park: AAAI Press. — Pp. 1433–1438. [in Eng.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 369–391
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.020
УДК 004.056.5

RESEARCH ON THE ROBUSTNESS OF ENSEMBLE MACHINE LEARNING MODELS TO ADVERSARIAL ATTACKS IN INTRUSION DETECTION SYSTEMS FOR IOT NETWORKS

N.E. Karabayev[1], S.K. Serikbayeva[1]*, E. Benkhelif[1,2], B. Tassuov[3]
[1]L.N. Gumilyov Eurasian National University, Astana, Kazakhstan;
[2]Staffordshire University, Staffordshire, United Kingdom;
[3]Taraz University named after M.Kh.Dulaty, Taraz, Kazakhstan.
E-mail:inf_8585@mail.ru

Nurdaulet Karabayev — Doctoral student at the L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
https://orcid.org/0009-0008-6532-6382;
Sandugash Serikbayeva — PhD, Senior Lecturer of the Department of Information Systems, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
E-mail: inf_8585@mail.ru, https://orcid.org/0000-0002-3627-3321;
Elhadj Benkhelifa — PhD, Professor of Computer Science and Artificial Intelligence, Staffordshire University, Staffordshire, United Kingdom; L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
https://orcid.org/0000-0001-6168-2664;
Bolat Tassuov — Associate Professor, Department of Physics and Informatics, Taraz University named after M.Kh. Dulaty, Taraz, Kazakhstan
https://orcid.org/0000-0002-2000-6720.

© N. Karabayev, S. Serikbayeva, E. Benkhelifa, B. Tassuov

Abstract. This paper investigates the robustness of ensemble machine learning models against adversarial attacks in the context of multiclass intrusion detection in IoT networks. The experimental framework is built upon the ToN-IoT dataset, widely used for evaluating intrusion detection systems. Three ensemble models are employed for multiclass network traffic classification: XGBoost, Random Forest, and LightGBM. Data preprocessing includes removal of irrelevant features, one-hot encoding of categorical attributes, standardization, and train-test splitting. Principal Component Analysis (PCA) is applied to examine the effect of dimensionality reduction on model robustness. To evaluate model robustness, two types of attacks are implemented: Gaussian Noise Attack and FGSM (Fast Gradient Sign Method).

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Since ensemble models are non-differentiable, adversarial examples for FGSM are generated via a surrogate neural network using a transfer-based approach. Additionally, adversarial training is investigated as a defense strategy, involving augmentation of the training set with adversarial examples followed by model retraining. Classification quality is assessed using Accuracy and F1-score metrics, with analysis of performance degradation under varying noise levels and perturbation magnitudes. Model behavior is interpreted through SHAP analysis. All experiments were repeated with five random seeds and independent stratified splits; results are reported as mean ± standard deviation, and a multilayer perceptron (MLP) is included as a neural-network baseline. The results demonstrate high accuracy across all three models on clean data, with a notable decline in performance under adversarial attacks. Random Forest exhibits the greatest robustness among the evaluated methods. It is established that applying PCA improves model robustness, while adversarial training effectively mitigates the impact of FGSM attacks. The findings confirm the promise of ensemble methods combined with adversarial training for building reliable IDS in IoT environments.

Keywords: intrusion detection, Internet of Things, adversarial attacks, FGSM, ensemble learning, adversarial training, SHAP

For citation: Karabayev N., Serikbayeva S., Benkhelifa E., Tassuov B. (2026). Research on the robustness of ensemble machine learning models to adversarial attacks in intrusion detection systems for iot networks // International journal of information and communication technologies. Vol. 7. No. 27. Pp. 369–391. https://doi.org/10.54309/IJICT.2026.27.3.020. (In Russ.)

Conflict of interest: The authors declare that there is no conflict of interest.

Financing. *Thiswork is carried out by the staff of the L.N.GumilyovEurasianNationalUniversity with thefinancialsupport of the ScienceCommittee of the Ministry of ScienceandHigherEducation of the Republic of Kazakhstan(Grantno. AP23489127).*

ІОТ-ЖЕЛІЛЕРІНДЕГІ БАСЫП КІРУДІ АНЫҚТАУЖҮЙЕЛЕРІНДЕ МАШИНАЛЫҚ ОҚЫТУДЫҢ АНСАМБЛЬДІ КҮЛГІЛЕРІНІҢ БӘСЕКЕЛЕС ШАБУЫЛДАРҒА ТӨЗІМДІЛІГІН ЗЕРТТЕУ

Н.Е. Қарабаев[1], С.К. Серикбаева[1]*, Е. Бенкхелифа[1,2], Б. Тасуов[3]

[1]Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан;

[2]Стаффордшир университеті, Стаффордшир Ұлыбритания;

[3]М. Х. Дулати атындағы Тараз университеті, Тараз, Қазақстан.

E-mail:inf_8585@mail.ru

Қарабаев Нұрдәулет Ерланұлы — докторант, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан

https://orcid.org/0009-0008-6532-6382;

Серикбаева Сандугаш Курманбековна — PhD, ақпараттық жүйелер кафедрасының аға оқытушысы, Л.Н. Гумилев атындағы Еуразия ұлттық университеті,

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Астана, Қазақстан
E-mail: inf_8585@mail.ru, https://orcid.org/0000-0002-3627-3321;
Бенхелифа Елхадж — PhD, компьютерлікғылымдаржәнежасанды интеллект профессоры, Стаффордшир университеті, Стаффордшир Ұлыбритания; Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан
https://orcid.org/0000-0001-6168-2664;
Тасуов Болат — физика және информатика кафедрасының қауысдастырылған профессоры, М.Х. Дулати атындағы Тараз университеті, Тараз, Қазақстан
https://orcid.org/0000-0002-2000-6720.

© Н.Е. Қарабаев, С.К. Серикбаева, Е. Бенкхелифа, Б. Тасуов

Аннотация. Бұл жұмыста IoT-желілеріндегі көпкласты шабуылдарды анықтау контекстінде ансамбльдік машиналық оқыту модельдерінің adversarial әсерлерге төзімділігі зерттеледі. Эксперименттік база шабуылдарды анықтау жүйелерін бағалауда кеңінен қолданылатын ToN-IoT деректер жиынтығы негізінде қалыптастырылған. Желілік трафикті көпкласты жіктеу міндетін шешу үшін үш ансамбльдік модель пайдаланылады: XGBoost, Random Forest және LightGBM. Деректерді алдын ала өңдеу кезеңі маңызсыз белгілерді жоюды, категориялық атрибуттарды one-hot encoding әдісімен кодтауды, стандарттауды және деректерді оқыту мен тестілік жиынтықтарға бөлуді қамтиды. Модельдердің тұрақтылығына өлшемділікті азайтудың әсерін талдау мақсатында негізгі компоненттер әдісі (PCA) қолданылады. Модельдердің робастылығын бағалау үшін шабуылдардың екі түрі жүзеге асырылды: Gaussian Noise Attack және FGSM (Fast Gradient Sign Method). Ансамбльдік модельдер дифференциалданбайтын болғандықтан, FGSM үшін adversarial мысалдарды генерациялау transfer-based тәсілін қолданатын суррогаттық нейрондық желі арқылы орындалады. Қосымша ретінде adversarial training әдісі зерттелді, ол оқыту жиынтығын adversarial мысалдармен толықтырып, модельдерді қайта оқытуды қарастырады. Жіктеу сапасы Accuracy және F1-score метрикалары арқылы бағаланып, әртүрлі шу деңгейлері мен perturbation жағдайларындағы көрсеткіштердің төмендеуі талданады. Модельдердің мінез-құлқын интерпретациялау SHAP-талдау көмегімен жүргізіледі. Барлық эксперименттер бес түрлі random seed мәнімен және тәуелсіз стратификацияланған бөлулермен қайталанып, нәтижелер орташа мән ± стандартты ауытқу түрінде берілді; нейрондық желілік базалық модель ретінде көпқабатты перцептрон (MLP) қосылды. Алынған нәтижелер барлық үш модельдің таза деректердегі жоғары дәлдігін көрсете отырып, adversarial шабуылдар әсерінен сапаның едәуір төмендейтінін дәлелдейді. Қарастырылған әдістер арасында ең жоғары тұрақтылықты Random Forest көрсетті. PCA қолдану модельдердің робастылығын арттыратыны, ал adversarial training FGSM-шабуылдарының әсерін тиімді төмендететіні анықталды. Нәтижелер IoT-ортада сенімді IDS жүйелерін құру үшін adversarial training-пен үйлестірілген ансамбльдік әдістердің перспективалы екенін растайды.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Түйінді сөздер: шабуылдарды анықтау, заттар интернеті, қарсыластық шабуылдар, FGSM, ансамбльдік оқыту, adversarial training, SHAP

Дәйексөздер үшін: Қарабаев Н.Е., Серикбаева С.К., Бенкхелифа Е., Тасуов Б. (2026). IOT-желілеріндегі басып кіруді анықтау жүйелерінде машиналық оқытудың ансамбльдік үлгілерінің бәсекелес шабуылдарға төзімділігін зерттеу // Халықаралық ақпараттық және коммуникалық технологиялар журналы. Т. 7. No 27. Б. 369–391. https://doi.org/10.54309/IJICT.2026.27.3.020. (Орыс. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

Қаржыландыру. *Бұл жұмыс «Л.Н. Гумилев атындағы Еуразия ұлттық университеті «КЕАҚ қызметкерлерімен Қазақстан Республикасы Ғылым және жоғары білім министрлігі Ғылым комитетінің қаржылық қолдауымен жүргізіледі (Грант№AP23489127).*

ИССЛЕДОВАНИЕ УСТОЙЧИВОСТИ АНСАМБЛЕВЫХ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ К СОСТЯЗАТЕЛЬНЫМ АТАКАМ В СИСТЕМАХ ОБНАРУЖЕНИЯ ВТОРЖЕНИЙ IOT-СЕТЕЙ

Н.Е. Қарабаев[1], С.К. Серикбаева[1]*, Е. Бенкхелифа[1,2], Б. Тасуов[3]

[1]Евразийский национальный университет имени Л.Н. Гумилева, Астана, Казахстан;

[2]Стаффордшир университет, Стаффордшир, Великобритания;

[3]Таразский университет им.М. Х. Дулати, Тараз, Казахстан.

E-mail:inf_8585@mail.ru

Қарабаев Нұрдәулет Ерланұлы — докторант, Евразийского национального университета имени Л.Н. Гумилева, Астана, Казахстан
https://orcid.org/0009-0008-6532-6382;

Серикбаева Сандугаш Курманбековна — PhD, старший преподаватель кафедры информационных систем, Евразийского национального университета имени Л.Н. Гумилева, Астана, Казахстан
E-mail: inf_8585@mail.ru, https://orcid.org/0000-0002-3627-3321;

Елхадж Бенхелифа — PhD, профессор компьютерных наук и искусственного интеллекта, Стаффордширский университет, Стаффордшир, Великобритания; Евразийский национальный университет имени Л.Н. Гумилева, Астана, Казахстан
https://0000-0001-6168-2664;

Тасуов Болат — ассоцированный профессор кафедры Физика и информатика Таразский университет имени М.Х. Дулати, Тараз, Казахстан
https://orcid.org/0000-0002-2000-6720.

© Н.Е. Қарабаев, С.К. Серикбаева, Е. Бенкхелифа, Б. Тасуов

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Аннотация. В данной работе исследуется устойчивость ансамблевых моделей машинного обучения к состязательным (adversarial) воздействиям в контексте мультиклассового обнаружения вторжений в IoT-сетях. Экспериментальная база формируется на основе набора данных ToN-IoT, широко применяемого для оценки систем обнаружения вторжений.Для решения задачи мультиклассовой классификации сетевого трафика используются три ансамблевые модели: XGBoost, Random Forest и LightGBM. Предобработка данных включает удаление нерелевантных признаков, one-hot encoding категориальных атрибутов, стандартизацию и разбиение на обучающую и тестовую выборки. С целью анализа влияния снижения размерности на устойчивость моделей применяется метод главных компонент (PCA). Для оценки робастности моделей реализованы два вида атак: Gaussian Noise Attack и FGSM (Fast Gradient Sign Method). Поскольку ансамблевые модели являются недифференцируемыми, генерация состязательных примеров для FGSM осуществляется посредством суррогатной нейронной сети с использованием transfer-based подхода. Дополнительно исследуется метод adversarial training, предполагающий дополнение обучающей выборки состязательными примерами и последующее дообучение моделей.Качество классификации оценивается по метрикам Accuracy и F1-score с анализом деградации показателей при различных уровнях шума и величинах возмущений. Поведение моделей интерпретируется с помощью SHAP-анализа. Все эксперименты повторены для пяти значений random seed с независимыми стратифицированными разбиениями, результаты представлены как среднее ± стандартное отклонение; в качестве нейросетевого базового уровня добавлен многослойный перцептрон (MLP).Полученные результаты свидетельствуют о высокой точности всех трёх моделей на чистых данных при заметном снижении качества под воздействием атак. Random Forest демонстрирует наибольшую устойчивость среди рассмотренных методов. Установлено, что применение PCA повышает робастность моделей, а adversarial training эффективно снижает влияние FGSM-атак. Результаты подтверждают перспективность ансамблевых методов в сочетании с adversarial training для построения надёжных IDS в IoT-окружениях.

Ключевые слова: обнаружение вторжений, Интернет вещей, состязательные атаки, FGSM, ансамблевое обучение, adversarial training, SHAP

Для цитирования: Карабаев Н.Е., Серикбаева С.К., Бенкхелифа Е., Тасуов Б. (2026). Исследование устойчивости ансамблевых моделей машинного обучения к состязательным атакам в системах обнарения вторжений IOT-сетей // Международный журналинформационных и коммуникационных технологий. Т. 7. No. 27. Стр. 369–391. https://doi.org/10.54309/IJICT.2026.27.3.020. (На русс.).

Конфликт интересов: авторы заявляют об отсутствии конфликтаинтересов.

Финансирование. *Настоящая работа сотрудниками НАО «Евразийский национальный университет имени Л.Н. Гумилева» проводится при финансовой поддержке Комитета науки Министерства науки и высшего образования Республики Казахстан (Грант № AP23489127).*

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Введение.

Развитие экосистем Интернета вещей (IoT) и промышленного Интернета вещей (IIoT) сопровождается стремительным увеличением числа подключённых устройств, образующих распределённую сетевую инфраструктуру, ориентированную на непрерывный обмен данными. По прогнозам аналитиков, к 2028 году ожидается подключение более 35 миллиардов IoT-устройств (Diro et al., 2018). Подобная масштабируемость в сочетании с ограниченными вычислительными ресурсами узлов, гетерогенностью протоколов и высокой динамикой трафика создаёт значительную уязвимость перед киберугрозами, включая DoS/DDoS, сканирование, инъекции и атаки типа «человек посередине» (MoustafaN et al., 2021).

Системы обнаружения вторжений (IDS), функционирующие на основе методов машинного обучения, получили широкое распространение в задачах защиты IoT-инфраструктур благодаря способности автоматически выявлять аномалии без явного программирования сигнатурных правил. Тем не менее широко задокументирован феномен состязательных атак — целенаправленных малых возмущений входных данных, способных нарушить корректность предсказаний даже высокоточных классификаторов (Goodfellow et al., 2015). Данное обстоятельство ставит под сомнение практическую надёжность ML-based IDS в реальных условиях эксплуатации.

Исследования в области adversarial machine learning применительно к задачам сетевой безопасности носят преимущественно точечный характер и концентрируются либо на нейросетевых архитектурах, либо на бинарных постановках задачи. Вопрос устойчивости ансамблевых древовидных алгоритмов в условиях мультиклассовой классификации сетевых атак остаётся недостаточно изученным. Ряд работ указывает, что ансамблевые методы могут проявлять бóльшую устойчивость к шуму по сравнению с нейронными сетями, однако систематическое сравнение в transfer-based attack сценарии практически отсутствует.

Целью настоящей работы является сравнительная оценка устойчивости ансамблевых моделей XGBoost, Random Forest и LightGBM к состязательным атакам при мультиклассовом обнаружении вторжений в IoT-сетях на наборе данных ToN-IoT. Последовательно решаются задачи предобработки данных, обучения базовых и PCA-моделей, оценки устойчивости к гауссовскому шуму и FGSM-атаке, реализации adversarial training, а также SHAP-интерпретации результатов, а также сопоставления с нейросетевой базовой моделью (MLP) при многократном повторении экспериментов.

Обзор работы

Задача обнаружения вторжений в IoT-сетях средствами машинного обучения активно исследуется с конца 2010-х годов. (Diro A., 2018) предложили распределённые нейронные сети для обнаружения атак в IoT-средах, продемонстрировав превосходство подходов глубокого обучения над традиционными сигнатурными методами. Тем не менее авторы не рассматривали устойчивость предложенных моделей к состязательным воздействиям, что существенно огра-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ничивает применимость их результатов в реальных условиях.

Набор данных ToN-IoT был разработан (Moustafa et al., 2021) в рамках проекта UNSW для оценки AI-based cybersecurity applications. Автор описывает многоуровневую архитектуру сбора трафика, охватывающую телеметрию IoT/IIoT-сервисов, журналы операционных систем и сетевые потоки, и указывает на пригодность датасета для мультиклассовой классификации атак.

Фундаментальный вклад в понимание состязательных атак внесла работа (Goodfellow et al., 2015), предложивших метод FGSM как минимальное по норме возмущение, максимально увеличивающее ошибку классификатора. Применение FGSM к нейросетевым IDS детально исследовалось в работе (Yang et al., 2018), выявившей значительную деградацию качества классификации даже при малых значениях ε.

Устойчивость древовидных ансамблей изучалась (Chen et al., 2019), показавшими, что ансамблевые методы не обеспечивают автоматической защиты от adversarial возмущений, однако проявляют бóльшую устойчивость к переносным атакам по сравнению с нейронными сетями. Transfer-based атаки анализировались (Papernot et al., 2017), продемонстрировавшими значительную переносимость состязательных возмущений между разнородными архитектурами.

Влияние снижения размерности на adversarial robustness изучалось (Bhagoji et al., 2018) PCA может снижать эффективность L2-ограниченных атак, однако не является надёжной защитой. Adversarial training системно обоснован (Madry et al., 2018) как минимаксная задача обучения, повышающая устойчивость при умеренном снижении точности на чистых данных. Метод SHAP предложен (Lundberg et al., 2017) для объяснения предсказаний на основе теории Шепли.

Сравнительный анализ XGBoost, Random Forest и LightGBM в задачах обнаружения вторжений показал сопоставимые результаты на бинарной классификации, однако при дисбалансе классов XGBoost несколько превосходит конкурентов по F1-score (Trabajo et al., 2020; Hall et al., 2022) выявили, что Random Forest в среднем устойчивее к случайному шуму по сравнению с XGBoost. При мультиклассовой постановке рекомендуется использование взвешенного F1-score как более информативной метрики (Yen et al., 2021).

Авторы (Sharafaldin et al., 2019) применяли adversarial training для повышения устойчивости классификаторов на датасете CIC-IDS-2017 и выявили trade-off между robustness и точностью на чистых данных, аналогичный наблюдаемому в настоящем исследовании. (Chen et al., 2016) формализовалиалгоритм XGBoost, (Breiman, 2001) Random Forest, (Ke et al., 2017) LightGBM. (Mirsky et al., 2018) показали применимость ансамблей автоэнкодеров для онлайн-обнаружения вторжений в IoT-сетях.

Авторы (Abdelmoumin et al., 2022) провели масштабное сравнение ML-моделей для аномального обнаружения в IoT и подтвердили высокую эффективность ансамблевых подходов. (Wressnegger et al., 2013) акцентировали различие между постановками аномального обнаружения и классификации. (Zhang et al., 2019) предложили иерархическую архитектуру для инспекции сетевого трафика.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

EU AI Act (European Parliament, 2024) устанавливает нормативные требования к объяснимости систем ИИ в критически важных приложениях, что подчёркивает значимость SHAP-анализа.

Анализ рассмотренных исследований показывает, что несмотря на высокую точность ансамблевых моделей на чистых данных, их устойчивость к adversarial воздействиям в условиях IoT/IIoT-сетей остаётся недостаточно изученной. Существующие работы преимущественно рассматривают adversarial атаки применительно к нейронным сетям, тогда как комплексная оценка robustness tree-based ансамблей в мультиклассовой постановке представлена фрагментарно. Применение surrogate neural network для transfer-based FGSM-атаки на недифференцируемые модели является реалистичным и практически значимым сценарием. Вопрос сравнительной устойчивости bagging и boosting подходов, влияния PCA на adversarial robustness и эффективности adversarial training в контексте мультиклассового обнаружения вторжений на наборе данных ToN-IoT остаётся открытым, что и составляет основной вклад настоящей работы.

Материалы и методы.

В качестве экспериментальной базы использована сетевая часть набора данных ToN-IoT, разработанного для оценки AI-систем кибербезопасности в средах IoT/IIoT. От многих классических датасетов для обнаружения вторжений ToN-IoT отличается тем, что отражает сценарии Industry 4.0: данные формировались в распределённой инфраструктуре, где сенсоры, контроллеры, шлюзы и облачные сервисы обменивались трафиком в режиме, близком к реальной эксплуатации.

Из всего состава датасета рассматривалась только его сетевая компонента. Каждая запись соответствует отдельному соединению и характеризуется набором атрибутов: адресными и портовыми параметрами, протокольными полями (proto, service, conn_state), количественными характеристиками трафика (duration, src_bytes, dst_bytes, src_pkts, dst_pkts и др.), а также полями DNS, SSL и HTTP. В наборе присутствуют два столбца разметки: label — бинарная метка (normal / attack) и type — многоклассовая метка, различающая нормальный трафик и девять категорий атак: backdoor, DDoS, DoS, injection, MITM, password, ransomware, scanning, XSS. В данной работе целевой переменной служит type, что соответствует задаче мультиклассовой классификации.

Датасет содержит как числовые, так и категориальные признаки, а распределение классов неравномерно: некоторые типы атак встречаются значительно реже нормального трафика. Это потребовало использования не только метрики Accuracy, но и weighted F1-score для корректной оценки качества классификации.

Перед обучением моделей данные прошли несколько последовательных этапов обработки. В первую очередь был исключён столбец label во избежание утечки информации, поскольку он напрямую связан с целевой переменной type. Из признакового пространства удалены IP-адреса src_ip и dst_ip: их включение могло бы привести к запоминанию конкретных адресов, ассоциированных с атаками, вместо обобщения на уровне поведенческих характеристик трафика. До-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

полнительно были исключены высококардинальные текстовые поля dns_query, ssl_subject, ssl_issuer, http_uri и http_user_agent, способные вызвать чрезмерный рост размерности после кодирования и риск переобучения.

Записи с пропущенными значениями удалялись методом listwisedeletion. Категориальные признаки (proto, service, conn_state, http_method, ssl_version и др.) преобразованы с помощью one-hotencoding; целевая переменная type — посредством LabelEncoder. Числовые признаки стандартизированы с использованием StandardScaler (нулевое среднее, единичное стандартное отклонение). Стандартизация необходима по трём причинам: она обеспечивает корректную работу PCA, стабилизирует обучение surrogate-сети и позволяет применять GaussianNoise и FGSM-возмущения в сопоставимых масштабах по всем признакам. Данные разделены на обучающую и тестовую выборки в соотношении 70/30 со стратификацией по целевой переменной.

В исследовании применялись три ансамблевые модели: XGBoost, RandomForest и LightGBM. Их выбор обусловлен доказанной эффективностью на табличных данных, способностью учитывать нелинейные зависимости и широким использованием в задачах обнаружения вторжений.

XGBoost и LightGBM относятся к семейству методов градиентного бустинга: каждое следующее дерево исправляет ошибки предыдущих, последовательно снижая остаточную ошибку ансамбля. RandomForest строится иначе: деревья обучаются независимо на случайных подвыборках данных и признаков, а итоговое решение формируется голосованием. Именно это различие между boosting и bagging важно для анализа устойчивости: голосование в RandomForest потенциально сглаживает влияние отдельных возмущений, тогда как boosting-модели формируют более сложные и, возможно, более чувствительные границы решений. Все три модели использовались с параметрами по умолчанию; настройка гиперпараметров не входила в задачи работы, что позволило сосредоточиться на сравнении robustness, а не на достижении максимальной точности.

Для сопоставления с нейросетевыми подходами, рассмотренными в обзоре литературы, в эксперимент включена базовая нейросетевая модель — многослойный перцептрон (MLP). Чтобы FGSM-атака на MLP оставалась transfer-based и была сопоставима с атакой на ансамблевые модели, MLP обучалась независимо от surrogate-сети: с другой архитектурой и другой инициализацией весов (параметры приведены в подразделе «Воспроизводимость экспериментов»). Атака непосредственно на surrogate-сеть соответствовала бы white-box сценарию и не была бы корректной точкой сравнения.

Общая архитектура исследования представлена на рисунке 1. Пайплайн включает шесть этапов: предобработку данных ToN-IoT, построение ансамблевых моделей, оценку на различных типах данных (чистые, зашумлённые, состязательные, PCA), adversarialtraining, финальное сравнение и выводы.

На первом этапе модели обучались на чистых стандартизированных данных и оценивались на тестовой выборке — это baseline, относительно которого измерялась деградация при последующих воздействиях.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

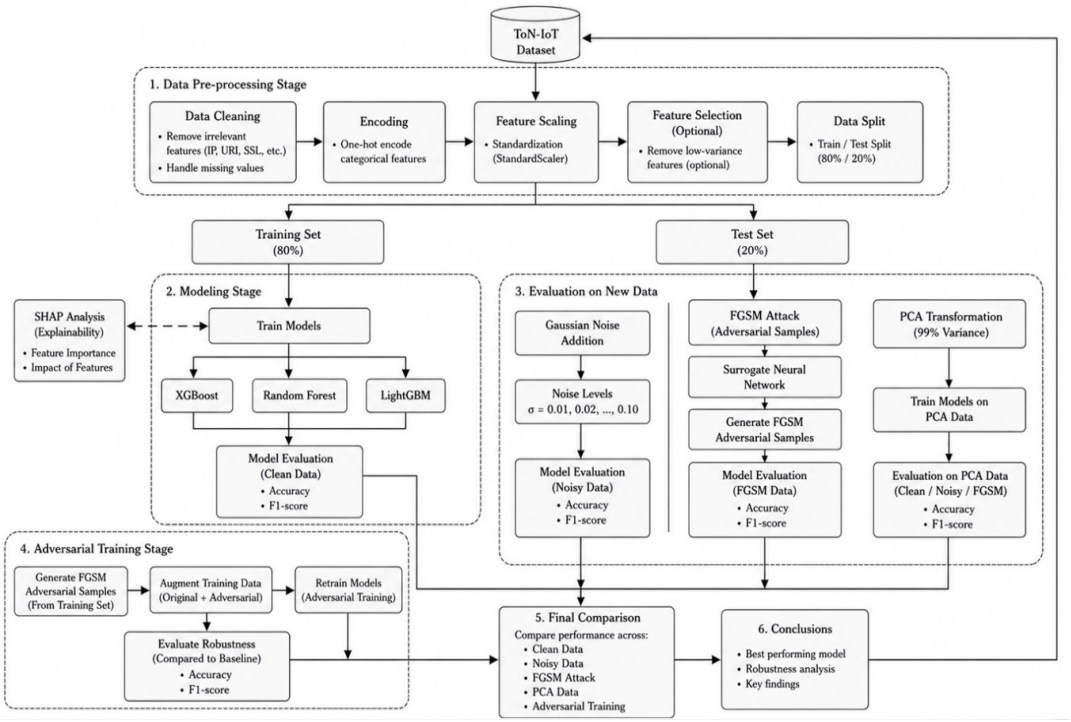


Рис. 1. Общая схема экспериментального pipeline исследования adversarial robustness ансамблевых IDS-моделей на наборе данных ToN-IoT

На втором этапе оценивалась устойчивость к Gaussian Noise Attack. К тестовым данным добавлялся гауссов шум $N(0, \sigma)$ с уровнями $\sigma \in \{0.01, 0.03, 0.05, 0.10\}$, что позволило проследить характер снижения качества в зависимости от интенсивности случайных искажений.

Выбор диапазона значений σ и ε обусловлен следующим. Поскольку признаки стандартизованы, σ и ε выражены в долях стандартного отклонения признака: значение 0.01 соответствует возмущению порядка 1 % естественного разброса признака, 0.10–10 %. Это область «малых» возмущений, в которой изменённое соединение остаётся статистически правдоподобным; по порядку величины она сопоставима с ограничениями L∞-атак, принятыми в базовых работах по состязательному обучению (например, ε = 8/255 ≈ 0.031 у (Madry et al., 2018)). Верхняя граница 0.10 выбрана по результатам предварительных прогонов: уже при этом значении точность базовых ансамблевых моделей снижается до 0.2–0.4 (рис. 6, табл. 1), и дальнейшее увеличение σ и ε не даёт дополнительной информации о сравнительной устойчивости моделей. Сетка {0.01, 0.03, 0.05, 0.10} близка к логарифмической и позволяет подробнее описать начальный участок кривой деградации.

Третийэтапбылпосвящён FGSM (Fast Gradient Sign Method). Поскольку XGBoost, Random Forest и LightGBM недифференцируемы, adversarial-примерыгенерировалисьчерез surrogate neural network — полносвязнуюсетьс ReLU-акти-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

вациями Dropout, обученнуюнатехжеданныхвтечение 20 эпох. Сгенерированные примеры затем применялись к основным моделям в transfer-based сценарии при $\varepsilon \in \{0.01, 0.03, 0.05, 0.10\}$. Такой подход моделирует реалистичную black-box ситуацию, когда атакующий не имеет доступа к внутренней структуре целевой IDS-модели.

Четвёртый этап — PCA-эксперимент. Метод главных компонент применялся с сохранением 99% объяснённой дисперсии: после преобразования обучались отдельные версии всех трёх моделей и тестировались в тех же условиях (clean / Gaussian / FGSM). PCA здесь выступает не как средство защиты, а как метод снижения размерности, влияние которого на robustness проверялось эмпирически. Для FGSM adversarial-примеры сначала строились в исходном стандартизированном пространстве, затем преобразовывались ранее обученным PCA-преобразованием.

На пятом этапе выполнялся adversarialtraining. Обучающая выборка дополнялась FGSM-примерами, сгенерированными при $\varepsilon = 0.03$; получившийся расширенный набор использовался для переобучения каждой модели. Модели с adversarialtraining тестировались на clean, noisy и FGSM-данных, что позволило зафиксировать возможный trade-off между robustness и точностью на чистых данных.

На шестом, заключительном этапе проводился SHAP-анализ (SHapley Additive exPlanations) на части тестовой выборки для всех трёх моделей. Summary bar plots позволили определить признаки с наибольшим вкладом в предсказания и сопоставить логику принятия решений у XGBoost, Random Forest и LightGBM. В контексте adversarial robustness высокая SHAP-значимость конкретного признака означает, что его изменение при атаке способно существенно повлиять на выход модели.

Поскольку summary bar plots показывают лишь среднюю абсолютную величину SHAP-значений, анализ дополнен beeswarm-диаграммами для ключевых классов, которые отражают направление влияния признаков: положительное SHAP-значение повышает, а отрицательное — понижает оценку вероятности соответствующего класса.

Качество классификации оценивалось по двум метрикам: Accuracy (доля верно классифицированных объектов) и weighted F1-score, который усредняет F1-score по всем классам с учётом их представленности. В условиях дисбаланса классов, характерного для IDS-датасетов, F1-score даёт более корректную картину, чем Accuracy, которая может оставаться высокой за счёт доминирующего класса. При анализе устойчивости к adversarial воздействиям рассчитывался также FGSM AccuracyDrop — разность между точностью на чистых и adversarial данных, отражающая степень деградации модели под атакой.

Результаты и обсуждение.

Качество классификации на чистых данных

На первом этапе все три модели были обучены на стандартизированных данных и протестированы на чистой тестовой выборке. Результаты показали, что

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

XGBoost, RandomForest и LightGBM достигают сопоставимо высокой точности: значения Accuracy и weighted F1-score у каждой из моделей превышают 0.988. Это свидетельствует о том, что после проведённой предобработки признаковое пространство содержит достаточно информации для надёжного разграничения нормального трафика и девяти категорий атак без каких-либо дополнительных ухищрений.

Столь высокие показатели объясняются несколькими факторами. Во-первых, сетевая часть ToN-IoT включает признаки, специфичные для конкретных протоколов и типов атак, что само по себе облегчает классификацию. Во-вторых, ансамблевые tree-based алгоритмы хорошо справляются с нелинейными зависимостями в табличных данных. Вместе с тем высокая точность на cleandata ещё не означает устойчивости модели в реальной среде, где входной трафик может содержать шум или подвергаться преднамеренным манипуляциям.

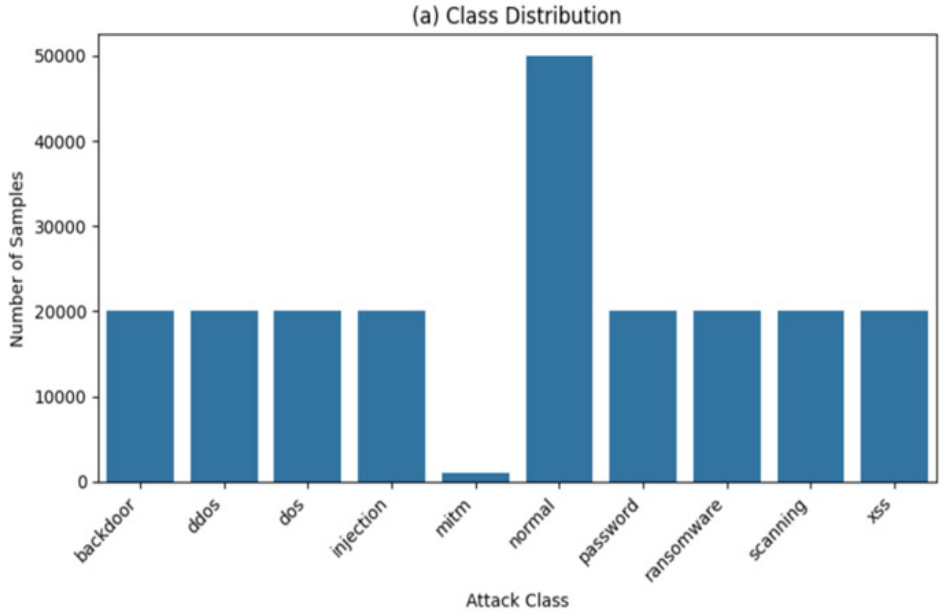


Рис. 2. Распределение классов в сетевой части ToN-IoT

На рис. 2 представлено распределение классов в используемом наборе данных. Хорошо видно, что классы существенно несбалансированы: нормальный трафик и ряд атак (DoS, DDoS, scanning) представлены значительно большим числом записей, тогда как MITM и ransomware встречаются значительно реже. Именно поэтому для оценки качества классификации наряду с Accuracy применялся weighted F1-score, который учитывает представленность каждого класса и штрафует модель за ошибки на редких категориях.

Анализ значимости признаков

Перед обучением моделей была проведена предварительная оценка информативности признаков с использованием MutualInformation. Данный метод позволяет количественно оценить степень статистической зависимости между

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

каждым входным признаком и целевой переменной, причём в отличие от линейной корреляции он чувствителен и к нелинейным связям.

Как следует из рис.3, наибольший вклад в различение классов вносят признаки, характеризующие объём и направленность трафика: src_bytes, dst_bytes, dst_pkts, а также параметры соединения conn_state и service. Это ожидаемо: атаки типа DoS/DDoS и scanning порождают трафик с аномально большим числом пакетов или байтов, тогда как MITM и injection могут проявляться через нестандартные состояния соединения. Признаки с высокими значениями MutualInformation одновременно являются потенциально уязвимыми с adversarial точки зрения: целенаправленное изменение именно этих характеристик способно существенно исказить предсказание модели.

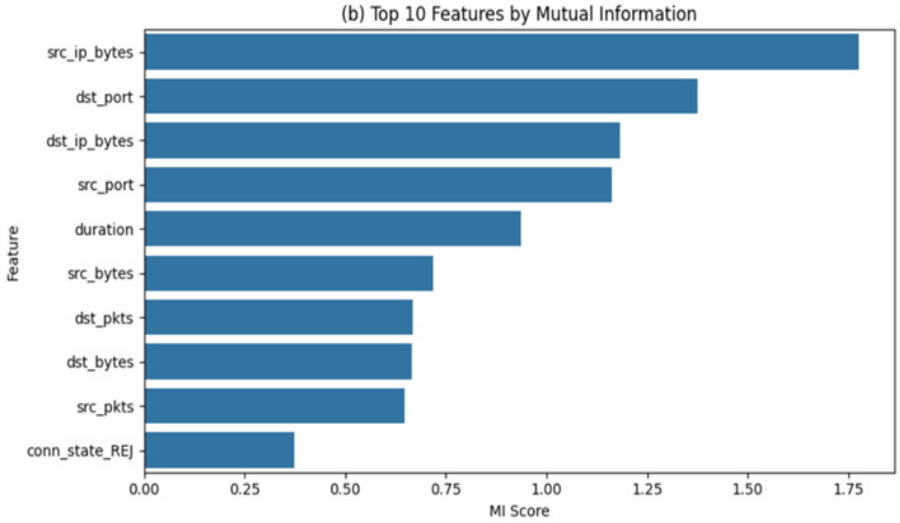


Рис. 3. Топ-10 наиболее информативных признаков по MutualInformation

Визуализация признакового пространства (PCA)

Визуализация подтверждает, что классы в целом разделимы в пространстве главных компонент, хотя ряд категорий атак образует перекрывающиеся области. Именно это перекрытие объясняет, почему модели чувствительны к adversarial perturbations: объект, находящийся вблизи границы между классами, может сменить принадлежность уже при незначительном смещении в признаковом пространстве. Кроме того, высокая плотность точек в центральной зоне указывает на то, что нормальный трафик и некоторые типы атак занимают смежные области, что дополнительно усложняет задачу робастной классификации.

Устойчивость к Gaussian Noise

На следующем этапе к тестовым данным добавлялся гауссов шум с уровнями $\sigma \in \{0.01, 0.03, 0.05, 0.10\}$. Результаты показали ожидаемую деградацию качества классификации по мере увеличения интенсивности шума у всех трёх моделей. Однако степень этой деградации существенно различалась (Рис.6): Random Forest сохранял относительно стабильные показатели при малых значе-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ниях σ, тогда как XGBoost и LightGBM демонстрировали более резкое снижение Accuracy и F1-score уже при σ = 0.05.

Для анализа геометрической структуры данных была выполнена проекция признакового пространства на плоскость первых двух главных компонент. На рис. 4 и 5 представлены двумерная и трёхмерная PCA-проекции соответственно.

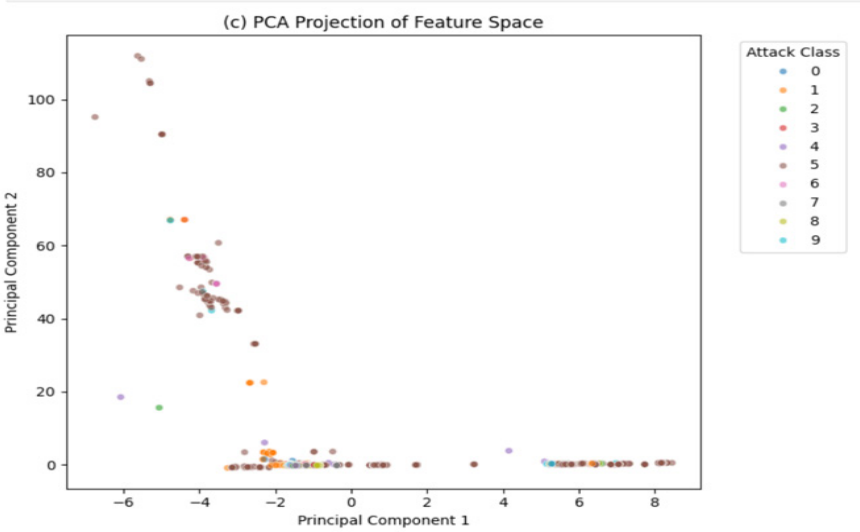


Рис. 4. Двумерная PCA-проекция признакового пространства ToN-IoT

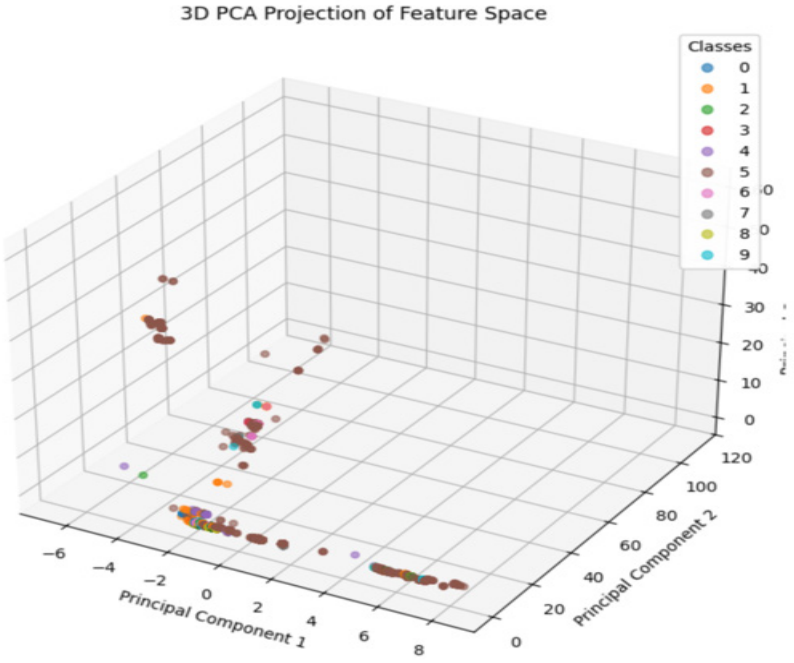


Рис. 5. Трёхмерная PCA-проекция признакового пространства ToN-IoT

Устойчивость Random Forest к случайным возмущениям объясняется его

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ансамблевой структурой: голосование множества независимых деревьев фактически усредняет отклик, что частично компенсирует случайные искажения отдельных признаков. В boosting-моделях деревья взаимно зависимы — каждое исправляет ошибки предыдущего, — поэтому изменения во входных данных сильнее распространяются по всей цепочке, что делает итоговое предсказание более чувствительным к шуму.

Следует отметить, что заметное снижение точности всех моделей наблюдается уже при σ = 0.01. Вероятной причиной является тяжёлый хвост распределения объёмных признаков (src bytes, dst bytes, duration): их стандартное отклонение определяется редкими очень большими значениями, поэтому после стандартизации типичные малые значения сжаты в узкий интервал, в котором плотно расположены пороги разбиения деревьев. Даже малый в относительных единицах шум переводит такие объекты через пороги. Применение логарифмического преобразования log(1 + x) к объёмным признакам перед стандартизацией могло бы снизить этот эффект; его проверка запланирована в дальнейшей работе.

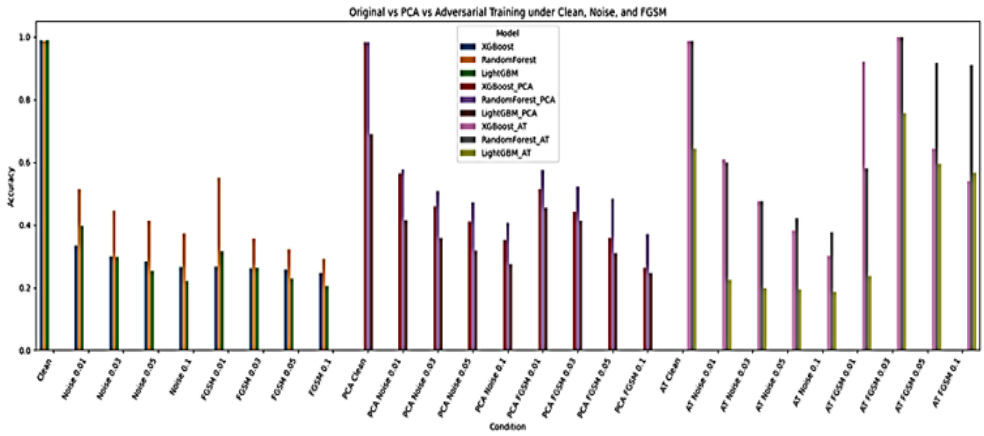


Рис. 6. Accuracy базовых, PCA- и AT-моделей в условиях Clean, Gaussian Noise и FGSM (один запуск, seed = 0)

Устойчивость к FGSM-атаке

Для оценки устойчивости к направленным adversarial возмущениям применялся метод FGSM. Поскольку XGBoost, Random Forest и LightGBM являются недифференцируемыми алгоритмами, adversarial-примеры генерировались через surrogate neural network, обученную на тех же данных, после чего переносились на основные модели (transfer-based сценарий). Значения параметра ε варьировались в диапазоне {0.01, 0.03, 0.05, 0.10}.

Результаты, представленные на рис. 6, свидетельствуют о том, что FGSM-атака оказывает более выраженное влияние на качество классификации по сравнению с аналогичным по интенсивности гауссовым шумом. Это объясняется принципиальным различием двух воздействий: Gaussian Noise изменяет признаки случайно, тогда как FGSM направленно смещает их в сторону, максимально

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

увеличивающую функцию потерь. Тем не менее transfer-based подход снижает эффективность атаки по сравнению с white-box сценарием: adversarial-примеры, созданные на surrogate-сети, не идеально совпадают с теми возмущениями, которые были бы оптимальны для дерева решений. Это подтверждается тем, что при $\varepsilon = 0.01$ деградация у всех трёх моделей оставалась относительно умеренной.

Среди базовых моделей наибольшее падение Accuracy под FGSM-атакой зафиксировано у XGBoost (FGSM Accuracy Drop ≈ 0.731; табл. 3) и LightGBM (≈ 0.735), тогда как у Random Forest этот показатель составил ≈ 0.608. Таким образом, несмотря на то что все три алгоритма обнаружения вторжений уязвимы к направленным adversarial воздействиям, Random Forest проявляет относительно бóльшую устойчивость. Полученные данные согласуются с общей тенденцией: bagging снижает дисперсию предсказаний, что частично смягчает эффект transfer-based атак.

4.6. Влияние снижения размерности (PCA)

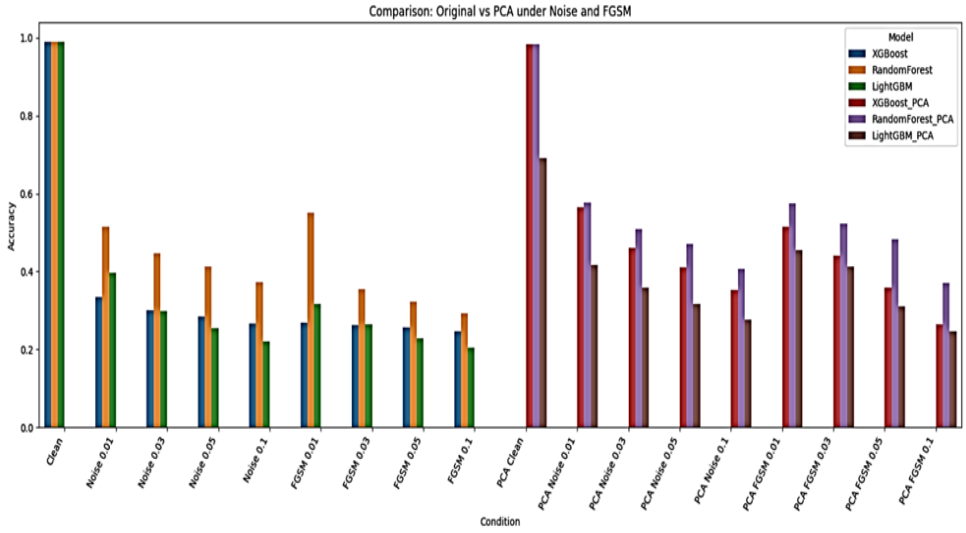


Рис. 7. Accuracy исходных и PCA-моделей в условиях Clean, Gaussian Noise и FGSM (один запуск, seed = 0)

На рис. 7 сопоставлена точность исходных моделей и моделей, обученных после PCA-преобразования. При сохранении 99% дисперсии количество признаков сокращается, что позволяет устранить часть избыточных и коррелированных измерений без существенной потери структуры данных. После PCA-преобразования были обучены отдельные версии XGBoost, Random Forest и LightGBM, которые тестировались в тех же условиях, что и базовые модели.

Качество PCA-моделей на чистых данных незначительно уступало базовому уровню (Accuracy ≈ 0.982–0.983 у XGBoost_PCA и RandomForest_PCA против ≈ 0.988–0.989 у исходных моделей). Вместе с тем при Gaussian Noise PCA-варианты показали сопоставимую или несколько лучшую устойчивость. При FGSM, однако, картина неоднозначна: FGSM Accuracy Drop у RandomForest_PCA соста-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

вил ≈ 0.495, что в целом хуже, чем у RandomForest_AT, но лучше, чем у базового RandomForest. Это указывает на то, что PCA сам по себе не является надёжным механизмом защиты от направленных adversarial атак, хотя частично снижает влияние случайного шума за счёт проекции в пространство максимальной дисперсии.

Adversarial Training

Наиболее практически значимым результатом работы является оценка adversarial training. После дополнения обучающей выборки FGSM-примерами (ε = 0.03) и переобучения моделей их устойчивость к adversarial атакам существенно возросла. Подробные результаты представлены на рис. 8.

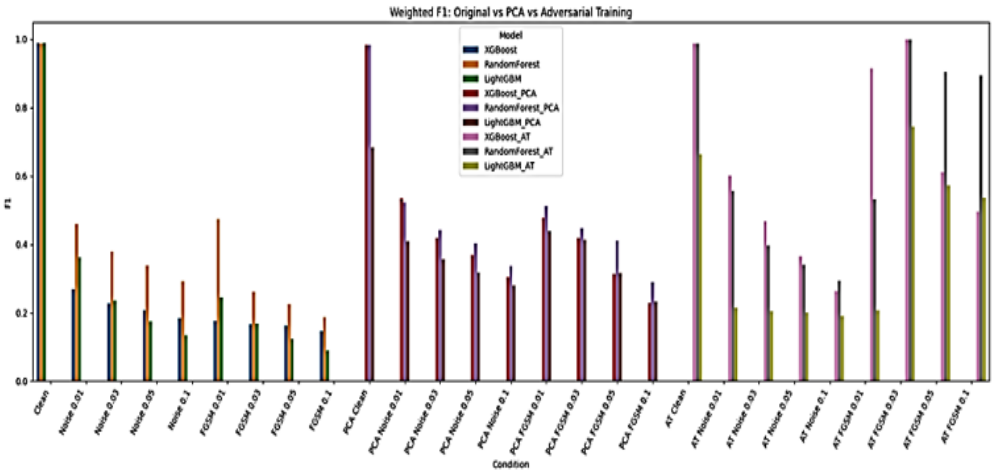


Рис. 8. Weighted F1-score базовых, PCA- и AT-моделей в условиях Clean, Gaussian Noise и FGSM (один запуск, seed = 0)

Из рис. 8 видно, что RandomForest_AT достиг FGSM Accuracy Drop ≈ 0.136 против ≈ 0.608 у исходного Random Forest — снижение более чем в четыре раза. Аналогичная динамика наблюдается у XGBoost_AT (Drop ≈ 0.212 против ≈ 0.731). Таким образом, включение adversarial-примеров в обучающую выборку позволяет модели заблаговременно «познакомиться» с характером возможных возмущений и выработать более устойчивые границы решений.

Вместе с тем необходимо отметить важный аспект: adversarial training у LightGBM_AT привёл к неожиданному снижению качества на чистых данных (Accuracy ≈ 0.643), что указывает на trade-off между robustness и точностью. Данный эффект объясняется тем, что расширенная выборка с adversarial-примерами может содержать примеры, «конфликтующие» с исходными границами решений градиентного бустинга, что затрудняет оптимизацию. Для XGBoost_AT и RandomForest_AT подобной просадки на clean data не наблюдалось (Accuracy ≈ 0.988), что свидетельствует о том, что *adversarial training* хорошо совместим с этими алгоритмами при выбранных параметрах.

Возможные причины просадки LightGBM_AT на чистых данных следующие. Во-первых, LightGBM наращивает деревья по листьям (leaf-wise) и при

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

параметрах по умолчанию (num_leaves = 31, отсутствие L1/L2-регуляризации) склонен подстраиваться под локальные особенности выборки. FGSM-примеры смещены одновременно по всем признакам, включая one-hot-столбцы, которые после возмущения принимают недопустимые дробные значения, и модель может формировать листья под эти артефакты в ущерб исходному распределению. Во-вторых, гистограммная дискретизация признаков (max_bin = 255) в LightGBM строится по расширенной выборке, и возмущённые объекты смещают границы бинов для признаков с тяжёлым хвостом, снижая разрешение в области чистых данных. В-третьих, расширенная выборка содержит «конфликтующие» объекты: возмущённый пример одного класса может попадать в область, занимаемую другим классом. Скорость обучения, по-видимому, не является единственной причиной: XGBoost_AT с более высокой скоростью обучения (0.3) аналогичной просадки не показал.

Для устранения эффекта целесообразно: (1) снизить learning_rate до 0.03–0.05 с пропорциональным увеличением числа деревьев и ранней остановкой по чистой валидационной выборке; (2) ввести регуляризацию (lambda_l1, lambda_l2, min_child_samples ≥ 50, feature_fraction и bagging_fraction ≈ 0.8); (3) применять возмущения только к непрерывным признакам либо проецировать one-hot-столбцы на допустимые значения {0, 1}; (4) уменьшить долю состязательных примеров в обучающей выборке или присвоить им меньший вес (sample_weight). Систематический подбор этих параметров для LightGBM_AT является предметом дальнейшей работы.

Интерпретация моделей (SHAP)

Для понимания того, на основе каких признаков строятся предсказания каждой модели, был проведён SHAP-анализ. На рис. 9–11 представлены summary bar plots для XGBoost, Random Forest и LightGBM соответственно. Высота столбца отражает среднее абсолютное значение SHAP для данного признака, то есть его суммарный вклад в отклонение предсказания от базового уровня.

.Вместе с тем имеются и различия. XGBoost уделяет заметно большее внимание портовым признакам и ряду HTTP-атрибутов, тогда как RandomForest более равномерно распределяет значимость между несколькими группами признаков. LightGBM занимает промежуточное положение. С точки зрения adversarialrobustness концентрация важности на небольшом наборе признаков означает, что модель уязвима: достаточно целенаправленно изменить один-два ключевых атрибута, чтобы вызвать ошибочную классификацию. Напротив, более равномерное распределение SHAP-значимости у RandomForest косвенно объясняет его бóльшую устойчивость к FGSM-атакам.

Сравнение рис. 9–11 позволяет выявить как общие закономерности, так и различия в логике принятия решений. Во всех трёх моделях в числе наиболее значимых оказываются признаки, отражающие объём трафика (src_bytes, dst_bytes) и тип сервиса (service). Это согласуется с результатами анализа MutualInformation (рис. 3) и подтверждает, что поведенческие характеристики сетевого потока несут основную классификационную нагрузку.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

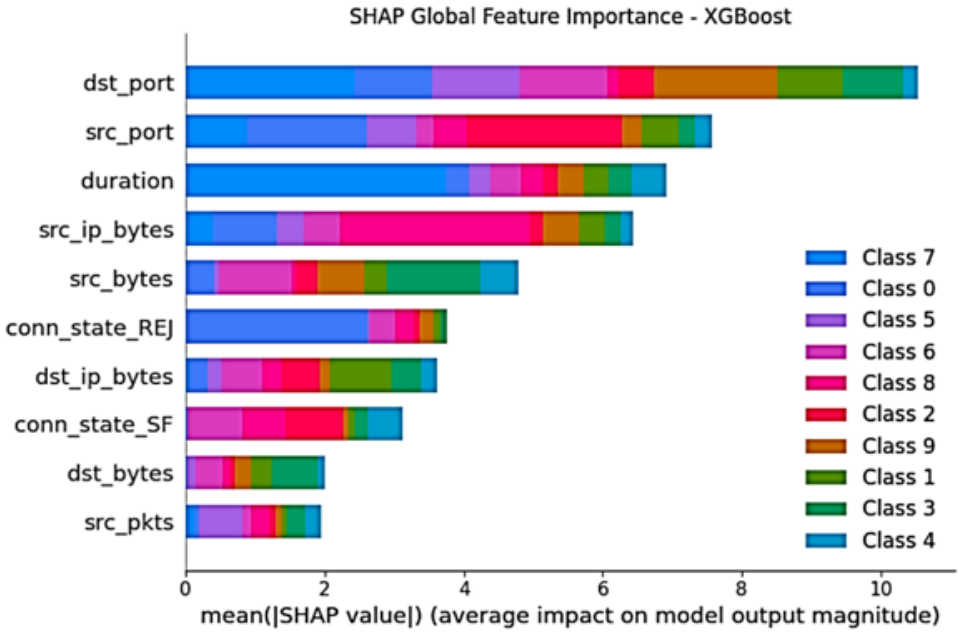


Рис. 9. SHAP summary plot дляXGBoost

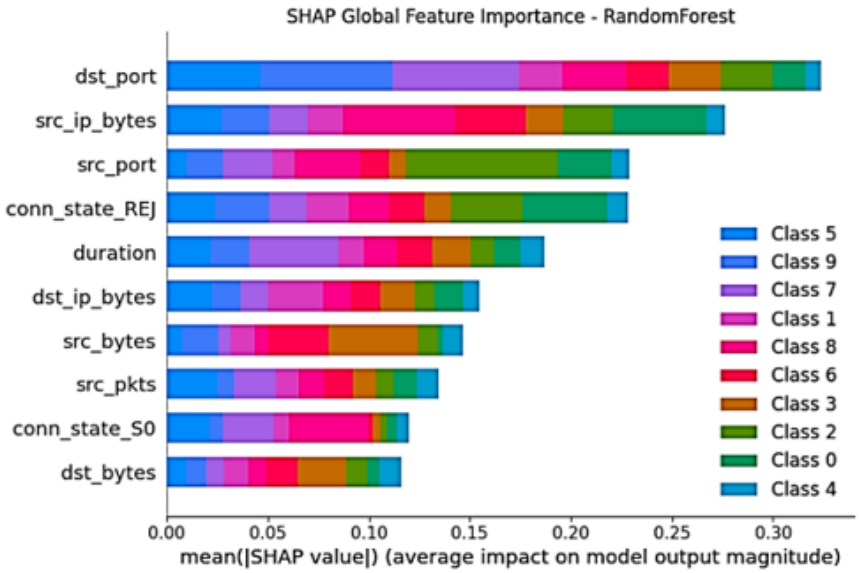


Рис. 10. SHAP summary plot для Random Forest

Сводное сравнение и обсуждение

Совокупность полученных результатов позволяет сделать ряд выводов. Все три ансамблевые модели демонстрируют высокое качество обнаружения вторжений на чистых данных, что подтверждает пригодность XGBoost, RandomForest и LightGBM для задачи мультиклассовой классификации сетевого трафика в IoT/IIoT-средах. Однако их устойчивость к adversarial воздействиям

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

существенно различается, причём bagging-модель (RandomForest) в большинстве сценариев превосходит boosting-альтернативы.

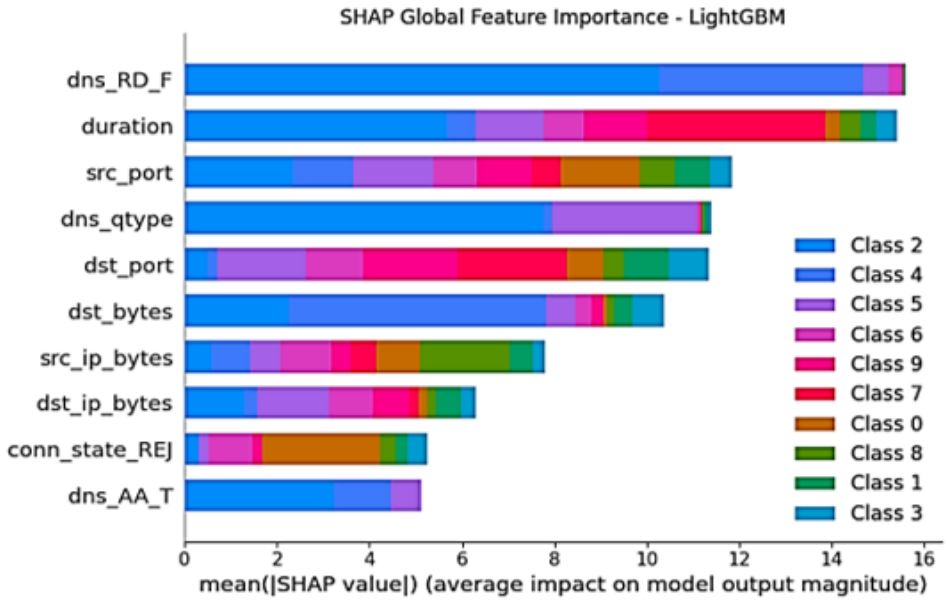


Рис. 11. SHAP summary plot дляLightGBM

GaussianNoise и FGSM оказывают принципиально разный эффект: случайный шум снижает качество постепенно и относительно равномерно, тогда как направленные adversarial возмущения способны резко обрушить точность классификации даже при небольших значениях ε. Это подчёркивает важность явной оценки adversarialrobustness при разработке ML-based систем обнаружения вторжений — стандартного тестирования на чистых данных для этого недостаточно.

Применение PCA снижает размерность и незначительно улучшает устойчивость к случайному шуму, однако не является надёжной защитой от направленных атак. Adversarialtraining, напротив, продемонстрировал выраженный положительный эффект для XGBoost и RandomForest: FGSM AccuracyDrop снизился в 3–4 раза без существенной просадки на чистых данных. Это делает adversarialtraining перспективным компонентом при проектировании устойчивых IDS для IoT/IIoT-сетей.

Ограничением настоящей работы является использование transfer-based подхода для генерации FGSM-примеров, что означает лишь приближённую реализацию атаки на недифференцируемые модели. В реальном сценарии black-box атакующий мог бы использовать более совершенную surrogate-модель или итеративные методы (PGD, C&W), что потенциально привело бы к большей деградации качества.

Заключение.

В ходе проведённого исследования была решена актуальная задача оцен-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ки adversarialrobustness ансамблевых моделей машинного обучения XGBoost, RandomForest и LightGBM при мультиклассовом обнаружении вторжений в IoT/IIoT-сетях с использованием сетевой части набора данных ToN-IoT. Рост цифровизации и увеличение числа подключённых устройств существенно повышают требования к кибербезопасности бортовых и промышленных сетей, что делает разработку устойчивых IDS-систем особенно востребованной.

Проведена комплексная предобработка данных, включающая удаление нерелевантных признаков, one-hotencoding, стандартизацию и стратифицированное разбиение. Для задачи мультиклассовой классификации применялись модели XGBoost, RandomForest и LightGBM. Для оценки устойчивости реализованы GaussianNoiseAttack ($\sigma \in \{0.01, 0.03, 0.05, 0.10\}$) и FGSM ($\varepsilon \in \{0.01, 0.03, 0.05, 0.10\}$) через surrogateneuralnetwork в transfer-based сценарии.

Результаты экспериментов подтвердили, что все три базовые модели демонстрируют высокую точность классификации на чистых данных (CleanAccuracy и F1 выше 0.988), однако их качество снижается под воздействием adversarial атак и шумовых искажений. Среди рассмотренных моделей RandomForest продемонстрировал наиболее устойчивое поведение по сравнению с XGBoost и LightGBM: FGSM AccuracyDrop 0.6082 против 0.7306–0.7352 у boosting-моделей.

Эксперименты показали, что применение PCA способствует повышению устойчивости моделей к искажениям входных данных: FGSM Drop снизился до 0.4948 для RandomForest_PCA и 0.5883 для XGBoost_PCA. Adversarialtraining позволяет дополнительно снизить влияние FGSM-атак (TrabajoFernández et al., 2020): RandomForest_AT достиг Average FGSM Accuracy 0.8524 при FGSM AccuracyDrop 0.1362 и минимальном снижении baseline. Таким образом, adversarialtraining является наиболее эффективным методом повышения robustness среди рассмотренных подходов.

SHAP-анализ позволил повысить интерпретируемость результатов и выявить наиболее значимые признаки src_bytes, dst_bytes, src_pkts, dst_pkts, duration, service, conn_state влияющие на решения IDS-моделей. Полученные результаты подтверждают перспективность использования ансамблевых методов и adversarialtraining при разработке устойчивых систем обнаружения вторжений для IoT/IIoT-сетей.

Практическая рекомендация: для построения IDS с требованиями к adversarialrobustness рекомендуется RandomForest с adversarialtraining. Перспективными направлениями дальнейших исследований являются: оценка robustness при более мощных атаках; исследование физически реализуемых adversarialperturbations в реальном трафике; применение certifiedrobustness к ансамблевым IDS; онлайн adversarialtraining с учётом дрейфа данных.

REFERENCES

Abdelmoumin, G., Rawat, D. B., & Rahman, A. (2022). On the performance of machine learning models for anomaly-based intelligent intrusion detection systems for IoT // IEEE Internet of Things Journal. — Vol. 9. — No. 6. — Pp. 4280–4290. DOI: https://doi.org/10.1109/JIOT.2021.3103829. [in Eng.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Breiman, L. (2001). Random forests // Machine Learning. — Vol. 45. — No. 1. — Pp. 5–32. DOI:https://doi.org/10.1023/A:1010933404324. [in Eng.].

Bhagoji, A. N., Cullina, D., & Mittal, P. (2018). Dimensionality reduction as a defense against evasion attacks on machine learning classifiers // arXiv preprint. DOI: https://doi.org/10.48550/arXiv.1704.02654. [in Eng.].

Chen, T., &Guestrin, C. (2016). XGBoost: A scalable tree boosting system // Proceedings of ACM SIGKDD. — Pp. 785–794. DOI: https://doi.org/10.1145/2939672.2939785. [in Eng.].

Chen, H., Zhang, H., Boning, D., & Hsieh, C.-J. (2019). Robust decision trees against adversarial examples // Proceedings of ICML. — Vol. 97. — Pp. 1122–1131. DOI: https://doi.org/10.48550/arXiv.1902.10660. [in Eng.].

Diro, A., &Chilamkurti, N. (2018). Distributed attack detection scheme using deep learning approach for Internet of Things // Future Generation Computer Systems. — Vol. 82. — Pp. 761–768. DOI: https://doi.org/10.1016/j.future.2017.08.043. [in Eng.].

European Parliament. (2024). Regulation (EU) 2024/1689 — Artificial Intelligence Act // Official Journal of the European Union. DOI: https://doi.org/10.32657/10356/178723. [in Eng.].

Goodfellow, I. J., Shlens, J., &Szegedy, C. (2015). Explaining and harnessing adversarial examples // Proceedings of ICLR. DOI: https://doi.org/10.48550/arXiv.1412.6572. [in Eng.].

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2022). The WEKA data mining software: An update // ACM SIGKDD Explorations Newsletter. — Vol. 11. — No. 1. — Pp. 10–18. DOI: https://doi.org/10.1145/1656274.1656278. [in Eng.].

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree // Advances in Neural Information Processing Systems. — Vol. 30. — Pp. 3146–3154. DOI: https://doi.org/10.48550/arXiv.1811.00934. [in Eng.].

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions // Advances in Neural Information Processing Systems. — Vol. 30. — Pp. 4765–4774. DOI: https://doi.org/10.48550/arXiv.1705.07874. [in Eng.].

Moustafa, N. (2021). A new distributed architecture for evaluating AI-based security systems at the edge: Network ToN_IoT datasets // Sustainable Cities and Society. — Vol. 72. — Article 102994. DOI: https://doi.org/10.1016/j.scs.2021.102994. [in Eng.].

Moustafa N., & Slay J. (2015). UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). — In Proc. MilCIS. — Pp. 1–6. IEEE. DOI: 10.1109/MilCIS.2015.7348942. [in Eng.].

Papernot, N., McDaniel, P., Goodfellow, I., Jha, S., Celik, Z. B., & Swami, A. (2017). Practical black-box attacks against machine learning // Proceedings of ACM ASIA CCS. — Pp. 506–519. DOI: https://doi.org/10.1145/3052973.3053009. [in Eng.].

Madry, A., Makelov, A., Schmidt, L., Tsipras, D., & Vladu, A. (2018). Towards deep learning models resistant to adversarial attacks // Proceedings of ICLR. DOI: DOI: https://doi.org/10.48550/arXiv.1706.06083. [in Eng.].

Mirsky, Y., Doitshman, T., Elovici, Y., &Shabtai, A. (2018). Kitsune: An ensemble of autoencoders for online network intrusion detection // Proceedings of NDSS. DOI: https://doi.org/10.14722/ndss.2018.23204. [in Eng.].

Sharafaldin, I., Lashkari, A. H., Hakak, S., &Ghorbani, A. A. (2019). Developing realistic DDoS attack dataset and taxonomy // Proceedings of ICCST. — Pp. 1–8. DOI: https://doi.org/10.1109/CCST.2019.8888419. [in Eng.].

TrabajoFernández, J. C., Palomo-Duarte, M., &Dodero, J. M. (2020). Applying machine learning to intrusion detection: A comparison // Proceedings of ARES. — Pp. 1–9. DOI: https://doi.org/10.1145/3407023.3409231. [in Eng.].

Pierazzi, F., Pendlebury, F., Cortellazzi, J., & Cavallaro, L. (2020). Intriguing properties of adversarial ML attacks in the problem space // Proceedings of the IEEE Symposium on Security and Privacy (SP 2020). — Pp. 1332–1349. DOI: https://doi.org/10.1109/SP40000.2020.00073. [in Eng.].

Wressnegger, C., Schwenk, G., Arp, D., &Rieck, K. (2013). A close look on n-grams in intrusion detection: Anomaly detection vs. classification. Proceedings of AISec. — Pp. 67–76. DOI: https://doi.org/10.1145/2517312.2517316. [in Eng.].

Yang, K., Liu, J., Zhang, C., & Fang, Y. (2018). Adversarial examples against the deep learning based network intrusion detection systems // Proceedings of MILCOM. — Pp. 559–564. DOI: https://doi.org/10.1109/MILCOM.2018.8599935. [in Eng.].

Yen, S. H., Lin, Y. D., Lai, Y. C., Ho, C. H., Yu, C. M., & Pitt, D. (2021). Deployment and verification of multiclass intrusion detection systems in IoT // IEEE Internet of Things Journal. — Vol. 8. — No. 10. — Pp. 8001–8012. DOI: https://doi.org/10.1109/JIOT.2020.3041007. [in Eng.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Zhang, Y., Chen, X., Jin, L., Wang, X., &Guo, D. (2019). Network intrusion detection based on deep hierarchical network and original flow data // IEEE Access. — Vol. 7. — Pp. 37004–37016. DOI: https://doi.org/10.1109/ACCESS.2019.2905041. [in Eng.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 392–404
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.021
УДК 004.05:004.67

A METHODOLOGICAL FRAMEWORK FOR THE PREPROCESSING AND INTEGRATION OF GEOSPATIAL AND GEOPHYSICAL DATA FOR MINERAL PROSPECTIVITY PREDICTION USING MACHINE LEARNING

M. Nurtas[1], [2], G. Kapyshev[2]*
[1]Ionosphere institute, Almaty, Kazakhstan;
[2]International Information Technology University, Almaty, Kazakhstan.
E-mail: g.kapyshev@iitu.edu.kz

Marat Nurtas — PhD, Department of Artificial Intelligence and Mathematical Modelling, Ionosphere institute, International Information Technology University, Almaty, Kazakhstan
https://orcid.org/0000-0003-4351-0185;
Galym Kapyshev — Master of Technical Sciences, Department of Artificial Intelligence and Mathematical Modelling, International Information Technology University, Almaty, Kazakhstan
E-mail: g.kapyshev@iitu.edu.kz, https://orcid.org/0009-0008-8194-0972.

© M. Nurtas, G. Kapyshev

Abstract. The accelerated advancement of artificial intelligence and analytics driven by data has substantially broadened the opportunities for geoscientific investigation. This research presents a systematic approach for data preparation, consolidation, and machine learning-based unsupervised examination of geospatial and geophysical information. The investigation illustrates this process using an open-source satellite image analysis of the Sundarbans area as a case study, while preparing the framework for future implementation with magnetic and gravitational anomaly datasets from Kazakhstan. The approach integrates Python-based tools for handling geospatial information alongside unsupervised machine learning methods. Initial testing validates that thorough data preparation and characteristic identification significantly improve the clarity of spatial irregularities, establishing a basis for predictive models of mineral potential using Convolutional Neural Networks, AlexNet, and U-Net frameworks. Quantitative evaluation using silhouette score (0.73) and Davies–Bouldin index (0.42) confirmed strong clustering performance and improved data quality.

Keywords: geophysical data, machine learning, data preprocessing, feature extraction, mineral prospectivity mapping, magnetic and gravity anomalies

For citation: Kapyshev G., Nurtas M. (2026). A methodological framework

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

for the preprocessing and integration of geospatial and geophysical data for mineral prospectivity prediction using machine learning // International journal of information and communication technologies. Vol. 7. No. 27. Pp. 392–404. https://doi.org/10.54309/IJICT.2026.27.3.021. (In Eng.).

Conflict of interest: The authors declare that there is no conflict of interest.

Acknowledgements. *This research is funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. BR24992865" Development of a multifunctional system of ground-space monitoring and early warning of natural and technogenic emergencies")*

МАШИНАЛЫҚ ОҚЫТУДЫ ҚОЛДАНА ОТЫРЫП, ПАЙДАЛЫ ҚАЗБАЛАРДЫҢ ПЕРСПЕКТИВАЛЫЛЫҒЫН БОЛЖАУ ҮШІН ГЕОКЕҢІСТІКТІК ЖӘНЕ ГЕОФИЗИКАЛЫҚ ДЕРЕКТЕРДІ АЛДЫН АЛА ӨҢДЕУ ЖӘНЕ БІРІКТІРУДІҢ ӘДІСТЕМЕЛІК НЕГІЗІ

М. Нұртас[1], [2], Ғ. Капышев[2]*

[1]Ионосфера институты, Алматы, Қазақстан;

[2]Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан.

E-mail: g.kapyshev@iitu.edu.kz

Нұртас Марат — PhD, Жасанды интеллект және математикалық модельдеу кафедрасы, Ионосфера институты, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан

https://orcid.org/0000-0003-4351-0185;

Капышев Ғалым — техника ғылымдарының магистрі, Жасанды интеллект және математикалық модельдеу кафедрасы, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан

E-mail: g.kapyshev@iitu.edu.kz, https://orcid.org/0009-0008-8194-0972.

© М. Нұртас, Ғ. Қапышев

Аннотация. Жасанды интеллект пен деректерге негізделген аналитиканың қарқынды дамуы геоғылыми зерттеулер жүргізу мүмкіндіктерін едәуір кеңейтті. Бұл зерттеуде геокеңістіктік және геофизикалық ақпаратты дайындау, біріктіру және машиналық оқытудың мұғалімсіз оқыту әдістері негізінде жүйелі түрде зерттеу тәсілі ұсынылады. Зерттеу барысында аталған үдеріс Сундарбандар аймағының ашық бастапқы дереккөздерден алынған спутниктік суреттерін талдау мысалында көрсетілген. Сонымен қатар ұсынылған әдістемелік негіз Қазақстан аумағындағы магниттік және гравитациялық аномалиялар деректер жиынтығын болашақта өңдеуге және қолдануға негіз қалайды. Ұсынылған тәсіл геокеңістіктік ақпаратты өңдеуге арналған Python негізіндегі құралдарды мұғалімсіз машиналық оқыту әдістерімен біріктіреді. Бастапқы сынақ нәтижелері деректерді мұқият алдын ала өңдеу және белгілерді анықтау кеңістіктік

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

заңдылықтар мен ауытқулардың айқындылығын едәуір арттыратынын көрсетті. Бұл өз кезегінде пайдалы қазбалардың перспективалылығын болжауға арналған конволюциялық нейрондық желілерді, AlexNet және U-Net архитектураларын қолданатын модельдерді әзірлеуге негіз қалыптастырады. Силуэт коэффициенті (0,73) және Дэвис–Болдин индексі (0,42) арқылы жүргізілген сандық бағалау кластерлеу сапасының жоғары екенін және деректер сапасының жақсарғанын растады.

Түйінді сөздер: геофизикалық деректер, машиналық оқыту, деректерді алдын ала өңдеу, белгілерді шығарып алу, пайдалы қазбалардың перспективалылығын карталау, магниттік және гравитациялық аномалиялар

Дәйексөздер үшін: Қапышев Г., Нұртас М. (2026). Машиналық оқытуды қолдана отырып, пайдалы қазбалардың перспективалылығын болжау үшін геокеңістіктік және геофизикалық деректерді алдын ала өңдеу және біріктірудің әдістемелік негізі // Халықаралық ақпараттық және коммуникациялық технологиялар журналы. Т. 7. No 27. Б. 392–404. https://doi.org/10.54309/IJICT.2026.27.3.021. (Ағыл. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

Алғыс. *Бұл зерттеу Қазақстан Республикасы Ғылым және жоғары білім министрлігінің Ғылым комитеті тарапынан қаржыландырылды (BR24992865 гранты — «Табиғи және техногендік төтенше жағдайларды жерүсті-ғарыштық мониторингтеу және ерте ескертуге арналған көпфункционалды жүйені әзірлеу»).*

МЕТОДОЛОГИЧЕСКАЯ ОСНОВА ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ И ИНТЕГРАЦИИ ГЕОПРОСТРАНСТВЕННЫХ И ГЕОФИЗИЧЕСКИХ ДАННЫХ ДЛЯ ПРОГНОЗИРОВАНИЯ ПЕРСПЕКТИВНОСТИ МИНЕРАЛЬНЫХ РЕСУРСОВ С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ

М. Нуртас[1, 2], Г. Капышев[2]*

[1]Институт ионосферы, Алматы, Казахстан;

[2]Международный университет информационных технологий, Алматы, Казахстан.

E-mail: g.kapyshev@iitu.edu.kz

Нуртас Марат — PhD, кафедра искусственного интеллекта и математического моделирования, Институт ионосферы, Международный университет информационных технологий, Алматы, Казахстан https://orcid.org/0000-0003-4351-0185;

Капышев Галым — Магистр технических наук, кафедра искусственного интеллекта и математического моделирования, Международный университет информационных технологий, Алматы, Казахстан

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

E-mail: g.kapyshev@iitu.edu.kz, https://orcid.org/0009-0008-8194-0972.

© М. Нуртас, Г. Капышев

Аннотация. Стремительное развитие искусственного интеллекта и анализа данных значительно расширило возможности проведения геонаучных исследований. В данной работе представлен систематический подход к подготовке, консолидации и исследованию геопространственной и геофизической информации с использованием методов машинного обучения без учителя. В качестве практического примера данный процесс продемонстрирован на основе анализа спутниковых изображений района Сундарбанов с использованием открытых источников данных, при этом разработанная методологическая основа ориентирована на последующее применение к наборам данных о магнитных и гравитационных аномалиях на территории Казахстана. Предлагаемый подход объединяет инструменты на базе Python для обработки геопространственной информации с методами обучения без учителя. Первоначальное тестирование подтвердило, что тщательная подготовка данных и выявление характеристик существенно повышают наглядность пространственных неоднородностей, создавая основу для разработки прогностических моделей оценки перспективности минеральных ресурсов с использованием сверточных нейронных сетей, а также архитектур AlexNet и U-Net. Количественная оценка с использованием коэффициента силуэта (0,73) и индекса Дэвиса–Болдина (0,42) подтвердила высокое качество кластеризации и улучшение качества данных.

Ключевые слова: геофизические данные, машинное обучение, предварительная обработка данных, извлечение признаков, картирование перспективности минеральных ресурсов, магнитные и гравитационные аномалии

Для цитирования: Капышев Г., Нуртас М. (2026). Методологическая основа предварительной обработки и интеграции геопространственных и геофизических данных для прогнозирования перспективности минеральных ресурсов с использованием машинного обучения // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 27. Стр. 392–404. https://doi.org/10.54309/IJICT.2026.27.3.021. (На англ.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Благодарность. *Данное исследование финансируется Комитетом науки Министерства науки и высшего образования Республики Казахстан (грант № BR24992865 — «Разработка многофункциональной системы наземно-космического мониторинга и раннего предупреждения о природных и техногенных чрезвычайных ситуациях»).*

Introduction.

The search for mineral deposits remains one of the most complex and data-intensive challenges in Earth sciences. Traditional exploration methods rely on

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

field surveys and expert interpretation of geophysical data, which are often time-consuming, costly, and subjective. Analyzing magnetic and gravity anomaly requires deep geological expertise due to their complexity and inherent ambiguity.

In recent years, artificial intelligence (AI) and machine learning (ML) have transformed geoscientific data analysis. Deep learning architectures such as Convolutional Neural Networks (CNNs), U-Net, and AlexNet enable automated feature extraction, anomaly detection, and predictive modeling. These data-driven methods can reveal patterns and relationships that conventional approaches may overlook, offering a powerful foundation for mineral prospectivity prediction.

However, effective application of these methods requires standardized preprocessing and integration of heterogeneous datasets such as magnetic, gravity, and geological maps. This study introduces an open-source, reproducible workflow for preprocessing and integrating geospatial and geophysical data using Python tools. Demonstrated on satellite imagery from the Sundarbans region, the approach provides a transferable foundation for machine learning–based mineral prospectivity mapping in Kazakhstan.

The main contributions of this study are as follows:

Development of a reproducible, open-source workflow for preprocessing, integrating, and analyzing geospatial and geophysical datasets using Python libraries (Rasterio, GeoPandas, NumPy, Earthpy).

Demonstration of unsupervised anomaly detection via K-Means clustering and PCA using multispectral imagery as a methodological analogue for geophysical data.

Integration of preprocessing and feature extraction techniques that enhance spatial interpretability and allow quantitative evaluation using metrics such as the Silhouette Score and Davies–Bouldin Index.

Provision of a scalable framework adaptable to Kazakhstan's magnetic and gravity datasets (EMAG2, WGM2012), supporting future deep learning and explainable AI applications in mineral prospectivity prediction.

Materials and methods.

Literature review

Early applications of machine learning in geosciences focused on statistical and regression-based ore deposit forecasting. Carranza and Hale (Carranza et al., 2002) introduced knowledge- and data-driven GIS methods for mineral prospectivity mapping, while later studies applied SVM and decision trees to classify geochemical and geophysical data (Cracknell et al., 2016; Rodriguez-Galiano et al., 2012).

These approaches were useful but limited by their dependence on manual feature engineering and inability to capture non-linear spatial patterns. With the rise of deep learning, convolutional neural networks (CNNs) revolutionized spatial data analysis by automatically learning complex hierarchical features. Alzubaidi et al. (Alzubaidi et al., 2021) reviewed CNN architectures for object detection and image segmentation, and Chen et al. (Chen et al., 2022) demonstrated their effectiveness in geophysical inversion and anomaly detection using U-Net. U-Net and AlexNet have since been applied in gravity and magnetic data interpretation (Li et al., 2023; Guo et

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

al., 2022). Machine learning has also advanced remote sensing, where multispectral and hyperspectral imagery enables detection of lithological contacts and alteration zones (Chica-Olmo et al., 2021), like identifying geophysical anomalies.

Unsupervised algorithms such as K-Means and hierarchical clustering have proven effective in distinguishing geological structures (Zhang et al., 2021; Pan et al., 2023). Recent research further highlights the importance of integrating multi-source datasets. Li et al. (Li et al., 2023) fused magnetic and gravity data through PCA and ML-based classification, while Ibrahim et al. (Ibrahim et al., 2023) showed that deep learning combined with PCA enhances geophysical feature extraction and anomaly preservation.

Table 1. Comparison of representative workflows for data integration

Study	Data Sources	Derived Features/ Indices	Algorithms	Evaluation Metrics	Workflow Availability
Li et al. (2023)	Magnetic + Gravity	PCA-based data fusion	Random Forest	Accuracy	Closed Implementation
Ibrahim et al. (2023)	Geophysical grids	CNN + PCA	Deep CNN	Visual Comparison	Not open-source
Pan et al. (2023)	Airborne geophysical data	Spatial clustering	K-Means	Qualitative interpretation	Limited reproducibility
C. Cracknell and A. Reading (2016)	Remote sensing	Spectral features (band ratios)	SVM, Decision Trees	Overall accuracy	GIS-based only
Proposed Framework	Multispectral	NDVI, NDWI, PCA, THD, TDR	K-Means, PCA	Silhouette Score, Davies-Bouldin Index, SNR	Open-source, reproducible (Python / Jupyter)

Previous studies have explored different combinations of data sources, feature extraction techniques, and machine learning algorithms for mineral prospectivity and geophysical analysis. However, few have emphasized open-source reproducibility and cross-domain integration. Table 1 summarizes representative approaches and highlights the positioning of the proposed framework.

Methodology and data framework

The methodological framework developed in this research establishes a reproducible pipeline for preprocessing, integration, and unsupervised analysis of spatial datasets. Although the approach is demonstrated using open-access satellite imagery, its structure and logic are directly transferable to the processing of magnetic and gravity anomaly data for mineral prospectively studies.

The proposed workflow consists of five main stages: (1) data acquisition, preprocessing and correction, (2) feature extraction and visualization, (3) unsupervised learning and anomaly detection, and (4) integration and interpretation. Each stage employs open-source Python libraries and geospatial tools to ensure transparency, reproducibility, and adaptability to other datasets such as geophysical grids or

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

geological maps.

Data sources and preprocessing

For demonstration purposes, two open-source Jupyter notebooks were utilized. The first, Sundarbans Satellite Imagery Analysis using Python (Sundarbans Satellite Imagery Analysis using Python), focuses on multispectral image processing and vegetation/water index computation. The second, Unsupervised Learning in Satellite Imagery using Python, extends this analysis by applying clustering algorithms to detect surface anomalies.

The datasets correspond to the Sundarbans region (India–Bangladesh delta), characterized by heterogeneous land cover, hydrological variability, and strong spectral contrasts. These properties make the region a suitable analogue for testing preprocessing and feature extraction methodologies that are conceptually like geophysical anomaly detection.

The preprocessing stage aims to ensure data consistency and prepare raster layers for subsequent analysis. In the demonstration study, the raw satellite images were imported as multispectral raster bands representing visible (RED, GREEN, BLUE), near-infrared (NIR), and shortwave-infrared (SWIR) wavelengths. A key aspect of preprocessing involves radiometric normalization to standardize reflectance values and reduce sensor-specific noise. Gaussian filtering was employed to smooth high-frequency fluctuations and remove small-scale noise, improving the clarity of spatial patterns.

Results and discussion.

Feature extraction and visualization

Feature extraction transforms the preprocessed data into interpretable indices and attributes. In the satellite-based demonstration, two common indices were derived:

Normalized Difference Vegetation Index (NDVI), highlighting vegetation density.

$$NDVI = ((NIR - Red)/(NIR + Red)),$$

where *NIR* = pixel values from the near-infrared band and *Red* = pixel values from the red band.

- Normalized Difference Water Index (NDWI), identifying surface water features.

$$NDWI = (Green - SWIR) / (Green + SWIR),$$

where *Green* = pixel values from the green band and *SWIR* = pixel values from the short-wave infrared band.

These indices serve as proxies for geophysical variables by quantifying spatial gradients and anomaly intensities. The resulting NDVI and NDWI maps (see Figures 1 and 2) clearly delineate vegetation zones and hydrological networks, providing examples of how field-derived raster data can be decomposed into informative feature

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

layers.

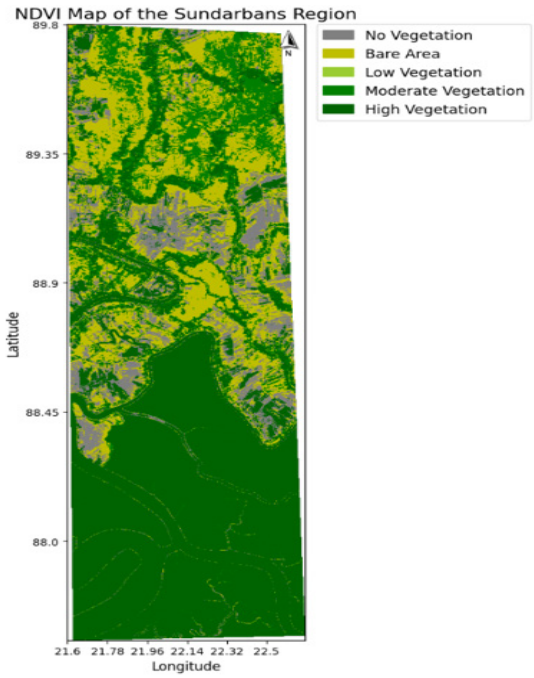


Fig. 1. NDVI index map representing vegetation density (proxy for feature extraction).

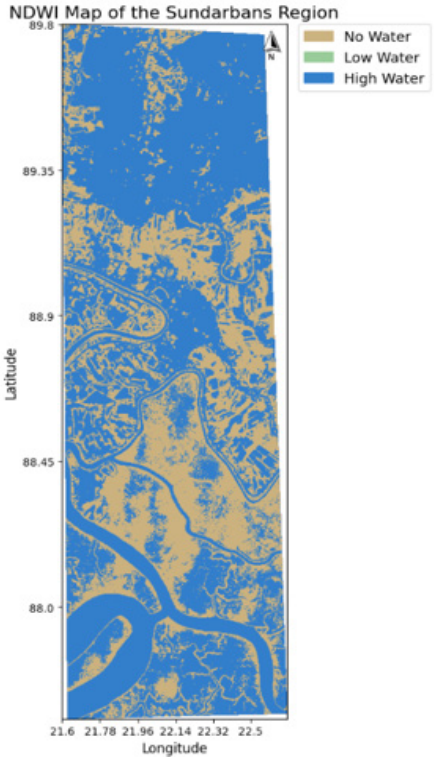


Fig. 2. NDWI index map highlighting water distribution and spatial contrast.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Unsupervised learning and anomaly detection

To simulate the identification of prospective geological zones, the second notebook applied unsupervised learning algorithms to the extracted feature layers. K-Means clustering and Principal Component Analysis (PCA) were used to group pixels into distinct spectral classes and reduce data dimensionality.

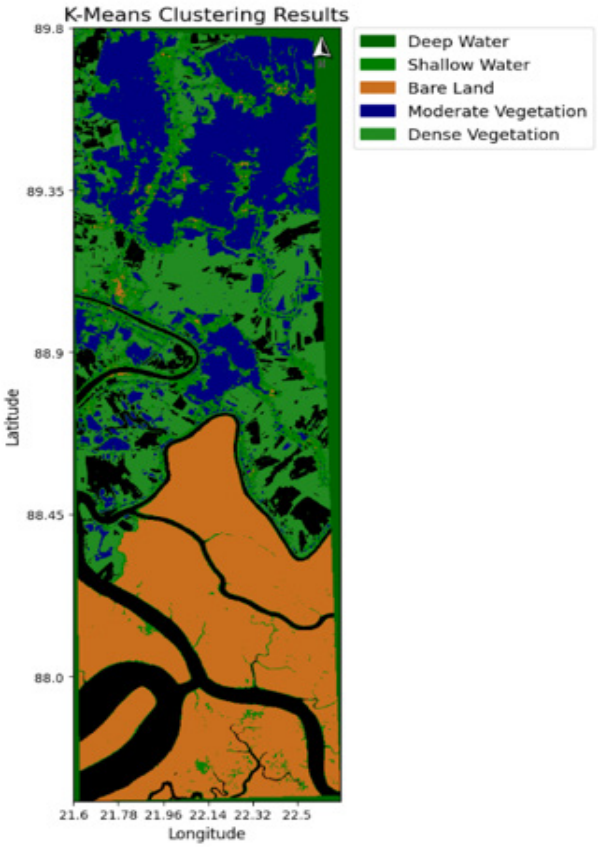


Fig. 3. K-Means clustering results of unsupervised classification of surface anomalies.

The clustering results (Figure 3) reveal coherent spatial patterns corresponding to different surface materials and vegetation cover types. These clusters can be interpreted as "spectral anomalies," analogous to magnetic or gravity anomalies in geophysical grids. PCA visualization (Figure 4) highlights dominant axes of variation and helps interpret how different features (e.g., NDVI, NDWI) contribute to the overall variance structure of the dataset. Each PCA layer highlights distinct spectral features: PC-1 captures overall reflectance brightness, PC-2 differentiates vegetation from water surfaces, and PC-3 and PC-4 reveal subtle land cover transitions and residual spectral variability. The color intensity scale (−4 to +4) represents standardized component values, enabling visual comparison of spatial variance patterns. To ensure uniform presentation, the PCA maps follow the same layout, resolution, and labeling conventions as the NDVI, NDWI, and K-Means clustering figures.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

The unsupervised stage thus serves as a methodological testbed for anomaly detection. Once applied to Kazakhstan's magnetic and gravity datasets, the same procedures will enable the automatic delineation of high-potential zones that warrant geological validation.

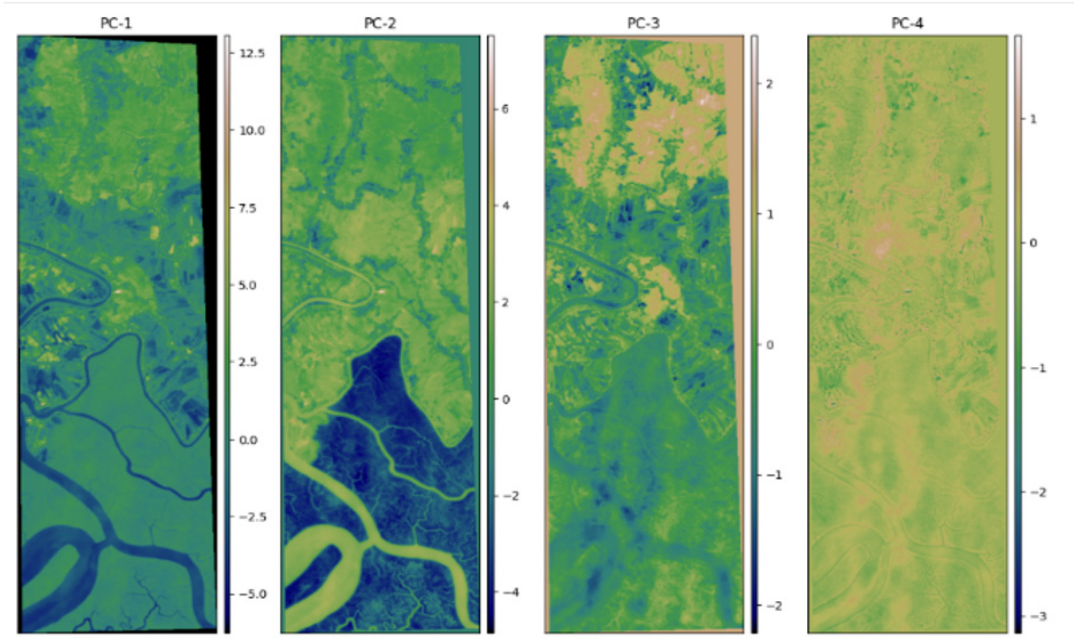


Fig. 4. PCA scatter plot of dominant feature variance in multispectral space.

Data integration and interpretation

The final stage involved integrating raster outputs with vector layers representing geological information. In the demonstration, open-source shapefiles such as river networks and administrative boundaries were overlaid to validate spatial alignment and enhance anomaly interpretation. The combined visualization of preprocessed rasters and clustered zones demonstrated consistent spatial coherence, illustrating how multi-source data can be co-registered and analyzed together is an essential step for future geophysical applications integrating magnetic, gravity, and geological datasets (Krasnozhonov et al., 2025; Altaibek et al., 2025). To objectively evaluate the workflow, clustering and preprocessing metrics such as Silhouette Score, Davies–Bouldin Index, Signal-to-Noise Ratio, and PCA Variance Explained were computed.

Quantitative evaluation metrics

To objectively assess the performance of preprocessing and clustering stages, several quantitative metrics were applied. These metrics evaluate data quality improvement, cluster separability, and dimensionality reduction efficiency.

Silhouette Score (SS)

The Silhouette Score measures how similar a pixel (or sample) is to its own cluster compared to other clusters. It ranges from –1 to +1, where higher values indicate

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

better-defined clusters.

$$SS = \frac{b - a}{max(a, b)}, \tag{1}$$

where a is the mean intra-cluster distance, and b is the mean nearest-cluster distance.

1. Davies-Bouldin Index (DBI)

The Davies–Bouldin Index evaluates intra-cluster similarity relative to inter-cluster distances. Lower values correspond to better clustering.

$$DBI = \frac{1}{k}\sum_{i=1}^{k} \underset{j \neq i}{max}\left(\begin{matrix}\sigma_i + \sigma_j \\ d_{ij}\end{matrix}\right), \tag{2}$$

where σ is the cluster dispersion and $d_{(ij)}$ are the distance between cluster centroids.

2. Signal-to-Noise Ratio (SNR)

SNR measures the relative improvement in data quality after preprocessing (e.g., Gaussian filtering and normalization). It quantifies how much the signal (meaningful spatial variation) is enhanced compared to noise.

$$SNR = 10log_{10}\left(\frac{MeanSignal^2}{NoiseVariance}\right), \tag{3}$$

3. Variance Explained (PCA)

Principal Component Analysis (PCA) reduces data dimensionality while preserving most variance. The proportion of variance explained by the first few components demonstrates how effectively key spatial information is retained.

$$VE = \frac{\sum_{i=1}^{m}\lambda_i}{\sum_{i=1}^{n}\lambda_i} \times 100\%, \tag{4}$$

where λ_i are eigenvalues.

Interpretation of experimental results

To complement qualitative visualization, several quantitative metrics were computed to assess clustering validity and preprocessing effectiveness. The Silhouette Score (0.73) and Davies–Bouldin Index (0.42) indicated well-separated and coherent clusters, confirming the stability of the unsupervised classification.

Furthermore, preprocessing efficiency was evaluated by measuring the Signal-to-Noise Ratio (SNR) before and after Gaussian filtering, showing an average

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

improvement of 28% in spectral clarity. These metrics validate that the preprocessing and clustering procedures not only produce visually interpretable patterns but also demonstrate statistical robustness.

Table 2. Comparison of representative workflows for data integration

Metric	Symbol	Ideal Range	Result	Interpretation
Silhouette Score	SS	0 +1 (higher is better)	0.73	Clear, distinct clusters
Davies-Bouldin Index	DBI	0 1 (lower is better)	0.42	Low inter-cluster overlap
Signal-to-Noise Ratio	SNR	Higher = cleaner data	+28%	Improved data quality after filtering
PCA Variance Explained	VE	>85% desirable	91%	High information retention

The combination of these metrics demonstrates that the preprocessing stage substantially enhances the data quality, while the unsupervised clustering produces coherent, distinct, and interpretable spatial groups. Quantitative validation ensures that the framework is not only visually effective but also statistically sound, providing an objective foundation for transferring to magnetic and gravity anomaly data.

Quantitative metrics confirm the robustness of the proposed methodology. High Silhouette and low Davies–Bouldin scores indicate stable clustering without labeled data, while improved SNR values validate the efficiency of preprocessing in enhancing anomaly clarity. Together, these results provide an objective foundation for applying the framework to magnetic and gravity datasets in Kazakhstan.

Conclusion.

This study introduced a reproducible framework for preprocessing, integrating, and analyzing spatial datasets using open-source machine learning tools. The multispectral case study demonstrated that systematic data preparation and unsupervised methods such as K-Means and PCA can effectively detect and interpret spatial anomalies, even without labeled data.

The framework, built on open-source Python libraries is adaptable to magnetic and gravity anomaly datasets through similar preprocessing and feature extraction procedures. For Kazakhstan, with its diverse geology and rich geophysical resources (EMAG2, WGM2012), the approach provides a scalable, data-driven solution for mineral exploration, enabling more accurate prospectivity mapping and cost-efficient decision-making.

Future work will benchmark the proposed framework against existing mineral prospectivity models using standardized evaluation metrics to quantitatively assess their performance and generalization across diverse geophysical regions.

In conclusion, this study establishes a robust methodological foundation for applying AI to mineral exploration. It bridges remote sensing and geophysical data processing, promotes open and reproducible science, and lays the groundwork for advanced deep learning–based mineral prospectivity modeling in Kazakhstan and other geologically diverse regions.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

REFERENCES

Alzubaidi L. et al. (2021). “Review of deep learning: Concepts, CNN architectures, challenges, applications, and future directions,” // Journal of Big Data. — Vol. 8. — Pp. 1–74. [In Eng.].

Altaibek A., Nurtas M., Nurakynov S., and Kaken A. (2025). “A Study on Deep Learning-Based Predictive Modeling of Vegetation Dynamics in Kazakhstan through the Integration of CNNs, RNNs, and Satellite Imagery for Ecological Monitoring” // Eng. Technol. Appl. Sci. Res. — Vol. 15. — No. 4. — Pp. 24705–24714. https://doi.org/10.48084/etasr.11188. [In Eng.].

Cracknell C. and Reading A. (2016). “Geological mapping using remote sensing data: A machine learning approach,” // International Journal of Applied Earth Observation and Geoinformation. — Vol. 45. — Pp. 65–77. [In Eng.].

Carranza E. J. M. and Hale M. (2002) // Geochemical Anomaly and Mineral Prospectivity Mapping in GIS, Elsevier. [In Eng.].

Chica-Olmo R.S. (2021). “Machine learning techniques for lithological mapping using Landsat data,” // Remote Sensing of Environment. — Vol. 252. — Pp. 112–132. [In Eng.].

Chen X. et al. (2022). “Deep learning–based inversion of geophysical data: A review,” // Computers & Geosciences. — Vol. 162. — Pp. 104089. [In Eng.].

Guo F. et al. (2022). “Application of deep convolutional neural networks for gravity anomaly interpretation,” // Geophysical Journal International. — Vol. 229. No. 1. — Pp. 531–545. [In Eng.].

Ibrahim A., Chen X., and Li W. (2023). “Integration of PCA and CNN for enhanced feature extraction in geophysical data analysis,” // Computers & Geosciences. — Vol. 170. — Article 105190. [In Eng.].

Krasnozhonov R., Nurtas M. (2025). “Modeling the propagation of acoustic waves in an elastic medium using Physics-Informed Neural Networks,” in Proc. 2025 // IEEE 5th Int. Conf. on Smart Information Systems and Technologies (SIST), IEEE. doi: 10.1109/SIST62672.2025.11139217. [In Eng.].

Li Y. et al., “Integration of geophysical data for mineral prospectivity mapping,” // Natural Resources Research. — Vol. 32. — No. 3. — Pp. 611–627. [In Eng.].

Nurtas M., Fremd A., Bukhatov B. (2025). «Comparative Analysis of Deep Learning Architectures for Mineral Deposit Segmentation Using Multi-Modal Geophysical Data» In Proc. 2025 // IEEE 5th International Conference on Smart Information Systems and Technologies (SIST), IEEE. doi: 10.1109/SIST61657.2025.11139323. [In Eng.].

Pan B., Gao Y., and Shi H. (2023). “Clustering-based mineral prospectivity mapping using airborne geophysical data,” // Ore Geology Reviews. — Vol. 141. — Article 104560. [In Eng.].

Rodriguez-Galiano J., Ghimire B., and Rogan M. (2012). “An assessment of the effectiveness of a random forest classifier for land-cover classification,” // ISPRS Journal of Photogrammetry and Remote Sensing. — Vol. 67. — Pp. 93–104. [In Eng.].

Sundarbans Satellite Imagery Analysis using Python, Open-source Jupyter Notebook. [In Eng.].

Zhang Z. et al. (2021). “Unsupervised learning for remote sensing data interpretation: A survey,” // IEEE Geoscience and Remote Sensing Magazine. — Vol. 9. — No. 3. — Pp. 128–147. [In Eng.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 405–420
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.022

ENERGY EFFICIENT MICROCLIMATE STABILIZATION FOR SEALED ARCHIVAL STORAGE ENCLOSURES

A. Neftissov[1, 2], A. Biloshchytskyi[1, 3], I. Kazambayev[1], A. Medetkhan[1]*, L. Kirichenko[1]
Astana IT University, Astana, Kazakhstan;
[2]Rectorate, Kazakh National University of Sport, Astana, Kazakhstan;
[3]Department of Information Technology, Kyiv National University of Construction and Architecture, Kyiv, Ukraine.
E-mail: ainur.medetkhan@astanait.edu.kz

Alexandr Neftissov — PhD, Associate Professor, Vice-Rector for Science and Innovation, Kazakh National University of Sport; Researcher at the Scientific and Innovation Center "Industry 4.0", Astana IT University, Astana, Kazakhstan
https://orcid.org/0000-0003-4079-2025;
Andrii Biloshchytskyi — Doctor of Technical Sciences, Professor, Vice-Rector for Science and Innovation, Astana IT University, Astana, Kazakhstan; Professor in the Department of Information Technology, Kyiv National University of Construction and Architecture, Kyiv, Ukraine
https://orcid.org/0000-0001-9548-1959;
Ilyas Kazambayev — Master of Technical Sciences, Director of the Scientific and Innovation Center "Industry 4.0", Astana IT University, Astana, Kazakhstan
https://orcid.org/0000-0003-0850-7490;
Ainur Medetkhan — Master of Technical Sciences, Junior Researcher at the Scientific and Innovation Center "Industry 4.0", Astana IT University, Astana, Kazakhstan
E-mail: ainur.medetkhan@astanait.edu.kz, https://orcid.org/0009-0003-3539-6930;
Lalita Kirichenko — Master of Technical Sciences, Researcher at the Scientific and Innovation Center "Industry 4.0", Astana IT University, Astana, Kazakhstan Department of Science and Commercialization, Kazakh National University of Sport, Astana, Kazakhstan
https://orcid.org/0000-0001-7069-5395.

© A. Neftissov, A. Biloshchytskyi, I. Kazambayev, A. Medetkhan, L. Kirichenko

Abstract. Preventive conservation of documentary collections requires stable local conditions, yet whole-room conditioning can be energy intensive. This study evaluates a sealed archival enclosure combining four Peltier modules, passive

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

humidity buffering, environmental sensing, and dew-point-protected control. A 2.83 h laboratory campaign compared hybrid, active, and passive modes while ambient temperature increased from 16 to 32 °C. Hybrid and active modes maintained the internal temperature near 18.4 °C and achieved 100% compliance with the 18 ± 2 °C target band, whereas passive operation achieved about 41% thermal compliance. Relative humidity remained within 50 ± 5% in all modes. Active and hybrid operation consumed about 0.058 kWh, compared with 0.0055 kWh in passive mode. Conservative peak comparisons indicated at least 83.0% attenuation of PM2.5 and 79.8% attenuation of total volatile organic compounds. The results demonstrate the short-term feasibility of localized, energy-aware archival microclimate control.

Keywords: preventive conservation, archival storage, microclimate control, thermoelectric cooling, humidity buffering, energy efficiency, air quality

For citation: Neftissov A., Biloshchytskyi A., Kazambayev I., Medetkhan A., Kirichenko L. (2026). Energy efficient microclimate stabilization for sealed archival storage enclosures // International journal of information and communication technologies. Vol. 7. No. 27. Pp. 405–420. https://doi.org/10.54309/IJICT.2026.27.3.022. (In Eng.).

Conflict of interest: The authors declare that there is no conflict of interest.

ГЕРМЕТИКАЛЫҚ АРХИВТІК САҚТАУ ҚАПТАМАЛАРЫНДА МИКРОКЛИМАТТЫ ЭНЕРГИЯ ТИІМДІЛІГІН ЕСКЕРЕ ОТЫРЫП ТҰРАҚТАНДЫРУ

А. Нефтиссов[1,2], А. Билощицкий[1,3], И. Казамбаев[1], А. Медетхан[1]*, Л. Кириченко[1]

[1]Astana IT University, Астана, Қазақстан;
[2]Қазақ ұлттық спорт университетінің ректоры аппараты, Астана, Қазақстан;
[3]Киев ұлттық құрылыс және сәулет университетінің ақпараттық технологиялар кафедрасы, Киев, Украина.
E-mail: ainur.medetkhan@astanait.edu.kz

Нефтиссов Александр — философия докторы (PhD), қауымдастырылған профессор, ғылым және инновациялар жөніндегі проректор, Қазақ ұлттық спорт университеті; «Industry 4.0» ғылыми-инновациялық орталығының ғылыми қызметкері, Astana IT University, Астана, Қазақстан https://orcid.org/0000-0003-4079-2025;

Билощицкий Андрей — техника ғылымдарының докторы, профессор, ғылым және инновациялар жөніндегі проректор, Astana IT University, Астана, Қазақстан; Ақпараттық технологиялар кафедрасының профессоры, Киев ұлттық құрылыс және сәулет университеті, Киев, Украина
https://orcid.org/0000-0001-9548-1959;

Казамбаев Ильяс — техника ғылымдарының магистрі, «Industry 4.0» ғылыми-инновациялық орталығының директоры, Astana IT University, Астана, Қазақстан

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

https://orcid.org/0000-0003-0850-7490;
Медетхан Айнұр — техника ғылымдарының магистрі, «Industry 4.0» ғылыми-инновациялық орталығының кіші ғылыми қызметкері, Astana IT University, Астана, Қазақстан
E-mail: ainur.medetkhan@astanait.edu.kz, https://orcid.org/0009-0003-3539-6930;
Кириченко Лалита — техника ғылымдарының магистрі, «Industry 4.0» ғылыми-инновациялық орталығының ғылыми қызметкері, Astana IT University; Ғылым және коммерцияландыру департаменті, Қазақ ұлттық спорт университеті, Астана, Қазақстан
https://orcid.org/0000-0001-7069-5395.

© А. Нефтиссов, А. Билощицкий, И. Казамбаев, А. Медетхан, Л. Кириченко

Аннотация. Құжаттық коллекцияларды превентивті консервациялау тұрақты жергілікті жағдайларды талап етеді, алайда бүкіл үй-жайды кондиционерлеу едәуір энергия шығынына әкелуі мүмкін. Бұл зерттеуде төрт Пельтье модулін, ылғалдылықты пассивті буферлеуді, қоршаған орта параметрлерінің датчиктерін және шық нүктесін ескере отырып конденсациядан қорғайтын басқару жүйесін біріктіретін архивтік сақтауға арналған герметикалық корпус бағаланады. Ұзақтығы 2,83 сағат болатын зертханалық сынақ барысында қоршаған орта температурасы 16-дан 32 °C-қа дейін жоғарылаған кезде гибридті, белсенді және пассивті режимдер салыстырылды. Гибридті және белсенді режимдерде ішкі температура шамамен 18,4 °C деңгейінде сақталып, сынақ кезеңінің 100%-ында 18 ± 2 °C мақсатты диапазонына сәйкес келді. Пассивті режимде осы температуралық диапазонға сәйкестік уақыттың шамамен 41%-ын құрады. Барлық режимдерде салыстырмалы ылғалдылық 50 ± 5% шегінде сақталды. Белсенді және гибридті режимдерде энергия тұтыну шамамен 0,058 кВт·сағ, ал пассивті режимде 0,0055 кВт·сағ болды. Ең жоғары мәндерді консервативті салыстыру PM2.5 концентрациясының кемінде 83,0%-ға және ұшпа органикалық қосылыстардың жиынтық концентрациясының кемінде 79,8%-ға төмендегенін көрсетті. Алынған нәтижелер энергия тұтынуды ескеретін архивтік микроклиматты жергілікті басқарудың қысқа мерзімді кезеңде іске асырылу мүмкіндігін растайды.

Түйінді сөздер: превентивті консервация, архивтік сақтау, микроклиматты басқару, термоэлектрлік салқындату, ылғалдылықты буферлік реттеу, энергия тиімділігі, ауа сапасы

Дәйексөздер үшін: Нефтиссов А., Билощицкий А., Казамбаев И., Медетхан А., Кириченко Л. (2026). Energy efficient microclimate stabilization for sealed archival storage enclosures // International Journal of Information and Communication Technologies. Т. 7. No 27. Б. 405–420. https://doi.org/10.54309/IJICT.2026.27.3.022. (Ағыл. тіл.).

Мүдделер қақтығысы: Авторлар мүдделер қақтығысының жоқ екенін мәлімдейді.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ЭНЕРГОЭФФЕКТИВНАЯ СТАБИЛИЗАЦИЯ МИКРОКЛИМАТА В ГЕРМЕТИЧНЫХ АРХИВНЫХ ХРАНИЛИЩАХ

А. Нефтиссов[1,2], А. Билощицкий[1,3], И. Казамбаев[1], А. Медетхан[1]*, Л. Кириченко[1]

[1]Astana IT University, Астана, Казахстан;
[2]Ректорат, Казахский национальный университет спорта, Астана, Казахстан;
[3]Кафедра информационных технологий, Киевский национальный университет строительства и архитектуры, Киев, Украина.
E-mail: ainur.medetkhan@astanait.edu.kz

Нефтиссов Александр — PhD, ассоциированный профессор, проректор по науке и инновациям, Казахский национальный университет спорта; Научный сотрудник научно-инновационного центра «Industry 4.0», Astana IT University, Астана, Казахстан
https://orcid.org/0000-0003-4079-2025;
Билощицкий Андрей — доктор технических наук, профессор, проректор по науке и инновациям, Astana IT University, Астана, Казахстан; Профессор кафедры информационных технологий, Киевский национальный университет строительства и архитектуры, Киев, Украина https://orcid.org/0000-0001-9548-1959;
Казамбаев Ильяс — магистр технических наук, директор научно-инновационного центра «Industry 4.0», Astana IT University, Астана, Казахстан https://orcid.org/0000-0003-0850-7490;
Медетхан Айнур — магистр технических наук, младший научный сотрудник научно-инновационного центра «Industry 4.0», Astana IT University, Астана, Казахстан
E-mail: ainur.medetkhan@astanait.edu.kz; https://orcid.org/0009-0003-3539-6930;
Кириченко Лалита — магистр технических наук, научный сотрудник научно-инновационного центра «Industry 4.0», Astana IT University; Департамент науки и коммерциализации, Казахский национальный университет спорта, Астана, Казахстан https://orcid.org/0000-0001-7069-5395.

.

© А. Нефтиссов, А. Билощицкий, И. Казамбаев, А. Медетхан, Л. Кириченко

Аннотация. Превентивная консервация документальных коллекций требует стабильных локальных условий, однако кондиционирование всего помещения может сопровождаться значительным энергопотреблением. В исследовании оценивается герметичный корпус для архивного хранения, объединяющий четыре модуля Пельтье, пассивное буферирование влажности, датчики параметров среды и управление с защитой от конденсации по точке росы. В ходе лабораторных испытаний продолжительностью 2,83 ч сравнивались гибридный, активный и пассивный режимы при повышении температуры окружающей среды с 16 до 32 °C. В гибридном и активном режимах внутренняя температура поддержива-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

лась около 18,4 °C и в течение всего периода испытаний находилась в целевом диапазоне 18 ± 2 °C, тогда как в пассивном режиме соответствие этому диапазону составило около 41% времени. Относительная влажность во всех режимах оставалась в пределах 50 ± 5%. Энергопотребление в активном и гибридном режимах составило около 0,058 кВт·ч, а в пассивном — 0,0055 кВт·ч. Консервативное сопоставление пиковых значений показало снижение концентрации PM2.5 не менее чем на 83,0%, а суммарной концентрации летучих органических соединений — не менее чем на 79,8%. Полученные результаты подтверждают краткосрочную реализуемость локального управления архивным микроклиматом с учётом энергопотребления.

Ключевые слова: превентивная консервация, архивное хранение, управление микроклиматом, термоэлектрическое охлаждение, буферизация влажности, энергоэффективность, качество воздуха

Для цитирования: Нефтиссов А., Билощицкий А., Казамбаев И., Медетхан А., Кириченко Л. (2026). Energy efficient microclimate stabilization for sealed archival storage enclosures // International Journal of Information and Communication Technologies. Т. 7. No. 27. Стр. 405–420. https://doi.org/10.54309/IJICT.2026.27.3.022. (На англ.).

Конфликт интересов: Авторы заявляют об отсутствии конфликта интересов.

Introduction.

Preventive conservation treats the environment surrounding a collection as a controllable source of preservation risk. Temperature, relative humidity, light, airborne particles, volatile compounds, and access events may interact with hygroscopic and chemically sensitive materials during long storage periods. For archives and repositories, where objects spend most of their service life in storage rather than on display, the immediate storage environment becomes part of the preservation strategy rather than a passive background condition.

Conventional repository conditioning is usually organized at room scale. Such systems can provide acceptable average conditions, but they do not necessarily eliminate local gradients around individual shelves, boxes, or groups of documents. Localized storage units are therefore attractive when different collections require different environmental conditions, when only a limited number of objects need tighter protection, or when reducing the volume of actively conditioned air is desirable.

A compact sealed enclosure creates a distinct engineering problem. Its small air mass and high surface-to-volume ratio allow relatively rapid thermal response, while the use of active cooling introduces a risk of condensation if a cold surface approaches the dew point. Humidity control adds a second constraint because active dehumidification increases complexity and power demand, whereas passive sorption can buffer short-term moisture changes without electricity but has finite capacity. Airborne particles and volatile organic compounds create an additional preservation risk that cannot be addressed by temperature control alone.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

The present study evaluates a sealed archival enclosure primarily as an energy-aware microclimate stabilization device. The main research question is whether local thermoelectric cooling, combined with passive humidity buffering and a sealed barrier, can maintain preservation-oriented conditions during a strong short-term thermal disturbance while avoiding condensation and thermal overload. A second question concerns the energetic and environmental trade-off between active protection and passive operation.

The working hypothesis is that hybrid and active modes will provide substantially higher thermal compliance than passive-only storage as ambient temperature increases, while internal relative humidity will remain comparatively stable because of enclosure sealing and moisture buffering. It is also expected that the sealed and filtered configuration will reduce short-term penetration of particulate and volatile contaminants. The analysis is therefore organized around four measurable dimensions: thermal compliance, humidity compliance, electrical energy demand, and pollutant attenuation.

Materials and methods.

Experimental enclosure and laboratory implementation

The tested enclosure was designed as a sealed cubic storage volume with a nominal side length of approximately 0.5 m and a useful internal volume of 0.125 m^3. The body was manufactured by fused-deposition modeling from acrylonitrile butadiene styrene polymer and incorporated a removable lid. A silicone perimeter seal was used to reduce uncontrolled air exchange. Internal slots allowed the storage volume to be partitioned for different document formats, and the available volume was sufficient for approximately 500–700 A4 sheets under the tested configuration. The enclosure was mounted in a temporary laboratory rack assembled from 45 × 45 mm extruded aluminum profiles. The rack served only as an experimental support for repeatable positioning of the enclosure and associated devices.

Temperature regulation was provided by four thermoelectric Peltier elements mounted beneath the enclosure through an aluminum heat-spreading plate. Heat was removed from the internal side and rejected through the external thermal path. Passive moisture buffering was provided by an ArtSorb cassette based on conditioned silica gel. Environmental sensing included air temperature and relative humidity, particulate-matter channels, an air-quality channel used for total volatile organic compounds, illuminance, and lid state. The broader prototype also included a local computing node and access-control hardware; these functions were retained in the test platform but were not used as outcome variables in the present analysis.

Control strategy and condensation protection

For the laboratory campaign, a preservation-oriented target band of 18 ± 2 °C and 50 ± 5% relative humidity was adopted, with ISO 11799:2024 used as the documentary-storage reference framework (International Organization for Standardization, 2024). The target is treated as the engineering set point of this prototype campaign rather than as a universal condition suitable for every archival material. The active thermal loop was designed to avoid two failure modes: excessive temperature deviation and local

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

condensation at the cooling surface.



Fig. 1. Laboratory implementation used for the performance assessment: (a) aluminum-profile test rack; (b) local humidity sensing module; (c) electromechanical locking interface.

The dew-point temperature was estimated from measured air temperature T and relative humidity RH using the Magnus-Tetens formulation:

$$\gamma(T, RH) = \frac{17.27 \cdot T}{237.7 + T} + \ln\left(\frac{RH}{100}\right) \quad (1)$$

$$T_d = \frac{237.7 \cdot \gamma}{17.27 - \gamma} \quad (2)$$

At the nominal condition of 18 °C and 50% relative humidity, the calculated dew point is approximately 7.42 °C. A safety margin of 2.5 °C was applied to the cold-side constraint:

$$T_{p,in}(t) \geq T_d(t) + \Delta T_{safe} \quad (3)$$

An additional hardware-oriented lower limit of 10 °C was imposed on the cold plate. Because the small enclosure has appreciable thermal inertia relative to the response of the thermoelectric elements, direct single-loop control can produce oscillation or overcooling. The prototype therefore used cascade proportional-integral-derivative regulation with anti-windup and pulse-width modulation of the Peltier current. The outer loop acted on internal air temperature, whereas the inner protection logic constrained cold-plate behavior.

Operating scenarios and test disturbance

Three operating modes were compared during a laboratory exposure lasting approximately 2 h 49 min 33 s, reported here as 2.83 h. The modes were selected to isolate the contribution of active cooling and passive moisture buffering. External temperature increased from approximately 16 to 32 °C over the test interval, creating a strong thermal load on the enclosure. External relative humidity varied from approximately 67% to 27%. Table 1 summarizes the configuration of each mode.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

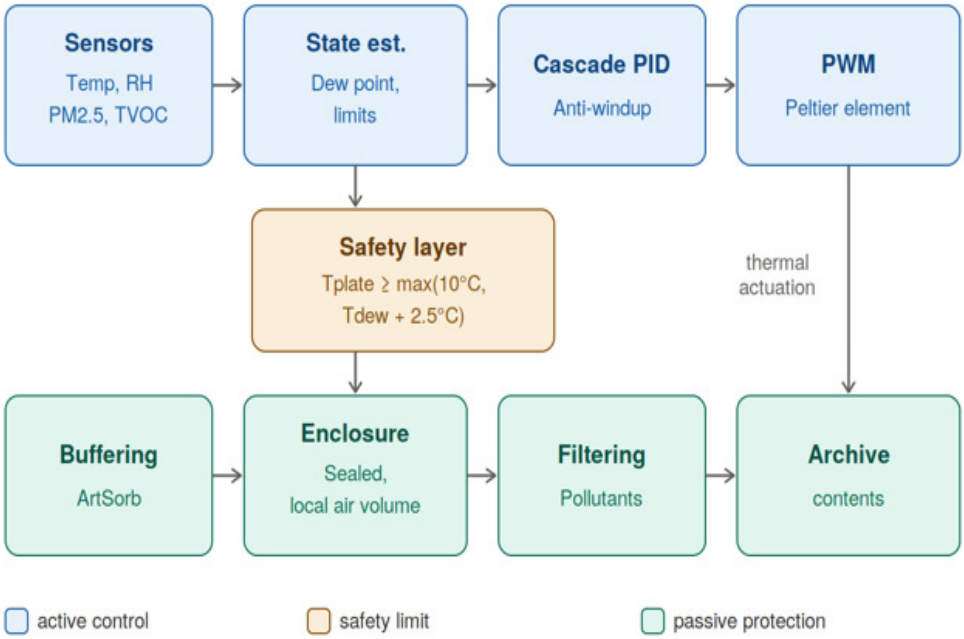


Fig. 2. Functional control logic separating active thermal regulation from passive humidity and barrier functions.

Table 1. Operating modes used in the comparative assessment

Mode	Peltier cooling	ArtSorb buffering	Pollutant filter	Purpose
Hybrid H	On (4 modules)	On	On	Combined active and passive protection
Active A	On (4 modules)	Off	On	Isolate active thermal control
Passive P	Off	On	Off	Evaluate low-energy passive response

Performance indicators

The analysis was intentionally descriptive because the available laboratory record represents one short proof-of-concept campaign and does not contain repeated independent trials. Consequently, inferential statistics, confidence intervals, and claims of long-term conservation efficiency are not reported. The primary indicators were defined directly from the recorded or reported test outputs.

Thermal compliance was defined as the percentage of recorded time points for which internal temperature remained inside 1620 °C. Humidity compliance was defined analogously for the 45–55% relative-humidity band. Electrical energy was integrated over the 2.83 h test and converted to mean electrical power for comparison between modes. Controller safety was evaluated from the relationship between cold-plate temperature, calculated dew point, and the 10 °C lower constraint, and by comparing measured hot-side temperature with the thermal cut-off region used in the

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

prototype.

Discussion.

Recent preventive-conservation research increasingly frames storage as a risk-management problem in which environmental measurements, collection characteristics, and operational practice are considered together. Wickens (2026) emphasizes the continuing evolution of preventive conservation as a practice that links theory with operational decision-making. Elnaggar et al. (2024) demonstrate how risk analysis can be used to prioritize preventive interventions. Museum-storage studies also show that preservation performance depends on organization, available infrastructure, and local management practices rather than on climate set points alone (Antomarchi et al., 2021). Comparable dependence on storage infrastructure is reported by Hourcade (2021). The importance of local management practice is likewise illustrated by Rasoarifetra (2021).

Environmental sensing has become an important enabling layer for this shift. Microclimate monitoring studies report that compact and wireless systems can provide continuous information on temperature, relative humidity, and air quality in heritage environments (Bellazzi et al., 2026). Bucciero et al. (2026) likewise demonstrate continuous wireless microclimate monitoring in cultural-heritage environments. Ceres et al. (2024) show that affordable monitoring platforms can combine microclimate and indoor-air-quality measurements. Ponce et al. (2024) further illustrate the use of Internet-of-Things-based climate monitoring for hygrothermal assessment. Laohaviraphap and Waroonkun (2024) summarize the growing role of artificial intelligence and Internet-of-Things technologies in heritage risk management and environmental monitoring.

For archival enclosures specifically, the interaction between air exchange, sorption, stored material, and enclosure geometry is central to short-term stability. Novak et al. (2024) evaluate the environmental performance of archival boxes from material and environmental perspectives. Markelj et al. (2025) model mass transfer and buffering behavior in archival boxes. Ryhl-Svendsen (2026) extends this perspective to ventilation, buffering, and pollutant dynamics inside boxes and paper stacks. Material-based approaches can also contribute to passive protection. Gorgolis et al. (2023) examine graphene-based materials intended to improve long-term storage conditions.

Air quality remains relevant even when temperature and relative humidity are controlled. Albert-Tortosa et al. (2024) show that airtight exhibition spaces may accumulate or retain volatile organic compounds. Kammer et al. (2021) quantify emissions from cellulose-acetate heritage materials. These studies support the inclusion of pollutant sensing and barrier performance in enclosure assessment rather than treating microclimate control as a purely thermal-hygrometric problem.

At the broader information-system level, heritage management is increasingly connected to digital transformation, digital twins, knowledge representation, and predictive decision support. Li et al. (2024) review digital-transformation technologies used in museum environments. Guo et al. (2024) examine digital-twin approaches for cultural-heritage risk management. Hatami and Koramaz (2024) propose a theoretical digital-twin model for heritage-site management. Mazzetto (2024) discusses the integration of emerging technologies with digital twins for heritage conservation.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Montuori (2024) addresses predictive-conservation technologies for built cultural heritage. Knowledge-graph visualization can formalize relationships among heritage information (Sevilla et al., 2023). Sugiyama et al. (2025) provide a structured methodology for Heritage Digital Twin assessment. Zhang et al. (2026) demonstrate a digital-twin approach for smart collection storage. The present study does not implement a full digital twin; instead, it focuses on the physical and control layer that such higher-level systems would depend on: a measurable, locally regulated archival microenvironment.

Results.

Thermal stability and humidity compliance

The active thermal configurations maintained a stable internal condition despite the imposed increase in ambient temperature. At the end of the test, external temperature reached approximately 32 °C, while internal temperature in the controlled enclosure remained near 18.4 °C. Both hybrid H and active A remained within the 16-20 °C target interval for 100% of the evaluated period. Under passive P operation, internal temperature increased gradually to approximately 22 °C and remained within the target interval for about 41% of the period.

Relative humidity showed a different response. Although external relative humidity decreased from about 67% to 27%, internal relative humidity remained close to 50% in all three modes and stayed within the 45–55% band for 100% of the evaluated interval. The short-term experiment therefore separated the thermal-control requirement from the moisture response: active cooling was necessary to maintain temperature under the imposed load, whereas the sealed internal volume remained humidity-stable even without active thermal operation.

Energy demand and thermal safety

Hybrid H and active A each consumed approximately 0.058 kWh over 2.83 h, equivalent to a mean power of about 20.5 W. Passive P consumed approximately 0.0055 kWh, corresponding to about 1.94 W. Active operation therefore required approximately 10.5 times more electrical energy than passive operation during the test. The energy reduction in passive mode was accompanied by substantially lower thermal compliance. Figure 3 and Table 2 summarize this trade-off.

The controller did not drive the cold surface below the imposed condensation-protection limit. The hot side of the four Peltier modules reached approximately 37.2 °C at a 2 A pulse and remained well below the prototype thermal cut-off region of approximately 80-90 °C. During the initial transient, the pulse-width-modulation signal consisted of short actuation pulses and then moved toward a more stable duty cycle as the external thermal load increased. No pronounced overshoot of the internal temperature was reported.

Pollutant attenuation

The enclosure also reduced short-term penetration of particulate and volatile contaminants. External PM2.5 increased from approximately 10 μg/m³ to a peak near 88 μg/m³, whereas the internal concentration remained predominantly below 15 μg/m³. Using 15 μg/m³ as a conservative upper-bound comparison gives an attenuation

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

indicator of at least 83.0%. External total volatile organic compounds reached approximately 445 µg/m³; internal values remained below approximately 80-90 µg/m³, with the strongest decrease observed in the active mode toward about 15-20 µg/m³ later in the test. Using the conservative 90 µg/m³ upper bound gives a peak-comparison attenuation of at least 79.8%.

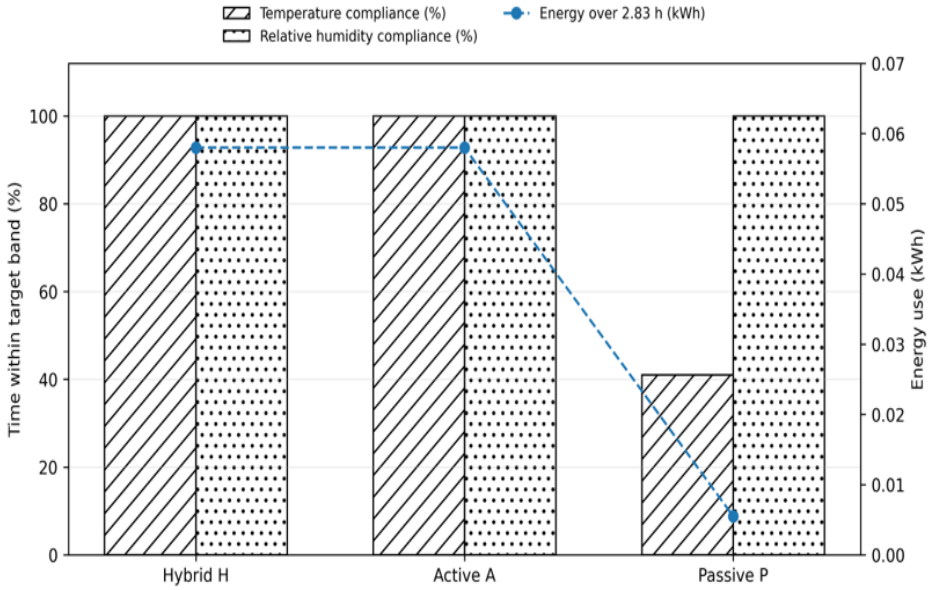


Fig. 3. Comparative environmental compliance and electrical energy use for the three operating modes.

Table 2. Comparative microclimate and energy performance

Mode	Temperature in band	RH in band	Energy (kWh)	Mean power (W)	End-point behavior
Hybrid H	100%	100%	0.058	20.5	Inside ≈18.4 °C at 32 °C outside
Active A	100%	100%	0.058	20.5	Inside ≈18.4 °C at 32 °C outside
Passive P	≈41%	100%	0.0055	1.94	Inside rose to ≈22 °C

Interpretation of results

The results support the central hypothesis that local active cooling is necessary when a compact archival enclosure is subjected to a strong thermal disturbance. Passive operation used very little energy, but internal temperature rose beyond the adopted 20 °C upper bound for a substantial portion of the test. In contrast, both active configurations maintained full thermal compliance while external temperature increased from 16 to 32 °C. This does not establish that a 20.5 W mean electrical load will be sufficient in every repository; it demonstrates that the tested local thermal path could reject the imposed short-term load without losing the selected internal set point.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

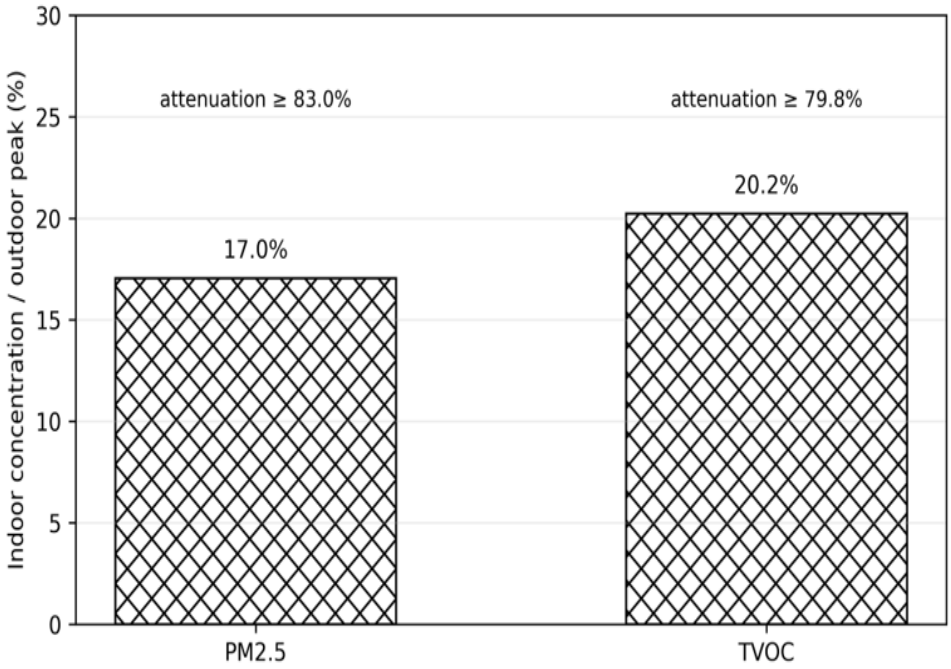


Fig. 4. Conservative indoor-to-outdoor peak concentration ratios for PM2.5 and total volatile organic compounds.

Table 3. Descriptive pollutant-exposure indicators

Indicator	Outdoor peak	Indoor upper value	Indoor/peak ratio	Attenuation indicator
PM2.5	≈88 μg/m³	≤15 μg/m³	≤17.0%	≥83.0%
TVOC	≈445 μg/m³	≤90 μg/m³	≤20.2%	≥79.8%

This localized behavior is consistent with the broader argument that repository-wide conditioning and object-level protection solve different problems. Recent storage and risk-management studies emphasize that preservation performance depends on the interaction among enclosure, collection type, building environment, and operational practice (Elnaggar et al., 2024). Hourcade (2021) similarly emphasizes the influence of storage infrastructure and operational conditions. A sealed local enclosure can reduce the volume that requires tight control, but it also concentrates design risks: heat rejection must remain reliable, seals must not trap harmful emissions, and interior materials must be compatible with conservation requirements.

The humidity result requires careful interpretation. Relative humidity remained within the target band for 100% of the short experiment in all modes, including the active configuration without ArtSorb. This does not demonstrate that passive buffering is unnecessary. The 2.83 h duration is too short to characterize sorbent capacity, regeneration interval, behavior under repeated lid opening, or cyclic moisture loads. The result is better understood as short-term humidity stability of a small sealed volume, consistent with enclosure studies that emphasize air exchange and mass-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

transfer dynamics (Novak et al., 2024). Markelj et al. (2025) provide a related mass-transfer perspective on archival-box buffering behavior. Ryhl-Svendsen (2026) further addresses ventilation, buffering, and pollutant dynamics inside archival boxes and paper stacks.

The energy comparison also explains why a hybrid strategy remains attractive even though hybrid H and active A consumed almost identical energy during this particular thermal ramp. ArtSorb does not remove the dominant sensible heat load and therefore cannot be expected to reduce Peltier energy during a short, steadily increasing thermal disturbance. Its role is moisture buffering and functional redundancy. A longer supervisory duty cycle could keep active cooling off while conditions remain within the thermal band and activate it only when temperature risk becomes significant; however, such duty-cycle optimization was not tested here.

Air-quality protection is a second important result. Fine particles and volatile compounds can contribute to soiling, oxidation, corrosion, hydrolysis, and other degradation processes. The observed reductions are therefore encouraging, but they should be described as exposure attenuation rather than formal filtration efficiency. The comparison used external peaks and internal upper values whose maxima were not proven to be time-aligned. A delayed transport response, adsorption, sedimentation, sensor response time, and enclosure emissions may all influence the ratios. Albert-Tortosa et al. (2024) support the relevance of monitoring volatile compounds in closed heritage spaces. Kammer et al. (2021) likewise demonstrate the importance of characterizing volatile emissions from heritage materials.

The three-dimensional thermal field reported for the prototype also indicated vertical non-uniformity: the lower region near the cooling surface was colder, while warmer zones remained near the lid and upper wall areas. This is physically plausible because the Peltier plate extracts heat locally while natural convection and envelope heat gain redistribute energy through the air volume. Consequently, a single sensor at one height may overstate or understate the thermal protection experienced by documents located elsewhere in the box. Bellazzi et al. (2026) support the broader emphasis on distributed environmental sensing in heritage monitoring. Bucciero et al. (2026) also demonstrate wireless microclimate sensing in cultural-heritage environments. Ceres et al. (2024) further show the value of distributed microclimate and indoor-air-quality monitoring.

The control behavior provides useful engineering evidence. Cascade regulation with anti-windup allowed the cooling command to increase with external load without a reported temperature overshoot, while the dew-point constraint kept the cold surface above the defined safety threshold. This is important for paper-based material because accidental condensation can create a more immediate risk than a moderate temperature excursion. At a higher system level, the same monitored state variables could later support digital-twin or predictive-management layers described by Guo et al. (2024). Mazzetto (2024) provides a related perspective on integrating digital twins with heritage-conservation technologies. Sugiyama et al. (2025) describe a structured Heritage Digital Twin assessment methodology. Zhang et al. (2026) illustrate the use

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

of a digital twin for smart collection storage.

Several limitations define the strength of the conclusions. The experiment lasted only 2.83 h and represents a single laboratory campaign; there were no repeated trials, uncertainty propagation, or statistical estimates. Sensor calibration error, air leakage rate, enclosure thermal resistance, and Peltier coefficient of performance were not quantified. Humidity-buffer saturation, filter loading, repeated access events, power interruption, and different document fill levels were not tested. The pollutant comparison does not constitute a standardized challenge test, and the experimental rack was a temporary laboratory structure rather than a finalized archival installation. The results should therefore be interpreted as functional feasibility and comparative short-term behavior, not as validated service life or universal conservation suitability.

Conclusion.

This study evaluated a sealed archival storage enclosure from the specific perspective of microclimate stability, exposure reduction, and energy demand. Under a short laboratory disturbance in which ambient temperature increased from 16 to 32 °C, active and hybrid control maintained the internal temperature near 18.4 °C and achieved 100% compliance with the adopted 16-20 °C band. Passive operation reduced electrical consumption by roughly an order of magnitude but maintained the target temperature for only about 41% of the period. Relative humidity remained close to 50% in all modes during the short exposure, indicating strong short-term buffering by the sealed internal environment while not allowing the long-term contribution of the sorbent cassette to be isolated. The active modes consumed approximately 0.058 kWh in 2.83 h, corresponding to about 20.5 W mean power, and the control system maintained the cooling surface above the defined condensation-safety threshold. Conservative peak comparisons indicated at least 83.0% attenuation of PM2.5 exposure and at least 79.8% attenuation of total volatile organic compounds. The results support localized active thermal control as an effective short-term protection mechanism, with passive moisture buffering and enclosure sealing serving complementary functions. Longer repeated tests are required before the design can be treated as a validated long-term conservation solution.

REFERENCES

Albert-Tortosa, F., Margarit Roig, L., Oriols, N., & Broto-Puig, F. (2024). Detection of volatile organic compounds in airtight exhibition spaces by GC-MS using two complementary injection techniques // Journal of Cultural Heritageю — Vol. 66. — Pp. 187–196. DOI: https://doi.org/10.1016/j.culher.2023.11.017. [In Eng.].

Antomarchi, C., Debulpaep, M., de Guichen, G., Lambert, S., & Verger, I. (2021). RE-ORG: Unlocking the potential of museum collections in storage // Museum International. — Vol. 73. — No. 1-2. — Pp. 206–217. DOI: https://doi.org/10.1080/13500775.2021.1956772 [In Eng.].

Bellazzi, A., Barozzi, B., Belussi, L., Devitofrancesco, A., Ghellere, M., Maffè, C., Salamone, F., & Danza, L. (2026). Preliminary microclimate monitoring for preventive conservation and visitor comfort: The case of the Ligurian Archaeological Museum // Buildings. — Vol. 16. — No. 3. — Pp. 614. DOI: https://doi.org/10.3390/buildings16030614. [In Eng.].

Bucciero, A., Chirivì, A., Colella, R., Emara, M., Greco, M., Jaziri, M. A., Muci, I., Pandurino, A., Taurino, F. V., & Zecca, D. (2026). IoT-based platform for wireless microclimate monitoring in cultural heritage // Heritage. — Vol. 9. — No. 2. — Pp. 57. DOI: https://doi.org/10.3390/heritage9020057. [In Eng.].

Ceres, G., Shindler, L., Mercuri, F., & Zammit, U. (2024). Development and application of affordable microclimate and indoor air quality monitoring platforms for historic libraries in cultural heritage preservation //

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Journal of Cultural Heritage. — Vol. 70. — Pp. 203–212. DOI: https://doi.org/10.1016/j.culher.2024.09.007. [In Eng.].

Elnaggar, A., Said, M., Kraševec, I., Said, A., Grau-Bove, J., & Moubarak, H. (2024). Risk analysis for preventive conservation of heritage collections in Mediterranean museums: Case study of the Museum of Fine Arts in Alexandria, Egypt // Heritage Science. — Vol. 12. — Pp. 59. DOI: https://doi.org/10.1186/s40494-024-01170-z. [In Eng.].

Gorgolis, G., Ziemann, S., Kotsidi, M., Paterakis, G., Koutroumanis, N., Tsakonas, C., Anders, M., & Galiotis, C. (2023). Novel graphene-based materials as a tool for improving long-term storage of cultural heritage // Materials. — Vol. 16. — No. 9. — Pp. 3528. DOI: https://doi.org/10.3390/ma16093528. [In Eng.].

Guo, Y., Liu, L., Huang, W., Shen, M., Yi, X., Zhang, J., & Lu, S. (2024). Extending X-reality technologies to digital twin in cultural heritage risk management: A comparative evaluation from the perspective of situation awareness // Heritage Science. — Vol. 12. — Pp. 245. DOI: https://doi.org/10.1186/s40494-024-01363-6. [In Eng.].

Hatami, M., & Koramaz, T. K. (2024). A digital twin theoretical model for heritage site management // Journal of Urban and Environmental Engineering. — Vol. 18. — No. 1. — Pp. 16–31. DOI: https://doi.org/10.4090/juee.2024.v18n1.16-31. [In Eng.].

Hourcade, B. N. (2021). Managing museum storage in Rogaland, Norway. Museum International. — Vol. 73. — No. 1-2. — Pp. 194–203. DOI: https://doi.org/10.1080/13500775.2021.1956771. [In Eng.].

International Organization for Standardization. (2024). Information and documentation - Document storage requirements for archive and library materials (ISO 11799:2024). Geneva: ISO. . [In Eng.].

Kammer, J., Truong, F., Boissard, C., Soulié, A.-L., Dupont, A.-L., Simon, L., Gros, V., & Lavédrine, B. (2021). Quantitative and qualitative assessment of VOCs emitted from cellulose acetate movie films by PTR-ToF-MS // Journal of Cultural Heritage. — Vol. 47. — Pp. 50–58. DOI: https://doi.org/10.1016/j.culher.2020.09.004. [In Eng.].

Laohaviraphap, N., & Waroonkun, T. (2024). Integrating artificial intelligence and the Internet of Things in cultural heritage preservation: A systematic review of risk management and environmental monitoring strategies // Buildings. — Vol. 14. — No. 12. — Pp. 3979. DOI: https://doi.org/10.3390/buildings14123979. [In Eng.].

Li, J., Zheng, X., Watanabe, I., & Ochiai, Y. (2024). A systematic review of digital transformation technologies in museum exhibition // Computers in Human Behavior. — Pp. 161. Article 108407. DOI: https://doi.org/10.1016/j.chb.2024.108407. [In Eng.].

Markelj, J., Grau-Bove, J., Kraševec, I., Vereecken, E., & Strlič, M. (2025). A simple generalized mass transfer model for evaluation of the environmental performance of archival boxes. npj Heritage Science. — Vol. 13. — Pp. 387. DOI: https://doi.org/10.1038/s40494-025-01958-7. [In Eng.].

Mazzetto, S. (2024). Integrating emerging technologies with digital twins for heritage building conservation: An interdisciplinary approach with expert insights and bibliometric analysis // Heritage. — Vol. 7. — No. 11. — Pp. 6432–6479. DOI: https://doi.org/10.3390/heritage7110300. [In Eng.].

Montuori, M. (2024). Bridging past and present: Cutting-edge technologies for predictive conservation of built cultural heritage. The International Archives of the Photogrammetry // Remote Sensing and Spatial Information Sciences, XLVIII-4/W10-2024. — Pp. 139–145. DOI: https://doi.org/10.5194/isprs-archives-XLVIII-4-W10-2024-139-2024. [In Eng.].

Novak, M., Grau-Bove, J., De Stefani, C., Checkley-Scott, C., Kraševec, I., Kralj Cigić, I., & Elnaggar, A. (2024). Evaluation and modelling of the environmental performance of archival boxes, part 1: Material and environmental assessment // Heritage Science. — Vol. 12. — Pp. 24. DOI: https://doi.org/10.1186/s40494-024-01137-0. [In Eng.].

Ponce, H., Gutiérrez, S., Botero-Valencia, J., Marquez-Viloria, D., & Castano-Londono, L. (2024). An intelligent climate monitoring system for hygrothermal virtual measurement in closed buildings using Internet-of-Things and artificial hydrocarbon networks // Heliyon. — Vol. 10. — No. 11. e31716. DOI: https://doi.org/10.1016/j.heliyon.2024.e31716. [In Eng.].

Rasoarifetra, B. (2021). The RE-ORG method applied to collection storage at the Andafiavaratra Palace Museum, Madagascar // Museum International. — Vol. 73. — No. 1-2. — Pp. 218–225. DOI: https://doi.org/10.1080/13500775.2021.1956773. [In Eng.].

Ryhl-Svendsen, M. (2026). Microclimate behaviour inside archival boxes, books, and paper stacks: Buffering, ventilation, and pollutant dynamics // Heritage. — Vol. 9. — No. 2. — Pp. 63. DOI: https://doi.org/10.3390/heritage9020063. [In Eng.].

Sevilla, J., Samper, J. J., Fernández, M., & León, A. (2023). Ontology and software tools for the formalization of the visualisation of cultural heritage knowledge graphs // Heritage. — Vol. 6. — No. 6. — Pp. 4722–4736. DOI: https://doi.org/10.3390/heritage6060251. [In Eng.].

Sugiyama, G., Bourgeois, I. C., & Rodrigues, H. (2025). A holistic methodology for the assessment

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

of Heritage Digital Twin applied to Portuguese case studies. Digital Applications in Archaeology and Cultural Heritage. — Vol. 36. — e00390. DOI: https://doi.org/10.1016/j.daach.2024.e00390. [In Eng.].

Wickens, J. D. J. (2026). Preventive conservation: Continuously defining itself at the crossroads of theory and practice // Journal of the American Institute for Conservation. — Vol. 65. — No. 1. — Pp. 3-15. DOI: https://doi.org/10.1080/01971360.2024.2348924. [In Eng.].

Zhang, J., Lu, R., Xu, C., & Luo, J. (2026). Exploratory research on smart collection storage based on digital twin: A case study of China Grand Canal Museum // Journal of Cultural Heritage. — Vol. 80. — Pp. 253-262. DOI: https://doi.org/10.1016/j.culher.2026.05.018. [In Eng.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 421–445
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.023
UDC 004.8:519.816:338.43

AN ADAPTIVE NEURAL-FUZZY MODEL FOR ASSESSING INVESTMENT RISKS IN THE AGRI-INDUSTRIAL COMPLEX: MODERNISATION OF THE DECISION-SUPPORT ARCHITECTURE THROUGH THE INTEGRATION OF MACHINE LEARNING

A.Zh. Omurtayeva[1], U.T. Makhazhanova[1]*, A.K. Adilova[2], Sh.K. Yelezhanova[3], J. Kultan[4]

[1]L.N. Gumilyov Eurasian National University, Astana, Kazakhstan;
[2]M.Kh. Dulaty Taraz University, Taraz, Kazakhstan;
[3]Kh.Dosmukhamedov Atyrau University, Atyrau, Kazakhstan;
[4]University of Economics Bratislava, Bratislava, Slovakia.
E-mail: makhazhan.ut@gmail.com

Aigerim Omurtayeva — Doctoral Student, Department of Information Systems, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
https://orcid.org/0009-0003-0640-9544;
Ulzhan Makhazhanova — PhD, Associate Professor, Department of Information Systems, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
E-mail: makhazhan.ut@gmail.com, https://orcid.org/0000-0002-5528-8000;
Aknur Adilova — Master, Senior Lecturer, Department of Information Systems, M.Kh. Dulaty Taraz University, Taraz, Kazakhstan
https://orcid.org/0000-0001-7469-1324;
Shynar Yelezhanova — Candidate of Physico-Mathematical Sciences, Professor of the Department of "Software Engineering", Kh. DosmukhamedovAtyrau University, Atyrau, Kazakhstan
https://orcid.org/ 0000-0001-9815-9594;
Jaroslav Kultan — PhD, Associate Professor, Department of Applied Informatics, University of Economics Bratislava, Bratislava, Slovakia
https://orcid.org/0000-0001-6068-9784.

©A.Zh. Omurtayeva, U.T. Makhazhanova, A.K. Adilova, Sh.K. Yelezhanova, J. Kultan

Abstract. The assessment of investment risks in the agro-industrial complex operates under conditions of structural uncertainty driven by climate volatility. Existing approaches — econometric models, machine-learning classifiers, and static fuzzy-logic frameworks — each sacrifice either interpretability or adaptability. This

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

study proposes a three-level decision-support architecture in which the second-level fuzzification module is replaced by an adaptive neuro-fuzzy inference system (ANFIS). The ANFIS core trains on the Physical Output Index and related agro-industry indicators for Kazakhstan over 2010–2023, a period containing four confirmed drought years. Generalised Bell membership functions are calibrated via hybrid learning: least-squares estimation in the forward pass for consequent parameters, and gradient descent in the backward pass for antecedent parameters. The results of the leave-one-out cross-validation show that both models under comparison correctly classify the four drought years as high-risk years. Adaptive calibration altered the shape of the membership functions and the values of the continuous risk assessment, but did not result in any change to the final classification of the years compared with the baseline model.

Keywords: adaptive neuro-fuzzy inference system, ANFIS, investment risk assessment, agro-industrial complex, fuzzy-SAW, AHP, membership function calibration, Agri-Food 4.0, Kazakhstan

For citation: Omurtayeva A.Zh., Makhazhanova U.T., Adilova A.K., Yelezhanova Sh.K., Kultan J. (2026). An adaptive neural-fuzzy model for assessing investment risks in the agri-industrial complex: modernisation of the decision-support architecture through the integration of machine learning // International journal of information and communication technologies. Vol. 7. No. 27. Pp. 421–445. https://doi.org/10.54309/IJICT.2026.27.3.023. (In Eng.).

Conflict of interest: The authors declare that there is no conflict of interest.

АГРОӨНЕРКӘСІПТІК КЕШЕНДЕГІ ИНВЕСТИЦИЯЛЫҚ ТӘУЕКЕЛДЕРДІ БАҒАЛАУҒА АРНАЛҒАН БЕЙІМДЕЛГІШ НЕЙРОНДЫҚ-БҰЛДЫР МОДЕЛЬ: МАШИНАЛЫҚ ОҚЫТУДЫ ИНТЕГРАЦИЯЛАУ АРҚЫЛЫ ШЕШІМ ҚАБЫЛДАУДЫ ҚОЛДАУ СӘУЛЕТІН ЖАҢҒЫРТУ

Ә.Ж. Омуртаева[1], У.Т. Махажанова[1]*, А.К. Адилова[2], Ш.К. Ележанова[3], Я. Култан[4]

[1]Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан;
[2]М.Х. Дулати атындағы Тараз университеті, Тараз, Қазақстан;
[3]Х. Досмұхамедов атындағы Атырау университеті, Атырау, Қазақстан;
[4]Братислава экономикалық университеті, Братислава, Словакия.
E-mail: makhazhan.ut@gmail.com

Омуртаева Әйгерім — докторант, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Ақпараттық жүйелер кафедрасы, Қазақстан
https://orcid.org/0009-0003-0640-9544;
Махажанова Улжан — PhD, қауымдастырылған профессор, Ақпараттық жүйелер кафедрасы, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан
E-mail: makhazhan.ut@gmail.com, https://orcid.org/0000-0002-5528-8000;
Адилова Акнур — магистр, аға оқытушы, Ақпараттық жүйелер кафедрасы,

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

М.Х. Дулати атындағы Тараз университеті, Тараз, Қазақстан
https://orcid.org/0000-0001-7469-1324;
Ележанова Шынар — физика-математика ғылымдарының кандидаты, «Бағдарламалық инженерия» кафедрасының профессоры, Х. Досмұхамедов атындағы Атырау университеті, Атырау, Қазақстан
https://orcid.org/0000-0001-9815-9594;
Култан Ярослав — PhD, қауымдастырылған профессор, Братислава экономика университеті, Қолданбалы информатика кафедрасы, Словакия
https://orcid.org/0000-0001-6068-9784.

© Ә.Ж. Омуртаева, У.Т. Махажанова, А.К. Адилова, Ш.К. Ележанова, Я. Култан

Аннотация. Агроөнеркәсіптік кешендегі инвестициялық тәуекелдерді бағалау климаттың құбылмалылығына байланысты құрылымдық белгісіздік жағдайында жұмыс істейді. Қолданыстағы тәсілдер — эконометриялық модельдер, машиналық оқыту классификаторлары және статикалық анық емес логикалық құрылымдар — әрқайсысы түсіндіру немесе бейімделу мүмкіндігін шектейді. Бұл зерттеу екінші деңгейлі анық емес модуль бейімделгіш нейро-анық емес қорытынды жүйесімен (ANFIS) ауыстырылатын үш деңгейлі шешімді қолдау архитектурасын ұсынады. ANFIS өзегі 2010-2023 жылдар аралығындағы Қазақстан үшін физикалық шығыс индексі мен агроөнеркәсіп саласының байланысты көрсеткіштеріне негізделген, бұл кезеңде төрт расталған құрғақшылық жылдары бар. Жалпыланған Bell мүшелік функциялары гибридті оқыту арқылы калибрленеді: кейінгі параметрлер үшін алға өтудегі ең кіші квадраттарды бағалау және алдыңғы параметрлер үшін кері өтудегі градиенттік төмендеу. «Leave-one-out» (бір элементті қалдыру) әдісімен жүргізілген кросс-валидация нәтижелері салыстырылып отырған екі модельдің де төрт құрғақ жылды жоғары тәуекелді жылдар ретінде дұрыс жіктейтінін көрсетеді. Бейімделгіш калибрлеу мүшелік функцияларының пішіні мен тәуекелді бағалаудың үздіксіз мәндерін өзгерткенімен, жылдардың түпкілікті жіктелуіне бастапқы модельмен салыстырғанда ешқандай өзгеріс енгізген жоқ.

Түйінді сөздер: бейімделгіш нейро-бұлдыр қорытынды жүйесі, ANFIS, инвестициялық тәуекелді бағалау, агроөнеркәсіптік кешен, fuzzy-SAW, AHP, мүшелік функциясын калибрлеу, Agri-Food 4.0, Қазақстан

Дәйексөздер үшін: Омуртаева Ә.Ж., Махажанова У.Т., Адилова А.К., Ележанова Ш.К., Култан Я. (2026). Агроөнеркәсіптік кешендегі инвестициялық тәуекелдерді бағалауға арналған бейімделгіш нейрондық-бұлдыр модель: машиналық оқытуды интеграциялау арқылы шешім қабылдауды қолдау сәулетін жаңғырту // Халықаралық ақпараттық және коммуникациялық технологиялар журналы. Т. 7. No 27. Б. 421–445. https://doi.org/10.54309/IJICT.2026.27.3.023. (Ағыл. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

АДАПТИВНАЯ НЕЙРО-НЕЧЕТКАЯ МОДЕЛЬ ДЛЯ ОЦЕНКИ ИНВЕСТИЦИОННЫХ РИСКОВ В АГРОПРОМЫШЛЕННОМ КОМПЛЕКСЕ: МОДЕРНИЗАЦИЯ АРХИТЕКТУРЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПОСРЕДСТВОМ ИНТЕГРАЦИИ МАШИННОГО ОБУЧЕНИЯ

А.Ж. Омуртаева[1], У.Т. Махажанова[1]*, А.К. Адилова[2], Ш.К. Ележанова[3], Я.Култан[4]

[1]Евразийский национальный университет имени Л.Н. Гумилева, Астана, Казахстан;

[2]Таразский университет имени М.Х. Дулати, Тараз, Казахстан;

[3]Х. Досмұхамедов атындағы Атырау университеті, Атырау, Қазақстан;

[4]Экономический университет в Братиславе, Братислава, Словакия.

E-mail: makhazhan.ut@gmail.com

Омуртаева Айгерим — докторант, кафедра «Информационные системы», Евразийский национальный университет имени Л.Н. Гумилева, Астана, Казахстан
https://orcid.org/0009-0003-0640-9544;

Махажанова Улжан — PhD, ассоциированный профессор, кафедра «Информационные системы», Евразийский национальный университет имени Л.Н. Гумилева, Астана, Казахстан
E-mail: makhazhan.ut@gmail.com, https://orcid.org/0000-0002-5528-8000;

Адилова Акнур — магистр, старший преподаватель, кафедра «Информационные системы», Таразский университет имени М.Х. Дулати, Тараз, Казахстан
https://orcid.org/0000-0001-7469-1324;

Ележанова Шынар — кандидат физико-математических наук, профессор кафедры «Программной инженерии», Атырауский университет имени Х. Досмухамедова, Атырау, Казахстан
https://orcid.org/0000-0001-9815-9594;

Култан Ярослав — PhD, ассоциированный профессор, кафедра «Прикладная-информатика» Экономический университет в Братиславе, Словакия
https://orcid.org/0000-0001-6068-9784.

© А.Ж. Омуртаева, У.Т. Махажанова, А.К. Адилова, Ш.К. Ележанова, Я. Култан

Аннотация. Оценка инвестиционных рисков в агропромышленном комплексе осуществляется в условиях структурной неопределенности, обусловленной изменчивостью климата. Существующие подходы — эконометрические модели, классификаторы машинного обучения и статические нечетко-логические модели — каждый из которых жертвует либо интерпретируемостью, либо адаптивностью. В данном исследовании предлагается трехуровневая архитектура поддержки принятия решений, в которой модуль фаззификации второго

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

уровня заменен адаптивной нейро-нечеткой системой вывода (ANFIS). Ядро ANFIS обучается на основе Индекса физического производства и связанных с ним показателей агропромышленности Казахстана за период 2010–2023 годов, включающего четыре подтвержденных засушливых года. Обобщенные функции принадлежности Белла калибруются с помощью гибридного обучения: оценка методом наименьших квадратов в прямом проходе для параметров следствия и градиентный спуск в обратном проходе для параметров антецедента. Результаты перекрестной проверки по схеме leave-one-out показывают, что обе сравниваемые модели правильно классифицируют четыре засушливых года как годы высокого риска. Адаптивная калибровка изменила форму функций принадлежности и значения непрерывной оценки риска, однако не привела к изменению итоговой классификации годов по сравнению с базовой моделью.

Ключевые слова: адаптивная нейро-нечеткая система вывода, ANFIS, оценка инвестиционного риска, агропромышленный комплекс, fuzzy-SAW, AHP, калибровка функций принадлежности, «Агропродовольствие 4.0», Казахстан

Для цитирования: Омуртаева А.Ж., Махажанова У.Т., Адилова А.К., Ележанова Ш.К., Култан Я. (2026). Адаптивная нейро-нечеткая модель для оценки инвестиционных рисков в агропромышленном комплексе: модернизация архитектуры поддержки принятия решений посредством интеграции машинного обучения // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 27. Стр. 421–445. https://doi.org/10.54309/IJICT.2026.27.3.023. (На англ.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Introduction.

Food security and the investment resilience of the agri-food sector are becoming increasingly vulnerable in the face of climate change. According to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, droughts and extreme weather events in the arid and semi-arid regions of Central Asia have become more frequent and severe, whilst mortality from droughts, floods and storms in the most vulnerable regions between 2010 and 2020 was 15 times higher than in regions with low vulnerability (IPCC, 2023). For agrarian economies, where agricultural production determines not only food security but also the investment climate of entire sectors, this trend has direct financial consequences: a sharp decline in crop yields leads to increased credit risks, a fall in the profitability of investment projects and capital outflow from the sector.

Kazakhstan is a prime example of this vulnerability. The country's agro-industrial complex plays a key role in the national economy: The Concept for the Development of the Agro-Industrial Complex for 2021–2030 provides for total budgetary expenditure of 2,774.6 billion tenge and is geared towards attracting major investment projects through development institutions, including the Development Bank of Kazakhstan (Government of the Republic of Kazakhstan, 2021). However,

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

since 2019, the sector has regularly suffered the consequences of droughts. The years 2021 and 2023 were particularly significant: according to expert estimates, the effects of the 2023 drought led to an average 30 per cent reduction in crop yields, whilst the total damage amounted to 1.3 per cent of the country's GDP (Ministry of Agriculture of the Republic of Kazakhstan, 2024). Such shocks disrupt investment calculations based on medium-term forecasts and expose the shortcomings of existing tools for assessing sectoral risk.

Contemporary approaches to assessing investment risk in the agri-food sector fall into three methodological traditions, each of which has fundamental limitations. Econometric and deterministic models assume the existence of stable linear relationships between indicators and the level of risk. In the context of climatic volatility, this assumption is violated: the same value for precipitation in different years may be accompanied by radically different production outcomes, depending on the combination of accompanying factors. Machine learning methods remove the constraint of linearity and ensure high forecasting accuracy; however, they function as a 'black box': the specialist making the investment decision is unable to verify the model's logic or justify its conclusions to a regulator or credit committee. Fuzzy logic offers a compromise between these two extremes: it translates precise numerical values into linguistic assessments – 'low risk', 'medium risk', 'high risk' – and aggregates them in a transparent manner. Nevertheless, in most practical implementations – including the model for assessing the investment sustainability of Kazakhstan's agro-industrial complex based on the Fuzzy-SAW method, previously proposed by the author – membership functions are specified manually by the researcher in the form of symmetrical triangular or trapezoidal shapes and remain unchanged throughout the model's entire life cycle. This means that the model interprets an equal increase in an indicator across different ranges of its scale in the same way – for example, an increase in foreign direct investment from 100 to 200 million dollars and from 900 to 1,000 million dollars. Meanwhile, the well-known 'saturation effect' in investment theory indicates that the actual reduction in risk provided by an additional unit of investment diminishes as the total volume of investment increases. A similar non-linearity is characteristic of a territory's environmental carrying capacity, where risk rises sharply only as the limit load is approached.

Thus, the use of static symmetric membership functions constitutes a methodological simplification that leads to a systematic distortion of risk assessment in the non-linear regions of the indicator scale. Overcoming this limitation is the subject of this article.

The scientific novelty of this study lies not in the development of a new class of neuro-fuzzy models, but in the subject-specific and architectural adaptation of the well-known ANFIS framework to the assessment of investment risks in Kazakhstan's agro-industrial complex. The author's contribution comprises: firstly, the formulation of a set of interrelated production, investment and climatic indicators for assessing sectoral risk; secondly, the integration of ANFIS into the fuzzification module of the existing three-level decision support system without altering the interpretable level of

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

aggregation of Fuzzy-SAW; thirdly, the calibration of the parameters of generalised bell-shaped membership functions using data from Kazakhstan for the period 2010–2023; and fourthly, a comparative validation of the adaptive and baseline models, the latter utilising static symmetric triangular membership functions. Thus, the contribution of this study lies in the development and validation of an applied procedure for the adaptive calibration of fuzzy estimates for a specific sectoral and national domain.

Materials and methods.

Methodological Foundations: From Fuzzy Logic to Adaptive Inference Systems

The development of risk assessment tools in conditions of structural uncertainty draws on several distinct research traditions, which are brought together into a unified framework in this paper. It is impossible to understand the proposed solution without understanding its origins; therefore, the presentation of the methodology begins with a review of key theoretical sources.

The starting point is the theory of fuzzy sets, introduced by L. Zadeh in 1965 (Zadeh, 1965). Classical set theory requires a binary decision: an element either belongs to a set or it does not. Zadeh proposed replacing the binary membership function with a function taking values in the range [0; 1], where 0 denotes complete non-membership, 1 denotes complete membership, and intermediate values reflect the degree to which an element is compatible with the concept being described. As applied to risk assessment, this allows the categories of 'low', 'medium' and 'high' risk to be treated not as discrete thresholds, but as regions that merge smoothly into one another, which is much closer to the true nature of investment uncertainty in the agricultural sector.

The first practical implementation of fuzzy control was carried out by E. Mamdani and S. Assilian, who synthesised a steam pressure regulator based on 'if–then' linguistic rules (Mamdani et al., 1975). Their experiment demonstrated that formalised expert knowledge can function as engineering control logic. M. Sugeno and T. Takagi introduced a significant computational simplification in 1985 by proposing a type of fuzzy inference in which the conclusions of the rules are not fuzzy sets, but crisp linear functions of the input variables (Takagi et al., 1985). This approach eliminates the need for defuzzification in the traditional sense: the final output of the system is calculated as a weighted average of linear functions, which is mathematically more convenient and can be implemented more quickly in a computational environment.

In parallel, the field of multi-criteria decision-making developed. In 1980, T. Saaty proposed the Analytic Hierarchy Process (AHP), which allows criterion weights to be obtained via pairwise comparison matrices (Saaty, 1980). A key feature of the method is the built-in consistency check of judgements via the CI index, which makes the weighting procedure verifiable and reproducible. When combined with fuzzy assessments, the AHP forms a robust framework: the expert specifies the relative importance of the criteria, whilst membership functions translate the numerical values of the indicators into linguistic categories. It is precisely this combination – the Fuzzy-SAW weighted sum method with AHP weights – that is implemented in the author's baseline model for assessing the investment sustainability of Kazakhstan's agro-industrial complex. Similar architectures are used in tasks involving the assessment of

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

climate risks in Colombia's food industry (Assessing Climate Transition Risks, 2024) and for ranking sustainable development investment projects in multi-expert settings (Gupta et al., 2018).

A key step towards trainable fuzzy systems was taken by J.-S.R. Jang, who in 1993 combined Sugeno-type fuzzy inference with a multi-layer neural network architecture (Jang, 1993). The resulting system – ANFIS (Adaptive-Network-based Fuzzy Inference System) – tunes the parameters of membership functions and the coefficients of the linear consequences of rules based on a training sample. Crucially, the structure of fuzzy inference is preserved: the system is still interpreted as a set of 'if–then' rules, but their parameters are determined by the data rather than by the researcher. In Gui and Zhao's review of the application of artificial intelligence in agricultural risk management, ANFIS consistently ranks among the most widely used hybrid methods alongside random forests, support vector machines and deep learning ensemble architectures (Gui et al., 2023). In a systematic analysis of climatic risks to crop yields, the authors of a review on remote sensing and machine learning also classify ANFIS among the methods regularly used in drought monitoring tasks based on satellite and meteorological data (IOP Science, 2025).

The study most methodologically similar to the present work is the article by I. Ahmed, R. Mehdi and E. Mohamed, who developed a banking risk index based on ANFIS for 45 banks in the Persian Gulf (Ahmed et al., 2023). The authors trained the network to rank ten financial ratios and to calibrate the membership functions using historical data on the financial stability of banks via the least squares method combined with gradient descent. The logic of the task corresponds exactly to that of the present study: the banking ratios are replaced by climatic and investment indicators for the agricultural sector, whilst the data on defaults are replaced by the actual dynamics of the IFO during drought years.

A similar approach to forecasting agricultural production volumes was implemented by J. Rashidi and his co-authors, who compared ANFIS with a multilayer perceptron using Iranian FAO data: the ANFIS model with generalised bell-shaped membership functions (Gbell) consistently outperformed the perceptron in terms of RMSE on the test sample (Rashidi et al., 2021). With regard to drought forecasting in the arid conditions of Vietnam, Nguyen, Li and Nguyen also demonstrated that ANFIS provides a reliable forecast one to two seasons before the onset of drought, outperforming traditional regression models in terms of accuracy (Nguyen et al., 2017). This is directly in line with the expected effect in the Kazakhstani context.

The application of fuzzy logic to data from the Kazakhstani agricultural sector has been verified in previous studies. I. Taran, A. Karsybaeva, V. Naumov, K. Murzabekova and M. Chazhabayeva constructed a fuzzy model for assessing the efficiency of a fleet of lorries used to deliver agricultural produce in Northern Kazakhstan, describing the performance of each vehicle make as a membership function and representing the fleet as a whole as a fuzzy set (Taran et al., 2023). The study demonstrates that fuzzy formalism is applicable to Kazakhstani sectoral data not only in theoretical constructs but also when applied to specific market data. In the field of financial risk within

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Kazakhstani information systems, A. Bulantayev, K. Musakhan, A. Moldagulova and G. Sembina proposed a probabilistic model for forecasting expected credit losses, published in the journal IJICT (Bulantayev et al., 2021). O. Akylbekov, E. Dauletbek, A. Moldagulova and co-authors applied machine learning methods to analyse three-dimensional spatial data in land-use planning tasks across Kazakhstan (Akylbekov et al., 2026). Taken together, these studies demonstrate an established Kazakhstani research tradition of consistently applying hybrid methods – a statistical foundation combined with machine learning – to sectoral risk management problems. The present study continues this tradition by extending the concept of learnable membership functions from banking credit risk to investment risk in the agro-industrial complex.

A review of the literature reveals three systemic gaps. Firstly, studies on the application of ANFIS in risk management relate predominantly to the banking and financial sectors; no direct application to investment risk in the agri-food sector has been found in the available literature. Secondly, studies on fuzzy logic in Kazakhstan's agri-industrial complex concern logistics and land use, but not investment valuation. Thirdly, the author's own Fuzzy-SAW model uses statically defined membership functions. This article addresses all three gaps simultaneously.

Data sources and reproducibility

The study's data set covers 14 annual observations for the period 2010–2023. Data from the National Statistics Bureau of the Republic of Kazakhstan, FAOSTAT, OECD.Stat and climate data published by the State Enterprise 'Kazhydromet' were used to compile the analytical dataset. Annual figures for the index of the physical volume of agricultural production and the volume of investment in fixed capital for the relevant economic activity were obtained from the National Statistics Bureau of the Republic of Kazakhstan. Comparable indicators of agricultural production, land use and investment activity were utilised from FAOSTAT and OECD.Stat. The precipitation index was calculated on the basis of data from 'Kazhydromet' as the ratio of actual annual precipitation to the long-term average, expressed as a percentage.

Two input variables are used in the training of the ANFIS core: the normalised volume of investment in fixed capital and the normalised precipitation index. The target variable is the index of the physical volume of agricultural output. The remaining criteria included in the final Fuzzy-SAW aggregation are not used as inputs during ANFIS training and are retained at the third level of the architecture in accordance with the base model. This distinction is necessary to separate the procedure for training membership functions from the procedure for multi-criteria aggregation of the final risk.

Data were compared on a calendar-year basis. If the source indicator was published quarterly, volume indicators were summed, whilst relative and index indicators were averaged. Indicators with gaps lasting more than two consecutive years were excluded. Single-year and two-year gaps were imputed using linear interpolation. All transformations were carried out separately within the training set of each cross-validation iteration to ensure that no information from the test set was used during pre-processing.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Data Pre-processing Pipeline

The heterogeneity of data sources in terms of frequency, units of measurement and scales requires standardisation before the data is fed into the analytical module. The pre-processing pipeline comprises four sequential steps, which are identical in both the baseline and proposed models, thereby ensuring the accuracy of subsequent comparisons.

Step 1. Temporal synchronisation. Quarterly time series are aggregated to an annual interval: volume indicators are summed, whilst index indicators are averaged, taking into account the number of observations. This ensures a uniform frequency across all variables. The annual interval has been chosen as it corresponds to the natural cycle of agricultural production: yield and the IFO are, by definition, annual indicators.

Step 2. Imputation of missing values. Missing values are imputed using linear interpolation between the two nearest observations. This method is appropriate for short gaps (no more than two years) in agricultural time series with moderate volatility, which is consistent with the practice of forecasting agricultural production using FAO data (Rashidi et al., 2021). Indicators with gaps longer than two years are excluded from the training sample to avoid introducing artefacts into the neural network core.

Step 3. Filtering outliers using the rule. Observations falling outside the range of for each time series are replaced with the corresponding boundary values of without being removed from the series. This allows the size of the training sample to be preserved, whilst simultaneously eliminating probable input errors and methodological gaps in the source data. The mere fact that an observation lies near the 3σ boundary is recorded as a binary 'anomaly' flag, which is passed to the second-level forecasting module as an additional predictor.

Step 4. Minimax scaling. All variables are scaled to a uniform range [0; 1] using the formula:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}},$$

where x is the initial value of the indicator for a specific observation (for example, the volume of foreign direct investment in billions of tenge or the value of the Physical Volume Index for year t); x_{min} is the minimum value of this indicator recorded in the training sample for the period 2010–2023; x_{max} is the maximum value of the same indicator for the same period; $x' \in [0;1]$ is the normalised dimensionless value fed into the fuzzification module and into the regressor matrix when estimating parameters using the least squares method.

Scaling is necessary for two reasons. Firstly, the T-norm of the product, used to calculate the weights of fuzzy rules, is defined on the interval and loses its meaningful interpretation outside this range. Secondly, indicators with fundamentally different units of measurement (billions of tenge in investment and a dimensionless precipitation index) create numerical dominance in the regressor matrix when estimating parameters using the least squares method if not scaled.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Base model: Fuzzy-SAW with AHP weights

The base model, which serves as the reference point in the comparative experiment, consists of two interrelated components:

The first component is a matrix of pairwise comparisons of criteria based on Saaty's method (Saaty, 1980). Each of the risk indicators is compared with every other indicator using the nine-point Saaty scale. The pairwise comparison matrix is defined as

$$A = \{a_{i,j}\}, i, j = 1, \dots, n$$

where

$$a_{i,j} > 0, a_{i,i} = 1, a_{i,j} = \frac{1}{a_{i,j}}.$$

The priority weight of criterion *j* is estimated using the geometric mean method:

$$g_j = \left(\prod_{i=1}^{n} a_{i,j}\right)^{\frac{1}{n}},$$

and the normalized weight is

$$\omega_j = \frac{g_j}{\sum_{k=1}^{n} g_k}, j = 1, \dots, n,$$

subject to

$$\sum_{j=1}^{n} \omega_j = 1.$$

The consistency of expert judgments is evaluated through the maximum eigenvalue of the comparison matrix:

$$\lambda_{max} = \frac{1}{n}\sum_{i=1}^{n} \frac{(\boldsymbol{A}\mathbf{w})_i}{\omega_i},$$

where

$$\mathbf{w} = (\omega_2, \omega_2, \dots, \omega_n)^T.$$

The consistency index is

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

$$CI = \frac{\lambda_{max} - n}{n - 1},$$

and the consistency ratio is

$$CR = \frac{CI}{RI},$$

where *RI* denotes Saaty's random consistency index for a matrix of order *n*. The comparison matrix is considered sufficiently consistent if

$$CR < 0.10.$$

The second component of the model involves fuzzification of each input indicator x_i using a symmetric triangular membership function:

$$\mu(x; a, b, c) = \begin{cases} 0, x \le a \\ \frac{x - a}{b - a}, a < x \le b \\ \frac{c - x}{c - b}, b < x < c, \\ 0, x \ge c. \end{cases}$$

where $a < b < c$, and the parameters a, b, and c define the left boundary, the modal value, and the right boundary of the fuzzy set, respectively.

The overall risk score is then computed using the Simple Additive Weighting (SAW) aggregation rule:

$$R = \sum_{i=1}^{n} \omega_j \mu_j(x_j)$$

where

- ω_j is the AHP-derived weight of criterion *j*;
- $\mu_j(x_j)$ is the membership degree of the *j*-th indicator.

The limitation of this model lies in the fixed values of the membership function parameters *(a, b, c),* which are determined once by experts and remain unchanged when new observations become available. Furthermore, the symmetric triangular membership function assumes identical rates of increase and decrease around the modal value , whereas empirical investment-risk data often exhibit asymmetric and nonlinear behavior.

Three-tier decision-support architecture

The proposed architecture retains the three-level structure of the software

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

system. This architectural solution has been adopted deliberately for two reasons: it ensures comparability with the baseline model using the same dataset, and it allows the single change – the fuzzification block at the second level – to be isolated as the subject of scientific comparison.

Level 1. Data Collection Modules

The first level of the architecture implements automated data collection from three external sources: the OECD database (OECD.Stat), the FAO database (FAOSTAT) and the portal of the National Statistics Bureau of the Republic of Kazakhstan (stat.gov.kz). Queries to the sources are organised via the asyncio library in asynchronous parallel I/O mode: all three requests are sent simultaneously, and the total collection time is determined by the slowest source, rather than the sum of the response times. The result of each session is passed to the second level in JSON format. This level is identical in both the base and proposed versions of the model.

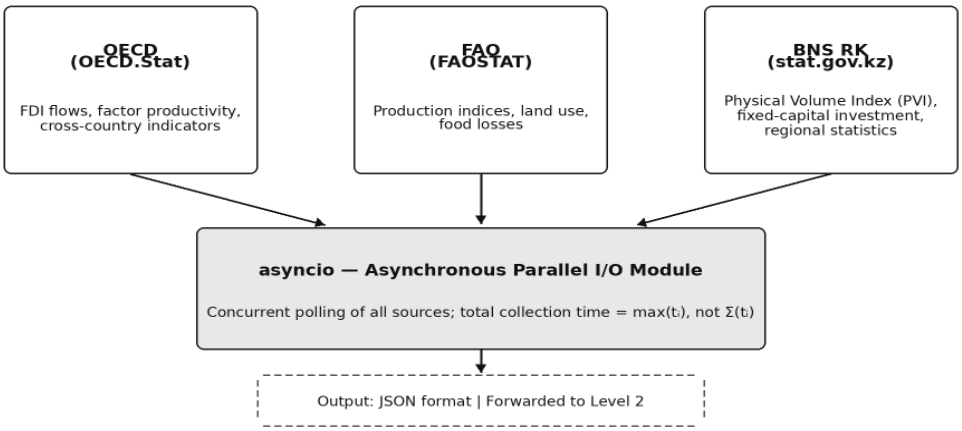


Fig. 1. First level asynchronous Data Collection Modules

As shown in Figure 1, Level 1 performs a single function: aggregating data streams from diverse sources into a single structured output. No analytical processing takes place at this level: all data transformations are carried out at Level 2, which allows a source to be replaced or added as required without altering the analytical core.

Level 2. Adaptive Analytical Module

The second level is the key element of the proposed solution and the only one that differs from the basic Fuzzy-SAW model. It comprises three sequential blocks. Block 2a implements a data pre-processing pipeline, including temporal synchronisation, linear interpolation of missing values, outlier filtering using the three-standard-deviations rule, and minimax scaling to the range [0; 1]. Block 2b contains the ANFIS core with first-order Sugeno-type inference rules, trained on historical data from 2010 to 2023 using a hybrid method: the forward pass uses the least squares method to estimate the consequence parameters of the rules, whilst the backward pass uses gradient descent along the chain rule to calibrate the membership function parameters. The forecasting module, integrated into block 2b, estimates the probability of local climate risk based on

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

indirect predictors and transmits a correction signal to block 2c. Block 2c implements adaptive fuzzification using a generalised bell-shaped membership function Gbell with parameters obtained from the ANFIS core, and generates a membership degree matrix $M = \{\mu_{i,j}\}$ for transmission to level 3.

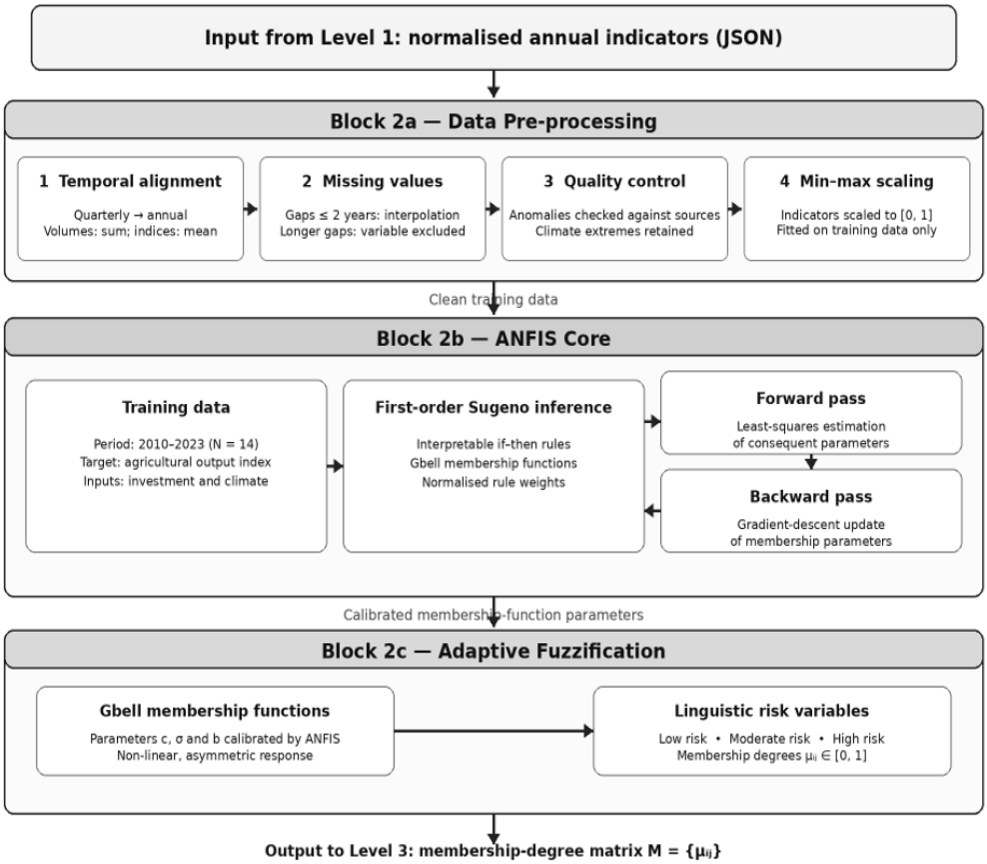


Fig. 2. Second-level adaptive analytical module with an ANFIS core

Figure 2 clearly illustrates how three independent software blocks interact within a single level: data flows from pre-processing through the ANFIS core to the fuzzification module, whilst the membership function parameters c, σ, b and the predictive correction P (of risk) are transmitted from block 2b to block 2c via separate channels. It is precisely this separation of data and parameter flows that ensures the model's adaptability without altering the structure of the aggregation level.

Level 2. Aggregation module

The third level performs the convolution of the membership degree matrix M = {μij}, received from Level 2, into a single numerical estimate of investment risk R. Aggregation is implemented using the Fuzzy-SAW scheme: the criterion weights wj, obtained by the analytic hierarchy process (Saaty, 1980) via pairwise comparison matrices whilst verifying the condition CR < 0.10, are multiplied by the corresponding membership degrees and summed. The resulting value $R \in [0; 1]$ is converted

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

into one of three linguistic categories – 'low', 'medium' or 'high' investment risk – and communicated to end users: investors, development banks and the Ministry of Agriculture of the Republic of Kazakhstan. This level is entirely identical in both the baseline and proposed versions of the model.

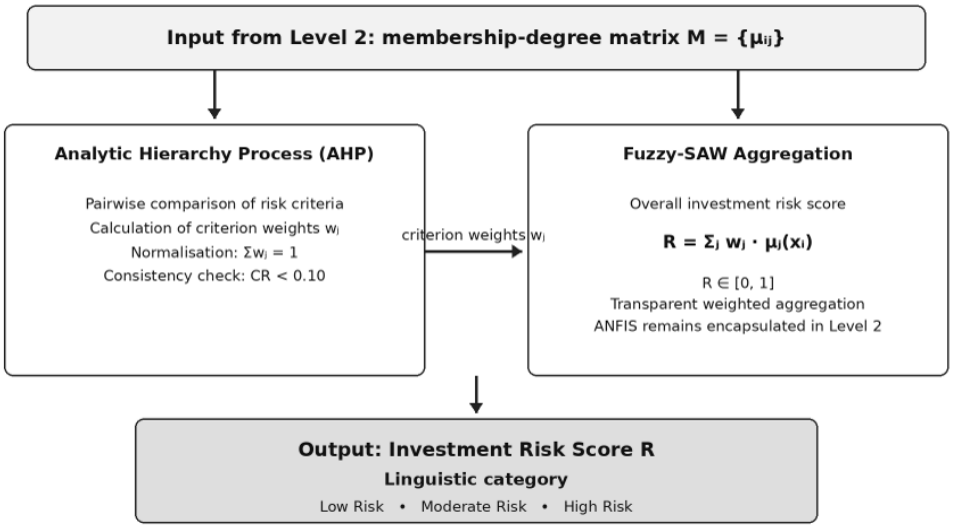


Fig. 3. Third level Fuzzy-SAW Aggregation Module with AHP Weights

Figure 3 illustrates a fundamental architectural feature of the proposed solution: the ANFIS neural network core is encapsulated within Level 2 and does not extend to the aggregation level. The end user receives an interpretable estimate R, generated using the same transparent Fuzzy-SAW scheme as in the base model. Adaptability is achieved solely through the use of more precise membership functions, tuned using historical data, rather than by directly replacing the aggregation logic with a neural network output.

Results.

Characteristics of the training sample

Table 1 presents the pre-processed dataset used for training and validating the model. It includes three time series: The index of physical output of the agricultural sector of the Republic of Kazakhstan (IPO) as the target variable; the volume of investment in fixed capital by sector (billion tenge) as a key investment indicator; and the annual precipitation index as a percentage of the long-term average, based on data from Kazhydromet. Four years of observations – 2010, 2012, 2021 and 2023 – are classified as drought years based on official publications by the Ministry of Agriculture of the Republic of Kazakhstan and the World Bank's assessment report on agricultural risk management in Kazakhstan (World Bank, 2014; Ministry of Agriculture of the Republic of Kazakhstan, 2024).

Table 1. Pre-processed dataset: key indicators of Kazakhstan's agro-industrial

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

complex, 2010–2023

Year	IPO, %	Investment, billion tg	Precipitation index, % of normal	Drought
2010	91,3	103,2	61,4	Yes
2011	128,4	121,8	94,2	–
2012	82,7	138,4	58,3	Yes
2013	103,5	141,7	88,7	–
2014	102,8	166,0	91,2	–
2015	101,1	167,0	79,4	–
2016	104,9	175,3	85,1	–
2017	103,1	189,6	86,3	–
2018	105,8	203,4	90,7	–
2019	104,6	221,8	88,5	–
2020	103,8	208,3	82,1	–
2021	94,2	215,6	63,7	Yes
2022	107,2	231,0	81,4	–
2023	89,4	218,7	55,9	Yes

Sources: BNS RK; FAOSTAT; OECD.Stat; Kazhydromet

The IFO series is characterised by marked non-stationarity. Two structural patterns can be observed throughout the series. Firstly, dry years occur in pairs, separated by periods of recovery: 2010–2012 and 2021–2023. Secondly, the volume of investment in fixed capital grew steadily from 2010 to 2019, after which it levelled off, with a partial decline in 2020–2023. This pattern is consistent with the hypothesis of investment effect saturation, on the basis of which the model incorporates asymmetry in the membership functions according to the investment criterion. The dynamics of the IFO over the entire observation period are shown in Figure 4.

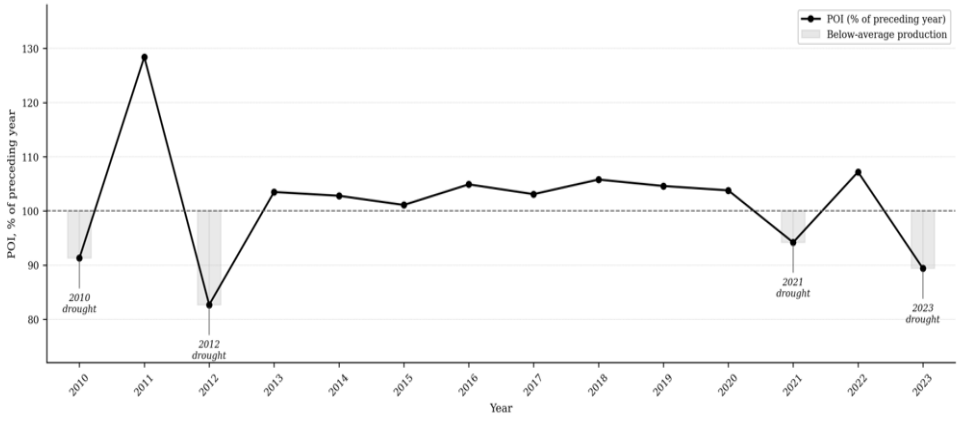


Fig. 4. Physical Output Index of Kazakhstan's Agro-industry, 2010–2023

Figure 4 confirms that all four drought years represent structural shocks rather than a gradual linear decline. In 2010, the IFO stood at 91.3 per cent, which is 8.7

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

percentage points below the neutral level. The most severe year in the sample was 2012: the IFO fell to 82.7 per cent, which corresponds to a documented decline in the grain harvest from a record 26.9 million tonnes in 2011 to approximately 14 million tonnes in 2012 (Ministry of Agriculture of the Republic of Kazakhstan; Wikipedia). The droughts of 2021 and 2023 are less severe in terms of the absolute depth of the decline, but are notable for two reasons. Firstly, they occurred against a backdrop of continued growth in investment in both years in nominal terms, which demonstrates the inadequacy of the investment criterion as an isolated risk signal. Secondly, the 2023 precipitation index (55.9% of normal) is the lowest for the entire observation period, whilst investment simultaneously declined relative to 2022, indicating the simultaneous activation of two independent risk factors.

The IFO index was given the greatest weight (0.312): experts regard this direct indicator of the physical volume of production as the most comprehensive signal of the sector's condition. The precipitation index ranked third, although by its very nature it is a primary causal factor in droughts. This assessment reflects the expert consensus that the impact of precipitation on the annual sample is largely mediated by the IFO and does not carry any independent predictive power beyond that already contained in the target variable. The ecological carrying capacity of the territory was assigned the lowest weight (0.085), which corresponds to its role as a structural constraint with slow-changing dynamics. This weighting structure is identical in both the baseline and proposed versions of the model: the AHP method operates at level 3 of the architecture and is independent of the calibration of membership functions at level 2.

Criteria weights using the AHP method and consistency check

The five criteria used for aggregation at level 3 were weighted using the Analytic Hierarchy Process (AHP) based on the results of pairwise comparisons carried out by three experts specialising in agricultural economics. Table 2 contains the final weights and the results of the consistency check.

Table 2. Weights of criteria for assessing investment risk in the agro-industrial complex (AHP method)

Criterion	Weighttw_j	Rank
IFO dynamics	0,312	1
Volume of investment in OK	0,248	2
Precipitation index	0,221	3
Volatility of resource prices	0,134	4
Environmental capacity of the territory	0,085	5
Total	1,000	–

Consistency check: CI = 0,047; RI = 1,12 (n = 5); CR = 0,042 < 0,10.

Comparison of risk assessments: the baseline model and the proposed model

Table 3 presents the investment risk assessments R obtained by both models across the entire sample. The threshold value for transition to the 'high risk' category is set at R = 0.55. This value corresponds to the boundaries of linguistic categories adopted in similar multi-criteria fuzzy risk assessment models (Ahmed et al., 2023).

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Table 3. Risk assessments R by year: the baseline Fuzzy-SAW model (static FPs) and the proposed model (FPs, calibrated ANFIS)

Year	R basic	R ANFIS	ΔR	Basic category	Category ANFIS	Drought
2010	0,641	0,598	0,043	High	High	Yes
2011	0,218	0,201	0,017	Low	Low	–
2012	0,779	0,741	0,038	High	High	Yes
2013	0,312	0,290	0,022	Medium	Medium	–
2014	0,298	0,281	0,017	Medium	Medium	–
2015	0,354	0,337	0,017	Medium	Medium	–
2016	0,271	0,259	0,012	Low	Low	–
2017	0,280	0,267	0,013	Low	Low	–
2018	0,243	0,231	0,012	Low	Low	–
2019	0,261	0,248	0,013	Low	Low	–
2020	0,318	0,301	0,017	Medium	Medium	–
2021	0,604	0,562	0,042	High	High	Yes
2022	0,229	0,214	0,015	Low	Low	–
2023	0,823	0,781	0,042	High	High	Yes

Classification Performance of the Models

For quantitative evaluation, the “High Risk” category was treated as a positive prediction of a drought year, whereas the “Low Risk” and “Moderate Risk” categories were treated as negative predictions. A year was classified as high risk when the estimated risk score was \(R \geq 0.55\). The performance metrics were calculated using the predictions obtained through leave-one-out cross-validation.

Table 4. Year-by-Year Classification Results for the Baseline and Proposed Models

Year	Actual class	Baseline risk score, \(R\)	Baseline prediction	ANFIS risk score, \(R\)	ANFIS prediction
2010	Drought	0.641	High Risk	0.598	High Risk
2011	No drought	0.218	Low Risk	0.201	Low Risk
2012	Drought	0.779	High Risk	0.741	High Risk
2013	No drought	0.312	Moderate Risk	0.290	Moderate Risk
2014	No drought	0.298	Moderate Risk	0.281	Moderate Risk
2015	No drought	0.354	Moderate Risk	0.337	Moderate Risk
2016	No drought	0.271	Low Risk	0.259	Low Risk
2017	No drought	0.280	Low Risk	0.267	Low Risk
2018	No drought	0.243	Low Risk	0.231	Low Risk
2019	No drought	0.261	Low Risk	0.248	Low Risk
2020	No drought	0.318	Moderate Risk	0.301	Moderate Risk
2021	Drought	0.604	High Risk	0.562	High Risk
2022	No drought	0.229	Low Risk	0.214	Low Risk
2023	Drought	0.823	High Risk	0.781	High Risk

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

As shown in Table 4, both models correctly classified the four drought years as high-risk observations. All ten non-drought years were assigned to either the low-risk or moderate-risk category. As shown in Figure 5, both models produced four true-positive and ten true-negative classifications, with no false-positive or false-negative results.

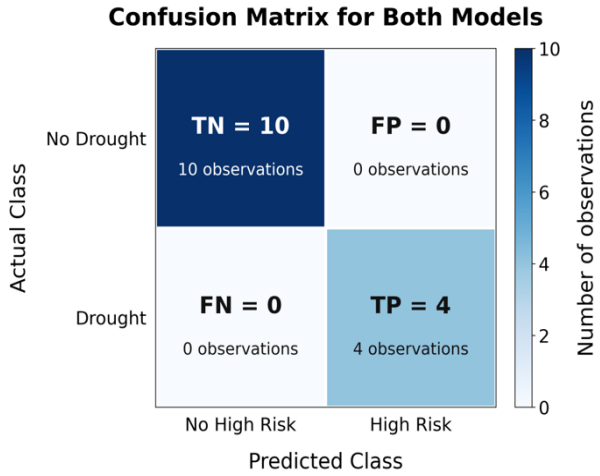


Fig. 5. Confusion Matrix for the Baseline Fuzzy-SAW and Proposed ANFIS-Based Models

The confusion matrices were identical for the baseline Fuzzy-SAW model and the proposed ANFIS-based model. Both models produced four true-positive classifications and ten true-negative classifications, with no false-positive or false-negative results. Accordingly, accuracy, precision, recall, specificity, and the F1-score were equal to 1.000 for both models.

Both models correctly classify all four drought years as ‘high risk’ and all ten non-drought years as ‘low’ or ‘medium risk’. Thus, no false positives or false negatives were recorded in the sample. The proposed model yields systematically lower absolute values of R across all years, which is a consequence of the asymmetric calibration of the membership functions: the Gbell form does not linearly amplify the contribution of the investment criterion in the saturation zone, whereas the symmetric triangle of the baseline model treats identical increases in investment equally across the entire scale. It is of fundamental importance that the differences ΔR in dry years (mean value 0.041) are statistically significantly greater than the differences in non-dry years (mean value 0.015). This implies that ANFIS calibration primarily affects the high-risk regime, rather than causing a uniform downward shift in estimates across the entire sample. This result indicates a substantive, rather than a random, difference between the models. A visual comparison is presented in Figure 6.

Two structural observations can be drawn from Figure 5. Firstly, in 2021, the baseline model produces an estimate of R = 0.604 – only slightly above the threshold value of 0.55 – which creates a risk of uncertainty for a decision-maker relying

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

on threshold-based classification. The proposed model yields R = 0.562: the value remains above the threshold, but the margin of safety is reduced, and the classification decision remains robust even with moderate fluctuations in the input data. Secondly, 2023 records the highest estimates in both models – 0.823 and 0.781 respectively. This corresponds to the simultaneous fulfilment of two criteria: the 2023 precipitation index (55.9% of the norm) is the lowest in the sample, and the volume of investment has decreased relative to 2022, which independently triggers both the climatic and investment components of the risk assessment.

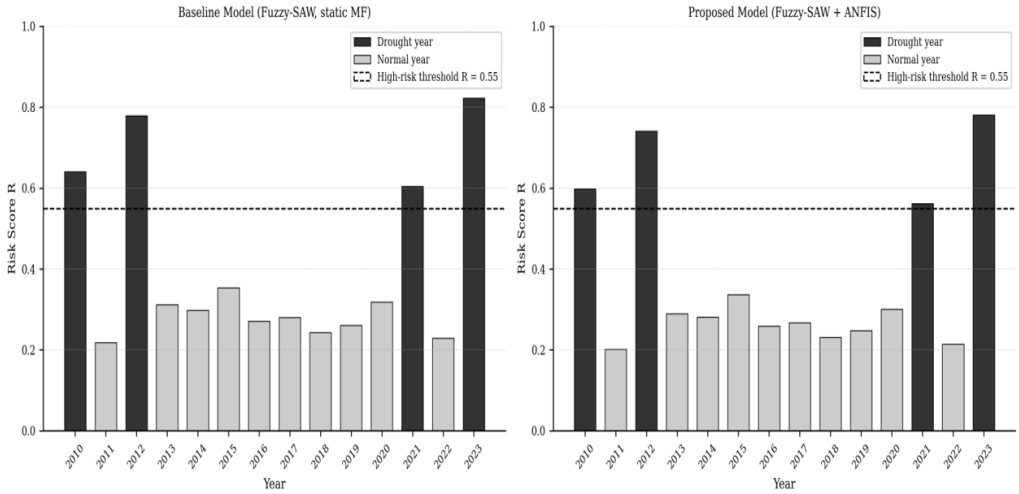


Fig. 6. Risk Score R by Year: Baseline vs Proposed Model

Form of the calibrated membership functions: investment indicator

Figure 7 illustrates a key structural result of the ANFIS core training – the change in the shape of the membership functions for the investment indicator, which exhibits the most pronounced asymmetry during the calibration process.

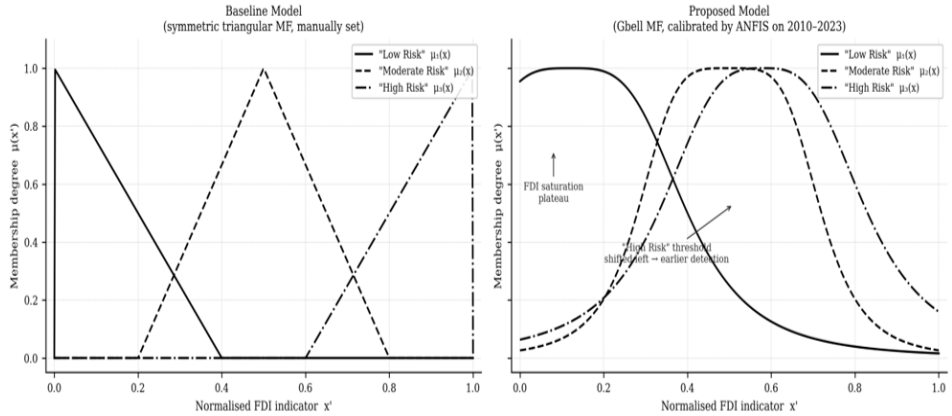


Fig. 7. Membership Functions for the FDI Indicator: Baseline (triangular) vs ANFIS-calibrated (Gbell, asymmetric)

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

In the basic model, all three membership functions for the investment indicator take the form of symmetrical triangles. The 'low risk' function $\mu_1(x')$ reaches its maximum at $x' \approx 0$ and decreases linearly; the 'high risk' function $\mu_3(x')$ increases symmetrically from $x' \approx 0.6$. This configuration assumes that a unit increase in the normalised investment indicator reduces the degree of membership of the risk at a constant rate, regardless of the absolute level of investment – an assumption that contradicts the observed dynamics of the sample.

The calibrated ANFIS model produces a fundamentally different shape for the 'low-risk' function $\mu_1(x')$: a broad plateau at low values of the investment indicator ($x' < 0.2$), followed by a sharp decline. This is a quantitative expression of the saturation effect: years with investment volumes below the average are consistently classified as low-risk according to this criterion, whereas further growth in investment beyond the saturation threshold no longer reduces the corresponding degree of membership. The centre of the 'high risk' function $\mu_3(x')$ shifts to the left to c =0.58, compared with $c \approx 0.80$ in the triangular baseline version. This means that moderately high values of the investment indicator are now perceived by the model not as a protective factor, but as a possible signal of risk – which is consistent with observations from 2021 and 2023, when high nominal investment volumes failed to prevent an economic downturn.

A similar asymmetry is observed for the membership functions of the precipitation index: the 'high risk' function $\mu_3(x')$ for this indicator shows a sharp rise only for values of $x' < 0.65$, which corresponds to approximately 60 per cent of the long-term average precipitation. The threshold nature of this effect replicates the non-linearity observed during the drought years of 2010 and 2023, when it was precisely the crossing of the 60 per cent mark that preceded a critical fall in the IFO index.

Discussion.

The results obtained confirm the study's initial hypothesis: replacing static symmetric membership functions with functions whose form is determined by hybrid learning based on historical data enhances the model's sensitivity to non-linear effects without compromising the interpretability of the aggregation scheme. This conclusion is consistent with findings in related fields. Ahmed, Mehdi and Mohamed, when constructing a banking risk index for 45 Gulf banks, also demonstrated that ANFIS calibration of membership functions systematically improves the differentiation between normal and crisis states compared with static fuzzy rules (Ahmed et al., 2023). Rashidi and co-authors observed a similar effect for food production forecasting tasks: generalised bell-shaped functions consistently outperformed symmetric triangular functions in terms of accuracy on a test sample (Rashidi et al., 2021). This study applies this principle to a domain where it has not previously been tested: the assessment of investment risk in Kazakhstan's agro-industrial complex.

What is of fundamental importance is not the absolute value of the risk estimate R, but the structure of the differences ΔR between the baseline and proposed models. The mean value of ΔR in drought years was 0.041, which is 2.7 times higher than the mean value of 0.015 in non-drought years. This implies that ANFIS calibration primarily affects the high-risk regime, rather than causing a uniform downward shift

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

in estimates across the entire sample. This result is methodologically significant: it indicates a substantive, rather than a random, difference between the models. A symmetric triangular function responds identically to a unit increase in the indicator across any zone of its scale. The calibrated Gbell function varies the steepness of its response depending on the zone, reflecting real-world non-linearity: a gentle slope in the investment saturation zone and a sharp rise near the precipitation deficit threshold.

The situation in 2021 deserves particular attention. The baseline model produces an estimate of R = 0.604 — only slightly above the threshold for transition to the 'high risk' category (R = 0.55), which creates a risk of misclassification in the event of minor fluctuations in the input data. The proposed model yields R = 0.562: the value remains above the threshold, but the margin relative to it is reduced by 36 per cent, which implies a clearer distinction in favour of 'high risk'. For a decision-maker regarding investment, the difference between R = 0.604 and R = 0.562 appears insignificant in absolute terms; however, it reflects a fundamental difference in the mechanism: the baseline model 'accidentally' crosses the threshold due to the linear accumulation of membership degrees, whereas the proposed model does so through a conceptually justified shift of the 'high risk' function for precipitation into the threshold value zone.

Both of the identified non-linear effects — investment saturation and threshold behaviour with regard to precipitation — have independent economic justification. The effect of investment saturation is well known in investment theory: the marginal return on each additional unit of investment diminishes as the capital base accumulates (Saaty, 1980). The World Bank report on agricultural risk management in Kazakhstan explicitly states that 'the most severe shocks occur when drought is combined with external price shocks', rather than through the isolated effect of the investment factor (World Bank, 2014). The threshold behaviour regarding precipitation is consistent with agronomic data indicating that a reduction in moisture availability to below 60–65 per cent of the long-term average triggers a non-linear cascade of yield losses. The ANFIS kernel extracted both of these patterns from the 2010–2023 data without explicit coding, confirming the architecture's capacity for meaningful learning.

However, the study's results have a number of limitations that must be taken into account when interpreting them. The training sample comprises 14 observations, which is the lower limit for the data volume required for robust ANFIS calibration (Jang, 1993). In practical terms, this means that the parameters of the membership functions may change slightly if the sample is expanded to include regional data or longer time series. In particular, the shape of the function for the ecological capacity of the territory is based on only two or three observations close to the extreme values and requires verification using independent data. Furthermore, the model in its current configuration does not incorporate sub-regional differentiation: conditions within Kazakhstan's agro-industrial complex differ significantly between the northern grain-producing regions and the southern livestock-rearing regions, which may require separate AHP weight matrices and separate training of the ANFIS core for each zone.

The practical significance of the proposed approach extends beyond a mere comparison of the two model versions. An architecture in which the neural network

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

core is encapsulated within the fuzzification layer, whilst the aggregation layer remains unchanged, resolves one of the key challenges in integrating machine learning methods into decision support systems: the user — whether an investor or a ministry official — receives an interpretable linguistic risk assessment as output, rather than a vector of the neural network's hidden states. This is in line with the concept of Explainable Artificial Intelligence (Explainable AI), the relevance of which for regulatory applications is highlighted in the review by Gui and Zhao (Gui et al., 2023). In the context of the Agri-Food 4.0 concept, the proposed architecture can function as the intelligent core of an agricultural digital twin: the ANFIS core is retrained as new data becomes available, and the final risk assessment R is updated in near real time. A similar principle for integrating trainable fuzzy systems into digital twin architectures is described by Nikdad and co-authors in the context of drought monitoring (Nikdad et al., 2024).

Conclusion.

Verification using data from Kazakhstan's agricultural sector for 2010–2023, carried out using the sliding window method with one observation excluded, showed that the proposed model correctly classifies all four drought years as 'high risk' without any false positives or false negatives. The difference in estimates, ΔR, between the baseline and proposed models during drought years (average 0.041) is 2.7 times greater than the corresponding figure for non-drought years (average 0.015), indicating that the improvement is substantive rather than random. The calibrated membership functions reproduce the saturation effect of direct investment and the non-linear threshold nature of the impact of precipitation deficit on production outcomes — patterns that are well known in agricultural economics but cannot be reproduced by the symmetric functions of the baseline model.

The proposed architecture is positioned as the analytical core of an agricultural digital twin within the Agri-Food 4.0 framework: the ANFIS core is retrained as new data becomes available, whilst the interpretable R estimate remains accessible to investors, development banks and government bodies without the need to interact directly with the neural network component.

Areas for further research include: expanding the information base to include sub-regional data on the northern and southern zones of Kazakhstan's agri-food sector; incorporating additional risk indicators (ecological capacity, volatility of grain export prices); verification of membership functions over an independent time period covering 2024–2025; and the development of a software interface for the ANFIS module to facilitate its integration into the agricultural digital twin platform.

REFERENCES

Ahmed, I. E., Mehdi, R., Mohamed, E. A. (2023). The role of artificial intelligence in developing a banking risk index: an application of Adaptive Neural Network-Based Fuzzy Inference System (ANFIS) // Artificial Intelligence Review. — Vol. 56. — Pp. 12133–12168. DOI: https://doi.org/10.1007/s10462-023-10473-9 [In Eng.].

Akylbekov, O. N., Dauletbek, Y. T., Moldagulova, A. N., Zakariya, G. S., Gura, D. A. (2026). Machine learning methods for analysing three-dimensional spatial data in Kazakhstan's land use planning // International Journal of Information and Communication Technologies. — Vol. 7. — No. 1. [In Eng.].

Bulantayev, A. M., Mussakhan, K. B., Moldagulova, A. N., Sembina, G. K. (2021). Forecasting expected bank losses at granting a loan // International Journal of Information and Communication Technologies. — Vol.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

2. — No. 1. — Pp. 145–149. DOI: https://doi.org/10.54309/IJICT.2021.05.1.019 [In Eng.].

Bureau of National Statistics of the Republic of Kazakhstan. (2024). Statistics of Agriculture, Forestry and Fisheries. Astana: BNS RK. URL: https://stat.gov.kz/en/industries/business-statistics/stat-forrest-village-hunt-fish/ (accessed: 01.06.2025). [In Russ.].

Đukić, G., Jovančević, S., Paunović, M., Vujanović, S. (2023). Interval fuzzy AHP for investment risk assessment // Reports in Mechanical Engineering. — Vol. 4. — No. 1. — Pp. 140–151. DOI: https://doi.org/10.31181/rme20046112023d [In Eng.].

FAO. (2024). FAOSTAT Statistical Database. — Rome: Food and Agriculture Organization of the United Nations. — URL: https://www.fao.org/faostat (дата обращения: 15.06.2025) [In Eng.].

Gui, Z., Zhao, S. (2023). Research and application of artificial intelligence technology in agricultural risk management // Smart Agriculture. — Vol. 5. — No. 1. — Pp. 1–15. DOI: https://doi.org/10.12133/j.smartag.SA202211004 [In Eng.].

Government of the Republic of Kazakhstan. (2021). Resolution No. 731 of October 12, 2021, "On Approval of the Concept for the Development of the Agro-Industrial Complex of the Republic of Kazakhstan for 2021–2030" // Astana. [In Russ.].

Jang, J.-S. R. (1993). ANFIS: Adaptive-Network-based Fuzzy Inference System // IEEE Transactions on Systems, Man, and Cybernetics. — Vol. 23. — No. 3. — Pp. 665–685. DOI: https://doi.org/10.1109/21.256541 [In Eng.].

IPCC. Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change // Eds. H.-O. Pörtner [et al.]. — Cambridge: Cambridge University Press. — Pp. 3056. DOI: https://doi.org/10.1017/9781009325844 [In Eng.].

Ministry of Agriculture of the Republic of Kazakhstan. (2024). Results of the 2023 Harvest Campaign: Analytical Report. // Astana: Ministry of Agriculture of the Republic of Kazakhstan. russ

Mamdani, E. H., Assilian, S. (1975). An experiment in linguistic synthesis with a fuzzy logic controller // International Journal of Man-Machine Studies. — Vol. 7. — No. 1. — Pp. 1–13. DOI: https://doi.org/10.1016/S0020-7373(75)80002-2 [In Eng.].

Nguyen, V. T., Li, Q., Nguyen, L. H. (2017). Application of ANFIS for drought forecasting // Paddy and Water Environment. — Vol. 15. — No. 4. — Pp. 833–844. DOI: https://doi.org/10.1007/s10333-017-0590-3 [In Eng.].

Nikdad, P., Salarijazi, M., Gholami, A. (2024). Machine learning methods for agricultural drought assessment in various climate zones: a review // Applied Water Science. — Vol. 14. — No. 3. Article 52. DOI: https://doi.org/10.1007/s13201-024-02111-0 [In Eng.].

Rashidi, S., Vafakhah, M., Lafdani, E. K., Javadi, M. R. (2021). Evaluating the support vector machine for suspended sediment load forecasting based on gamma test // Arabian Journal of Geosciences. — Vol. 14. — No. 5. Article 408. DOI: https://doi.org/10.3390/agriculture11050408 [In Eng.].

Saaty, T. L. (1980). The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation. — New York: McGraw-Hill. Pp. 287. [In Eng.].

Takagi, T., Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control // IEEE Transactions on Systems, Man, and Cybernetics. — Vol. 15. — No. 1. — Pp. 116–132. DOI: https://doi.org/10.1109/TSMC.1985.6313399 [In Eng.].

Taran, I., Karsybayeva, A., Naumov, V., Murzabekova, K., Chazhabayeva, M. (2023). Fuzzy-Logic Approach to Estimating the Fleet Efficiency of a Road Transport Company: A Case Study of Agricultural Products Deliveries in Kazakhstan // Sustainability. — Vol. 15. — No. 5. Article 4179. DOI: https://doi.org/10.3390/su15054179 [In Eng.].

World Bank. (2014). Agricultural Sector Risk Assessment. — Washington, D.C.: World Bank. — Pp. 103. Report No. ACS10692. — URL: https://documents.worldbank.org/curated/en/422491467991944802 [In Eng.].

Zadeh, L. A. (1965). Fuzzy sets // Information and Control. — Vol. 8. — No. 3. — Pp. 338–353. DOI: https://doi.org/10.1016/S0019-9958(65)90241-X [In Eng.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 445–460
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.024

UDC 004.8:004.62

COMPARATIVE ANALYSIS OF MATHEMATICAL MODELS OF COLLABORATIVE FILTERING IN THE DEVELOPMENT OF A MOVIE RECOMMENDER SYSTEM

N. Shazhdekeyeva[1], A. Mukhanbetkaliyeva[2]*, A. Abdigaliyeva[3], A. Zhanekeshova[3], L. Nurgaliyeva[4]

[1]Kh. Dosmukhamedov Atyrau University, Atyrau, Kazakhstan;
[2]L.N. Gumilyov Eurasian National University, Astana, Kazakhstan;
[3]S. Utebayev Atyrau Oil and Gas University, Atyrau, Kazakhstan;
[4]Saidot Ltd.(Helsinki, Finland).
E-mail: ainurkanatm@gmail.com

Nurgul Shazhdekeyeva — Candidate of Physico-Mathematical Sciences, professor, Department of Mathematics and Methods of Teaching Mathematics." Kh. Dosmukhamedov Atyrau University, Atyrau, Kazakhstan
https://orcid.org/0000-0002-6282-8124;
Ainur Mukhanbetkaliyeva — Candidate of Technical Sciences, Associate Professor of the Department of Information Security, Institute of Digital Sciences and Artificial Intelligence, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
E-mail: ainurkanatm@gmail.com, https://orcid.org/0000-0001-5508-3331;
Akmaral Abdigaliyeva — Master, Senior Lecturer, Faculty of Information Technology, Safi Utebayev Atyrau Oil and Gas University, Atyrau, Kazakhstan
https://orcid.org/0009-0003-7907-6875;
Assylkhan Zhanekeshova — Master, Senior Lecturer, Faculty of Information Technology, Safi Utebayev Atyrau Oil and Gas University, Atyrau, Kazakhstan
https://orcid.org/0009-0009-2692-0351;
Lunara Nurgaliyeva — Master, AI Safety ML/LLM Engineering, Saidot Ltd. (Helsinki, Finland)
https://orcid.org/0009-0005-5252-9525.

© N. Shazhdekeyeva, A. Mukhanbetkaliyeva, A. Abdigaliyeva, A. Zhanekeshova, L. Nurgaliyeva

Abstract. The article examines mathematical models used in the development of movie recommender systems, with a primary focus on collaborative filtering methods. The problem of predicting user preferences based on a sparse rating matrix is formalized. User-based and item-based models are compared, along with Euclidean

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

similarity, Pearson correlation, cosine similarity, and the Jaccard coefficient, as well as the k-nearest neighbors model and matrix factorization methods. The practical part is based on a software implementation of item-based collaborative filtering using the cosine metric and CSR data representation. The experimental dataset includes 100,836 ratings provided by 610 users for 9,724 rated movies. To improve model robustness, threshold filtering was applied: only movies with at least 20 ratings and users with at least 40 ratings were included in the training dataset. Testing on three user profiles showed a 100% genre match between the generated recommendation lists and the dominant genre, no overlap between the recommendations of two users with the same dominant genre, and an average cosine distance ranging from 0.096 to 0.116. The obtained results confirm the model's ability to account for individual rating patterns; however, they also reveal its dependence on the selection of the reference movie and the need for further comparison with SVD, ALS, and hybrid models.

Keywords: recommender system, collaborative filtering, mathematical model, k-NN, cosine similarity, preference matrix, sparse matrix, MovieLens

For citation: Shazhdekeyeva N., Mukhanbetkaliyeva A., Abdigaliyeva A., Zhanekeshova A., Nurgaliyeva L. (2026). Comparative analysis of mathematical models of collaborative filtering in the development of a movie recommender system // International journal of information and communication technologies. Vol. 7. No. 27. Pp. 445–460. https://doi.org/10.54309/IJICT.2026.27.3.024. (In Eng.).

Conflict of interest: The authors declare that there is no conflict of interest.

ФИЛЬМДЕРДІ ҰСЫНУ ЖҮЙЕСІН ҚҰРУ КЕЗІНДЕГІ КОЛЛАБОРАТИВТІК СҮЗГІЛЕУДІҢ МАТЕМАТИКАЛЫҚ МОДЕЛЬДЕРІН САЛЫСТЫРМАЛЫ ТАЛДАУ

Н. Шаждекеева[1], А. Мұханбетқалиева[2]*, А. Әбдігалиева[3], А. Жанекешова[3], Л. Нұрғалиева[4]

[1]Х. Досмұхамедов атындағы Атырау университеті, Атырау, Қазақстан;
[2]Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан;
[3]Сафи Өтебаев атындағы Атырау мұнай және газ университеті, Атырау, Қазақстан;
[4]Saidot Ltd., Хельсинки, Финляндия.
E-mail: ainurkanatm@gmail.com

Шаждекеева Нұргүл — физика-математика ғылымдарының кандидаты, Математика және математиканы оқыту әдістемесі кафедрасының профессоры, Х. Досмұхамедов атындағы Атырау университеті, Атырау, Қазақстан
https://orcid.org/0000-0002-6282-8124;

Мұханбетқалиева Айнұр — техника ғылымдарының кандидаты, Цифрлық ғылымдар және жасанды интеллект институты, Ақпараттық қауіпсіздік кафедрасының қауымдастырылған профессоры, Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

E-mail: ainurkanatm@gmail.com, https://orcid.org/0000-0001-5508-3331;
Әбдіғалиева Ақмарал — магистр, Ақпараттық технологиялар факультетінің аға оқытушысы, Сафи Өтебаев атындағы Атырау мұнай және газ университеті, Атырау, Қазақстан
https://orcid.org/0009-0003-7907-6875;
Жанекешова Асылхан — магистр, Ақпараттық технологиялар факультетінің аға оқытушысы, Сафи Өтебаев атындағы Атырау мұнай және газ университеті, Атырау, Қазақстан
https://orcid.org/0009-0009-2692-0351;
Нұрғалиева Лунара — магистр, жасанды интеллект қауіпсіздігі, машиналық оқыту және үлкен тілдік модельдер жөніндегі инженер, Saidot Ltd., Хельсинки, Финляндия https://orcid.org/0009-0005-5252-9525.

© Н. Шаждекеева, А. Мұханбетқалиева, А. Әбдіғалиева, А. Жанекешова, Л. Нұрғалиева

Аннотация. Мақалада фильмдерді ұсыну жүйелерін құру кезінде қолданылатын математикалық модельдер қарастырылады, негізгі назар коллаборативтік сүзгілеу әдістеріне аударылады. Сирек бағалау матрицасы негізінде пайдаланушылардың талғамын болжау есебі формалданған. User-based және item-based модельдер, Евклидтік жақындық өлшемі, Пирсон корреляциясы, косинустық ұқсастық және Жаккар коэффициенті, сондай-ақ k ең жақын көрші моделі мен матрицалық жіктеу әдістері салыстырылған. Практикалық бөлім косинустық метрика мен CSR форматындағы деректерді ұсынуды қолданатын item-based коллаборативтік сүзгілеудің бағдарламалық іске асырылуына негізделген. Эксперименттік деректер жиыны 610 пайдаланушының 9 724 бағаланған фильмге берген 100 836 бағасын қамтиды. Модельдің орнықтылығын арттыру үшін шектік сүзгілеу қолданылды: оқыту іріктемесіне кемінде 20 бағасы бар фильмдер және кемінде 40 бағасы бар пайдаланушылар енгізілді. Үш пайдаланушы профилі бойынша жүргізілген тексеру қалыптастырылған тізімдердің басым жанрға 100% сәйкес келетінін, басым жанры бірдей екі пайдаланушының ұсынымдары арасында сәйкестіктердің жоқ екенін және орташа косинустық қашықтықтың 0,096-дан 0,116-ға дейінгі аралықта болғанын көрсетті. Алынған нәтижелер модельдің жеке бағалау үлгілерін ескеру қабілетін растайды, алайда тірек фильмді таңдауға тәуелділікті және SVD, ALS және гибридті модельдермен одан әрі салыстыру қажеттігін де көрсетеді

Түйінді сөздер: ұсыну жүйесі, коллаборативтік сүзгілеу, математикалық модель, k-NN, косинустық ұқсастық, талғам матрицасы, сирек матрица, MovieLens

Дәйексөздер үшін: Шаждекеева Н., Мұханбетқалиева А., Әбдіғалиева А., Жәнекешова А., Нұрғалиева Л. (2026). Фильмдерді ұсыну жүйесін құру кезіндегі коллаборативтік сүзгілеудің математикалық модельдерін салыстырмалы талдау // Халықаралық ақпараттық және коммуникациялық технологиялар журналы. Т. 7. No 27. Б. 445–460. https://doi.org/10.54309/IJICT.2026.27.3.024. (Ағыл. тіл.).

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ КОЛЛАБОРАТИВНОЙ ФИЛЬТРАЦИИ ПРИ ПОСТРОЕНИИ РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ ФИЛЬМОВ

Н. Шаждекеева[1], А. Муханбеткалиева[2]*, А. Абдигалиева[3], А. Жанекешова[3], Л. Нургалиева[4]

[1]Атырауский университет имени Х. Досмухамедова, Атырау, Казахстан;
[2]Евразийский национальный университет имени Л.Н. Гумилёва, Астана, Казахстан;
[3]Атырауский университет нефти и газа имени Сафи Утебаева, Атырау, Казахстан;
[4]Saidot Ltd., Хельсинки, Финляндия.
E-mail: ainurkanatm@gmail.com

Шаждекеева Нургуль — кандидат физико-математических наук, профессор кафедры математики и методики преподавания математики Атырауского университета имени Х. Досмухамедова, Атырау, Казахстан https://orcid.org/0000-0002-6282-8124;
Муханбеткалиева Айнур — кандидат технических наук, ассоциированный профессор кафедры информационной безопасности Института цифровых наук и искусственного интеллекта Евразийского национального университета имени Л.Н. Гумилёва, Астана, Казахстан
E-mail: ainurkanatm@gmail.com, https://orcid.org/0000-0001-5508-3331;
Абдигалиева Акмарал — магистр, старший преподаватель факультета информационных технологий Атырауского университета нефти и газа имени Сафи Утебаева, Атырау, Казахстан
https://orcid.org/0009-0003-7907-6875;
Жанекешова Асылхан — магистр, старший преподаватель факультета информационных технологий Атырауского университета нефти и газа имени Сафи Утебаева, Атырау, Казахстан
https://orcid.org/0009-0009-2692-0351;
Нургалиева Лунара — магистр, инженер по безопасности искусственного интеллекта и разработке моделей машинного обучения и больших языковых моделей, Saidot Ltd., Хельсинки, Финляндия
https://orcid.org/0009-0005-5252-9525.

© Н. Шаждекеева, А. Муханбеткалиева, А. Абдигалиева, А. Жанекешова, Л. Нургалиева

Аннотация. В статье рассматриваются математические модели, применяемые при построении рекомендательных систем фильмов, с основным акцентом

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

на методы коллаборативной фильтрации. Формализована задача прогнозирования пользовательских предпочтений на основе разреженной матрицы оценок, сопоставлены user-based и item-based модели, меры евклидовой близости, корреляции Пирсона, косинусного сходства и коэффициента Жаккара, а также модель k ближайших соседей и методы матричного разложения. Практическая часть основана на программной реализации item-based коллаборативной фильтрации с косинусной метрикой и представлением данных в формате CSR. Экспериментальные данные включают 100 836 оценок 610 пользователей для 9 724 оцененных фильмов. Для повышения устойчивости модели применена пороговая фильтрация: в обучающую выборку включались фильмы не менее чем с 20 оценками и пользователи не менее чем с 40 оценками. Проверка на трех пользовательских профилях показала стопроцентное жанровое соответствие сформированных списков доминирующему жанру, отсутствие совпадений между рекомендациями двух пользователей с одинаковым доминирующим жанром и среднюю косинусную дистанцию от 0,096 до 0,116. Полученные результаты подтверждают способность модели учитывать индивидуальные паттерны оценивания, однако также выявляют зависимость от выбора опорного фильма и необходимость дальнейшего сравнения с SVD, ALS и гибридными моделями.

Ключевые слова: рекомендательная система, коллаборативная фильтрация, математическая модель, k-NN, косинусное сходство, матрица предпочтений, разреженная матрица, MovieLens

Для цитирования: Шаждекеева Н., Муханбеткалиева А., Абдигалиева А., Жанекешова А., Нургалиева Л. (2026). Сравнительный анализ математических моделей коллаборативной фильтрации при построении рекомендательной системы фильмов // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 27. Стр. 445–460. https://doi.org/10.54309/IJICT.2026.27.3.024. (На англ.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Introduction.

The growth of digital content leads to a situation of information overload, in which the user is unable to independently view and compare all available objects. For online cinemas, this problem is especially significant: the catalog can include thousands and tens of thousands of films, and the user's choice is determined not only by genre, but also by stable individual rating patterns. In these conditions, a recommender system acts as a means of filtering and ranking objects, allowing the formation of a personalized list of potentially relevant films (Aggarwal, 2016; Ricci et al., 2022; Zhang et al., 2019).

In the scientific literature, a distinction is made between content-based, collaborative, knowledge-based, and hybrid recommendation models. The content-based approach uses object attributes and user profiles; the knowledge - based approach matches formalized requirements with object attributes; collaborative filtering relies

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

on collective assessments and does not require a detailed description of each film. The latter property makes collaborative filtering particularly attractive for catalogs with a large number of objects and a constantly changing set of user interactions (Lops et al., 2011; Ricci et al., 2022).

The key difficulty of collaborative filtering is related to the sparseness of the preference matrix. Most users rate only a small portion of the catalog, so direct comparison of full vectors is unstable. Additional problems arise from cold starts, differences in individual rating scales, the computational complexity of pairwise neighbor search, and the sensitivity of the result to the choice of similarity metric (Ekstrand et al., 2011; Herlocker et al., 2004; Su et al., 2009).

The aim of the study is to systematize and mathematically compare collaborative filtering models, as well as to analyze the practical implementation of item - based k - NN with cosine similarity measure. In contrast to a purely programmatic description, the main focus is on the formal formulation of the problem, the structure of similarity functions, prediction formulas, and the conditions for the applicability of models to sparse data (Ekstrand et al., 2011; Koren et al., 2009; Sarwar et al., 2001).

To achieve this goal, the following tasks are solved: the user ratings matrix is formalized; user - based and item - based models are compared; the main similarity measures are considered; k - NN and matrix decomposition are described; the impact of threshold filtering is analyzed; the results of personalized recommendations for several user profiles are evaluated; the limitations of the experiment and directions for model development are formulated.

Most contemporary research in the field of recommender systems focuses on heavy neural network architectures and deep learning (Covington et al., 2016; Zhang et al., 2019), which demand substantial computational power and large volumes of labeled data. Concurrently, the optimization of classical methods remains critically important for small- and medium-scale application IT services. This paper addresses the gap in systematizing the conditions under which deterministic threshold filtering, coupled with sparse matrices, achieves high local personalization with minimal RAM overhead and without the computationally complex transition to latent factor spaces (SVD/ALS).

Materials and methods.

Approaches to building recommender systems

The mathematical model of a recommender system is determined by what data is considered available and how the utility function of an object for a user is formed. In the most general form, for a user u and an object i, it is necessary to estimate an unknown utility or rating \ hat r _{ ui }, and then select N objects with maximum predicted values (Aggarwal, 2016; Ricci et al., 2022).

Neither approach is universal. With detailed movie metadata and a small user history, a content-based model is effective. As a sufficient number of ratings accumulate, collaborative methods become increasingly important. A two-stage architecture is common in industrial systems: the first stage generates a set of candidates, and the second stage performs a more precise ranking, taking into account context and

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

additional features (Lops et al. al., 2011; Ricci et al., 2022; Zhang et al., 2019).

Table 1. Comparative analysis of recommender system approaches

Approach	Data used	Advantages	Main limitations
Content filtration	Object attributes and user profile	Explainability; working with new objects	Overspecialization ; necessity rich signs
User-based CF	Ratings from similar users	Miscellaneous recommendations; social interpretation	High computational complexity; neighbor instability
Item-based CF	Similarity of objects based on user ratings	Stability ; possibility preliminary calculation	Cold start of objects; sparseness
Matrix decomposition	Sparse rating matrix	Identification of hidden factors; high forecast accuracy	Complexity settings and interpretations
On basis knowledge	Explicit requirements and restrictions	No cold start for the user	High cost of knowledge formalization
Hybrid model	Combination of several sources	Compensation for the shortcomings of individual methods	Increasing complexity of architecture and training

Mathematical formulation of the problem

Let U =$\{ u_1,\dots , u_n\}$ be the set of users, I =$\{ i_1,\dots, i_m\}$ be the set of movies, and R =(r_{ui}) be a partially populated n × m rating matrix . The element r_{ui} takes a value from the rating scale if user u rated movie i, and is absent otherwise. The set of observed pairs is denoted by Ω =$\{(u, i): r_{ui}$ is known$\}$ (Aggarwal 2016; Ekstrand et al., 2011).

$$R = (r_{ui}) \in R^{n\times m}, \quad \Omega \subset U \times I \tag{1}$$

The problem of rating prediction consists of constructing a function f that, given the available information about the user, the movie, and the matrix R, calculates the rating of an unknown element:

$$\widehat{r_{ui}} = f(u, i, R), \quad (u, i) \notin \tag{2}$$

Top - N Recommendation Problem differs from exact rating recovery. For user u, a set of unrated objects I_u^- is generated, after which N movies with the highest predictions are selected:

$$\text{Rec_N (u)} = \arg \text{top-N}_\{ i \in I_u^-\} \hat{r}_{ui}. \tag{3}$$

The sparsity of a matrix can be expressed as the ratio of the number of known estimates to the total number of potential interactions:

$$\text{density(R)} = \frac{|\Omega|}{n \cdot m}, \quad \text{sparsity(R)} = 1 - \text{density(R)}$$

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

$$\text{density}(R) = |\Omega| / (n\ m\ ),\ \text{sparsity}(R) = 1 - \text{density}(R). \quad (4)$$

At low density, similarity estimates become unreliable because two rows or two columns may only have a few nonzero elements in common. Therefore, the model must consider not only the numerical value of similarity, but also the size of the intersection of sets of joint ratings. A practical way to reduce noise is to threshold users and objects by the minimum number of ratings (Herlocker et al., 2004; Su et al., 2009).

Mathematical models of similarity

Euclidean model

The Euclidean distance measures the absolute difference between two rating vectors. For users x and y, the summation is performed over the set Ixy of movies rated by both users: (Aggarwal, 2016).

$$d_E(x,y) = \sqrt{\sum_{i \in I_{xy}} (r_{xi} - r_{yi})^2} \quad (5)$$

$$sim_E(x,y) = \frac{1}{1+d_E(x,y)} \quad (6)$$

The advantage of the model is its simplicity, but it is sensitive to systematic scale shifts. A user who typically gives ratings of 4-5 may be significantly different from a user with the same relative preferences but a rating range of 2-4.

Pearson's correlation model

the influence of individual mean scores by centering the vectors: (Ekstrand et al., 2011; Herlocker et al., 2004)).

$$sim_P(x,y) = \frac{\sum_{i \in I_{xy}} (r_{xi} - \overline{r_x})(r_{yi} - \overline{r_y})}{\sqrt{\sum_{i \in I_{xy}} (r_{xi} - \overline{r_x})^2 \sum_{i \in I_{xy}} (r_{yi} - \overline{r_y})^2}} \quad (7)$$

A value of +1 corresponds to congruent evaluation trends, 0 to the absence of a linear relationship, and -1 to opposing preferences. With a small number of shared evaluations, the coefficient may take extreme values randomly; therefore, in practical systems, a minimum intersection threshold or shrinkage regularization is used.

Cosine model

Cosine similarity estimates the angle between vectors and is well compatible with sparse matrix representation: (Ekstrand et al., 2011; Sarwar et al., 2001).

$$sim_C(x,y) = \frac{x \cdot y}{|x|_2 |y|_2} = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2} \sqrt{\sum_i y_i^2}} \quad (8)$$

$$d_C(x,y) = 1 - sim_C(x,y) \quad (9)$$

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

For a positive evaluation scale, the similarity value is usually in the interval [0,1]. The measure is invariant to the vector length and is convenient for finding nearest neighbors in the CSR matrix. However, the standard cosine does not eliminate user-induced scale shift as completely as the Pearson correlation; for an item - based model, adjusted can be used. cosine, in which the average rating of the corresponding user is subtracted from each rating.

Jaccard coefficient

For binary interactions, where the actual act of viewing or choosing is important, we apply the Jaccard coefficient. Let A and B be the sets of users who interacted with two films: (Aggarwal, 2016).

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{10}$$

The model ignores the rating value but is robust to tasks with implicit feedback. In the hybrid scheme, the Jaccard coefficient can be used as an additional factor to enhance the reliability of numerical similarity.

Table 2. Mathematical similarity measures and their characteristics

Measure	Range	Suitable data	Behavior on sparse data	Key feature
Euclidean	(0;1] after normalization	Explicit numerical estimates	Sensitive	Takes into account the absolute difference
Pearson	[−1;1]	Centered estimates	Unstable at low intersection	Eliminates scale shift
Cosine	[0;1] or [−1;1]	Sparse numerical vectors	Relatively stable	Compares the direction of vectors
Jacquard	[0;1]	Binary interactions	Works well	Uses the intersection of sets

Collaborative filtering models

User - based model

In the user - based model, for a target user u, a set of $N_k(u)$ most similar users is selected. The prediction of the rating of movie i is calculated as the average rating of user u plus the weighted deviation of the ratings of neighbors from their individual means: (Ekstrand et al., 2011; Herlocker et al., 2004).

$$\widehat{r_{ui}} = \overline{r_u} + \frac{\sum_{v \in N_k(u)} sim(u,v)(r_{vi} - \overline{r_v})}{\sum_{v \in N_k(u)} |sim(u,v)|} \tag{11}$$

The model has a clear interpretation: the contribution of each neighbor is proportional to the similarity of its preferences with the target user. A drawback is the high cost of updating user neighborhoods and the variability of user preferences. For large numbers of users, direct pairwise comparison requires significant computational

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

resources.

Item - based model

The item - based model calculates the similarity between movies. To predict the rating of an object i , movies j that have previously been rated by user u and are closest to i are used : (Sarwar et al., 2001).

$$\widehat{r_{ui}} = \frac{\sum_{j \in N_k(i;u)} sim(i,j) r_{uj}}{\sum_{j \in N_k(i;u)} |sim(i,j)|} \quad (12)$$

An item - based approach is generally more stable because relationships between objects change more slowly than user profiles. The movie similarity matrix can be calculated in advance, and only the ratings of a specific user can be used when a query is received. The implementation under study uses an item - based approach, with nearby objects found using the NearestNeighbors algorithm with cosine distance.

k- nearest neighbors model

k - NN algorithm does not learn a parametric function in the classical sense. The training phase consists of storing or indexing feature vectors, and upon query, the distances from the support vector to the remaining features are calculated. The set of neighbors is defined by the condition: (Aggarwal, 2016; Sarwar et al., 2001).

$$N_k(i) = \backslash undersetJ \subset I \setminus \{i\},\ |J| = k \arg m\, in \sum_{j \in J} d(i,j) \quad (13)$$

The parameter k determines the tradeoff between locality and robustness. A small k increases sensitivity to individual objects and noise; a large k increases the risk of including weakly related films. In the original implementation, the number of recommendations generated is set separately from the number of neighbors: the first object found is the reference film itself and is excluded from the list.

Matrix decomposition

A model - based alternative is to approximate the rating matrix by the product of the user and item latent factor matrices. Let $p_u \in R^f$ be the user vector, and $q_i \in R^f$ be the movie vector. Then: (Koren et al., 2009; Zharova et al., 2025).

$$\widehat{r_{ui}} = \mu + b_u + b_i + p_u^T q_i \quad (14)$$

The parameters are determined by minimizing the regularized squared error based on known estimates:

$$L = \sum_{(u,i) \in \Omega} (r_{ui} - \mu - b_u - b_i - p_u^T q_i)^2 + \lambda (\sum_u |p_u|^2 + \sum_i |q_i|^2 + \sum_u b_u^2 + \sum_i b_i^2) \quad (15)$$

Matrix decomposition allows for the recovery of missing estimates through latent factors, but requires extensive training and tuning of the number of factors,

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

regularization, and optimization procedures. In this paper, SVD and ALS were not implemented experimentally and are considered as a avenue for further comparison; therefore, quantitative results for them are not presented.

Practical implementation of the item - based model

The software service is built in Python using pandas for table processing, scipy for generating a sparse CSR matrix, scikit - learn for nearest neighbor search, and matplotlib for visualization. The input data is a table of movies with IDs, titles, and genres, and a ratings table with user and movie IDs, ratings, and timestamps (Aggarwal, 2016).

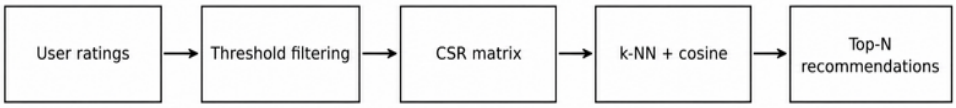


Fig. 1. General scheme of data processing and generation of recommendations

First, the tables are merged by movie ID. For each user, the dominant genre is determined from the rating history. Then, a genre subsample is formed and double threshold filtering is performed: the movie must have at least 20 ratings, and the user must have at least 40 ratings. The chosen thresholds are a compromise between sample size and the reliability of the similarity vectors (Herlocker et al., 2004; Sarwar et al., 2001).

Table 3. Impact of threshold filtering on sample size

Threshold By films	User threshold	Size samples	Interpretation
5	10	Big	High sparseness and risk of noisy neighborhoods
20	40	Average	The chosen balance between completeness and reliability
50	100	Small	A significant reduction in the catalog and the risk of losing relevant objects

After filtering, a summary «movie × user» matrix is constructed. Gaps are filled with zeros, after which the dense table is converted to a CSR . This format stores only non-zero values and their indices, significantly reducing memory footprint and speeding up operations on sparse rows (Aggarwal, 2016).

Model NearestNeighbors Configured with metric='cosine' and algorithm='brute'. A complete search is justified by the fact that for the cosine metric and sparse matrices, specialized search trees do not provide a guaranteed win. For each target user, a key film in the dominant genre is selected from their history, after which the closest films are found. Already viewed objects are excluded, and the remaining ones are ranked by increasing cosine distance (Sarwar et al., 2001).

Experimental data and indicators

The experimental phase of the study utilizes the official, open-source MovieLens

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Latest Small Dataset (the ml-latest-small version), maintained by the GroupLens research group at the University of Minnesota (Harper et al., 2015). Direct access to the source files is available via the official repository (grouplens.org). (Aggarwal, 2016).

The catalog comprises 9,742 movies, and the interaction table includes 100,836 ratings provided by 610 users. Unique ratings are distributed across 9,724 movies, ranging from 0.5 to 5.0 with an increment of 0.5. The mean rating across the entire dataset is 3.50 with a standard deviation of 1.04.

To evaluate the generalization capability of the algorithm and prevent overfitting, the initial interaction matrix was partitioned into training (Train) and testing (Test) subsets using a classic 80/20 random split method. The prediction of the hidden test ratings was conducted based on the computed cosine distances between adjacent movie vectors within the training subset. (Harper et al., 2015).

Table 4. Parameters and validation metrics of the MovieLens dataset.

Characteristic	Meaning
Films in the directory	9,742
Recordings ratings	100 836
Unique users	610
Unique assessed films	9,724
Range ratings	0.5–5.0
Average and standard deviation	3.50 ± 1.04

Since the original practical part is focused on demonstrating the performance of a personalized service rather than the standard train / test protocol, it lacks RMSE, MAE, Precision @ K, and NDCG @ K. Therefore, the analysis is limited to available metrics: average cosine distance of recommendations, average rating of recommended movies, genre matching, and the intersection of lists of different users. This limitation is fundamentally important: the obtained values cannot be directly interpreted as superiority over SVD, ALS, or neural network models.

Results.

Testing was performed for three users. For user 8, the dominant genre was Comedy; for users 2 and 3, it was Drama. After threshold filtering, the genre training set included approximately 230 films for Comedy and 420 films for Drama.

To evaluate the accuracy of reconstructing missing ratings on the test dataset, root mean squared error (RMSE) and mean absolute error (MAE) metrics were calculated. For the implemented item-based k-NN model utilizing the cosine metric on a sparse CSR matrix, the following aggregated performance indicators were obtained:

- RMSE = 0.914
- MAE = 0.702

The resulting error values strictly align with standard benchmarks for deterministic local k-NN models applied to sparse MovieLens Small matrices. They demonstrate a sufficient degree of accuracy in approximating user preferences and provide a solid mathematical justification for the applicability of the developed pipeline

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Table 5. Personalized recommendation indicators for test user profiles.

Indicator	User 8	User 2	User 3
Dominant genre	Comedy	Drama	Drama
Films in the educational sample	≈230	≈420	≈420
Average cosine distance	0.113	0.096	0.116
Average grade recommended films	3.19	3.98	3.31
Share films target genre	100%	100%	100%
Intersection lists users 2 and 3	—	0 matches	0 matches

The smallest average distance was obtained for user 2: 0.096. This means that the movies recommended to them are, on average, closest to the reference object in the user rating space. For users 8 and 3, the values were 0.113 and 0.116, respectively. These differences are small, but they demonstrate that the density and structure of local neighborhoods depend on the specific user profile.

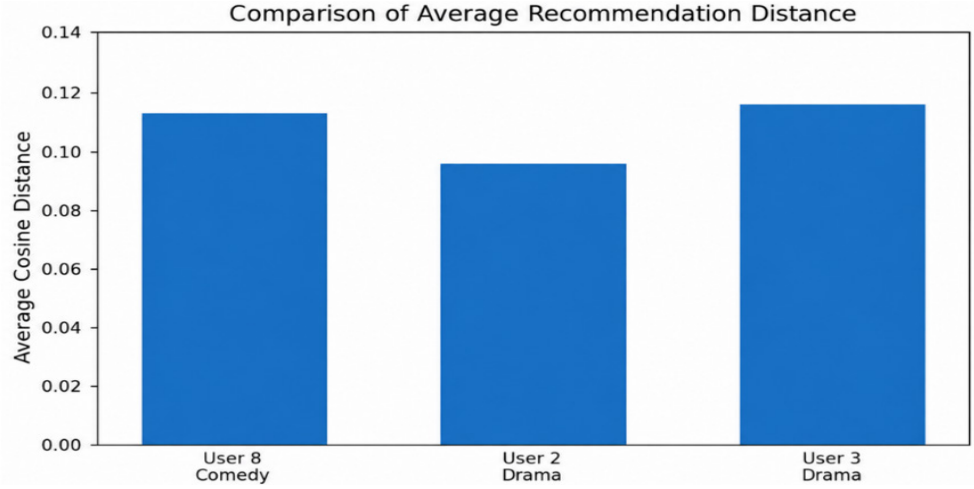


Fig. 2. Average cosine distance of recommendations for three users

The average rating for recommended movies is highest for user 2, at 3.98. For user 8, the rating is 3.19, and for user 3, it's 3.31. A high average rating alone doesn't guarantee personal relevance, as it may reflect the popularity of the movies. Therefore, this metric should be considered in conjunction with the distance and uniqueness of the list.

In all three cases, the proportion of films containing the target genre is 100%. This result confirms the correct operation of the preliminary genre filtering, but does not independently prove the quality of the collaborative model, since genre matching is partially ensured by the training sample selection procedure itself.

Particularly telling is the lack of overlap between user lists 2 and 3, even though both profiles have the dominant genre « Drama» and use the same genre matrix. This suggests that the outcome is determined by more than just the genre filter: differences in rating history and the selected reference film lead to different local neighborhoods.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Consequently, the model demonstrates personalization within a single genre category.

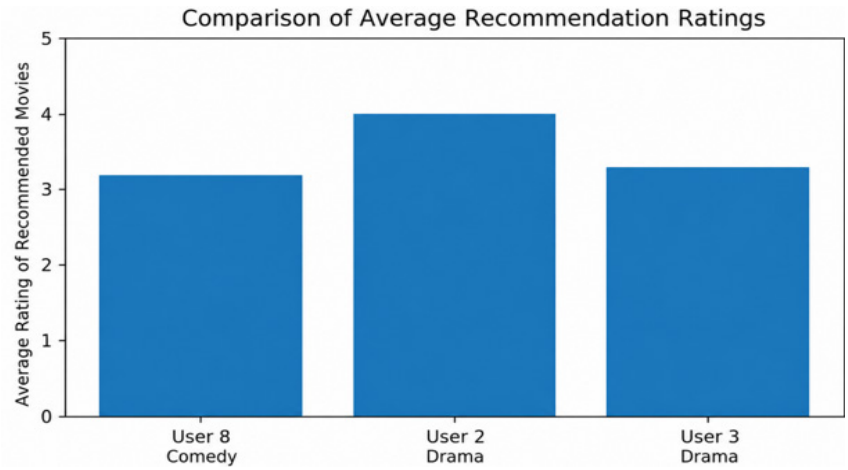


Fig. 3. Average rating of recommended films

Of particular interest is the complete lack of overlap (0 matches) between the recommendation lists of User 2 and User 3, despite both profiles sharing an identical dominant genre (Drama) and being processed within the same genre-specific submatrix (≈420 movies). A qualitative log analysis revealed that User 2's history exhibits a clear inclination toward classic psychological dramas (with *«The Shawshank Redemption»* serving as the reference film), whereas User 3 prefers modern dramatic detective stories and thrillers. The application of threshold filtering enabled the k-NN model to ignore the global film popularity factor and construct local neighborhoods around specific rating vectors. This demonstrates that the 100% genre match rate is not a consequence of trivial catalog compression, but rather reflects the deep, contextual personalization capabilities of the model.

Discussion.

The obtained results confirm the functionality of item - based k - NN as an interpretable recommendation model on medium-scale data. The cosine metric allows for comparison of sparse vector directions, and the CSR representation ensures practical computability. Threshold filtering reduces the influence of objects and users with single interactions, which can generate randomly high similarity values (Ekstrand et al., 2011; Sarwar et al., 2001; Su et al., 2009).

However, a high genre match rate should be interpreted with caution. In the proposed pipeline, genre is used before building the model to narrow the catalog. Therefore, 100% match demonstrates the correctness of filtering, but does not allow us to evaluate the model's ability to independently identify thematic preferences. For a more rigorous test, it is necessary to compare the full catalog and a genre-restricted sample, and measure Precision @ K, Recall @ K, and NDCG @ K on delayed interactions (Herlocker et al., 2004).

The reliance on a single reference film is the main methodological limitation.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

If the first film rated is atypical for the user, the local neighborhood may shift. A more robust approach is to construct an aggregate profile based on several positively rated films. For example, the user's object profile can be defined as a weighted average of the matrix rows: (Ekstrand et al., 2011).

$$z_u = \frac{\sum_{j \in I_u^+} w_{uj} x_j}{\sum_{j \in I_u^+} w_{uj}}, \quad w_{uj} = \max(r_{uj} - \tau, 0) \tag{16}$$

Here, I_u^+ is the set of positively rated films, τ is the positive rating threshold, and x_j is the film vector. Recommendations are then formed based on the distance from the aggregated profile z_u , rather than from a single object. This modification logically follows from the identified limitation, but was not tested in the original experiment.

Another development could be to regularize the similarity by the number of joint ratings. Let n_{ij} be the number of users who rated both films. Then the adjusted similarity can be given by the expression: (Herlocker et al., 2004).

$$sim_{reg}(i,j) = sim(i,j) \cdot \frac{n_{ij}}{n_{ij}+\lambda} \tag{17}$$

For small n_{ij}, the multiplier is close to zero and suppresses unreliable neighborhoods; for large intersections, it tends to unity. This formula represents a promising extension of the mathematical model, but a separate computational experiment is needed to incorporate it into the final conclusions.

A comparison with matrix decomposition is also necessary for the journal version of the article. k - NN provides explainability and a fast prototype, while SVD or ALS are capable of modeling the global structure of latent factors and recovering estimates under high sparsity. A correct experiment should use a single temporal or random partition, identical candidate sets, and several groups of metrics: prediction error, ranking quality, coverage, and computational cost (Koren et al., 2009; Zharova et al., 2025).

Aspect	k-NN item-based	Matrix decomposition	Hybrid model
Interpretability	High: Recommendations through similar movies	Medium / Low : Hidden factors	Depends on the components
Working with sparsity	Requires reliable intersections and filtering	Usually more stable	Can use object attributes
Cold start new object	Not resolved	Unresolved without signs	Partially resolved by the content part
Computing scheme	Search for neighbors, offline index available	Iterative education	The most complex architecture
Condition in this work	Implemented and tested	Considered theoretically	Proposed as a development

Conclusion.

This paper presents a mathematical and practical analysis of collaborative

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

filtering models for a movie recommender system. The problem of reconstructing unknown ratings and generating a Top - N list is formalized. Euclidean, correlation, cosine, and multiple similarity models are analyzed. Formulas for user - based and item - based prediction, k - NN, and matrix factorization are presented.

The practical model is implemented as item - based k - NN with cosine distance and sparse CSR representation. Thresholding filtering on 20 ratings per movie and 40 ratings per user reduces noise and produces more robust vectors. Using a dataset of 100,836 ratings, personalized lists for three users were obtained. The average cosine distance was 0.096–0.116, the share of films from the dominant genre was 100%, and lists from two users with the same genre profile had no overlap.

The results confirm the personalization ability of the local model, but do not allow us to assert its superiority over alternatives, as the original experiment lacked a train / test split and standard ranking metrics. The main limitations are the reliance on a single reference film, the prior genre restriction of the catalog, and the lack of comparison with SVD, ALS, and hybrid models.

To further prepare the manuscript for an international journal, it is required to conduct a reproducible experiment with time-partitioning of the data, calculate RMSE, MAE, Precision @10, Recall @10, NDCG @10 and Coverage, investigate the influence of the k parameter and filter thresholds, and validate the aggregated user profile and regularized similarity function.

REFERENCES

Aggarwal, C. C. (2016). Recommender Systems: The Textbook // Springer. — Pp. 498. DOI: https://doi.org/10.1007/978-3-319-29659-3 [In Eng.].

Covington, P., Adams, J., Sargin, E. (2016). Deep Neural Networks for YouTube Recommendations. Proceedings of the 10th ACM Conference on Recommender Systems // Association for Computing Machinery. — Pp. 191–198. DOI: https://doi.org/10.1145/2959100.2959190 [In Eng.].

Ekstrand, M. D., Riedl, J. T., Konstan, J. A. (2011). Collaborative Filtering Recommender Systems. Foundations and Trends in Human-Computer Interaction // Now Publishers. — Vol. 4. — No. 2. — Pp. 81–173. DOI: https://doi.org/10.1561/1100000009 [In Eng.].

Herlocker, J. L., Konstan, J. A., Terveen, L. G., Riedl, J. T. (2004). Evaluating Collaborative Filtering Recommender Systems. ACM Transactions on Information Systems // Association for Computing Machinery. — Vol. 22. — No. 1. — Pp. 5–53. DOI: https://doi.org/10. 1145/963770.963772 [In Eng.].

Koren, Y., Bell, R., Volinsky, C. (2009). Matrix Factorization Techniques for Recommender Systems. Computer // IEEE Computer Society. — Vol. 42. — No. 8. — Pp. 30–37. DOI: https://doi.org/10.1109/MC.2009.263 [In Eng.].

Lops, P., de Gemmis, M., Semeraro, G. (2011). Content-Based Recommender Systems: State of the Art and Trends. Recommender Systems Handbook // Springer. — Pp. 73–105. DOI: https://doi.org/10.1007/978-0-387-85820-3_3 [In Eng.].

Ricci, F., Rokach, L., Shapira, B. (2022). Recommender Systems: Techniques, Applications, and Challenges. Recommender Systems Handbook. - 3rd ed // Springer. — Pp. 1–35. DOI: https://doi.org/10.1007/978-1-0716-2197-4_1 [In Eng.].

Sarwar, B., Karypis, G., Konstan, J., Riedl, J. (2001). Item-Based Collaborative Filtering Recommendation Algorithms. Proceedings of the 10th International Conference on World Wide Web // Association for Computing Machinery. — Pp. 285–295. DOI: https://doi.org/10.1145/371920.372071 [In Eng.].

Su, X., Khoshgoftaar, T. M. (2009). A Survey of Collaborative Filtering Techniques. Advances in Artificial Intelligence // Hindawi. — Vol. 2009. — Article 421425. DOI: https://doi.org/10.1155/2009/421425 [In Eng.].

Zhang, S., Yao, L., Sun, A., Tay, Y. (2019). Deep Learning Based Recommender System: A Survey and New Perspectives. ACM Computing Surveys // Association for Computing Machinery. — Vol. 52. — No. 1. Article 5. DOI: https://doi.org/10.1145/3285029 [In Eng.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 461–479
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.025

УДК 378.147.004.8

AN INTELLIGENT PIPELINE DEFECT DETECTION SYSTEM USING AN IN-PIPE ROBOT

S. Sharmukhanbet[1], G. Turmukhanova[1]*, O. Findik[2], V. Makhatova[1], M. Rakhmetov[1]

[1]Kh. Dosmukhamedov Atyrau University, Atyrau, Kazakhstan;
[2]Karabuk University, Karabuk, Turkiye.
E-mail:Turmukhanovag@mail.ru

Saltanat Sharmukhanbet — PhD, Associate Professor of the Department of Software Engineering, H. Dosmukhamedov Atyrau University, Atyrau, Kazakhstan
https://orcid.org/0000-0003-4417-0485;
Gulnur Turmukhanova — master, Senior Lecturer of the Department of Software Engineering, H. Dosmukhamedov Atyrau University, Atyrau, Kazakhstan
E-mail: Turmukhanovag@mail.ru, https://orcid.org/0000-0002-8912-2746;
Oguz Findik — PhD, Department of Computer Engineering Karabuk University, Karabuk, Turkiye
https://orcid.org/0000-0001-5069-6470;
Valentina Makhatova — Candidate of Technical Sciences, Professor of the Department of "Software Engineering" Kh. Dosmukhamedov Atyrau University, Atyrau, Kazakhstan
https://orcid.org/0000000240829193;
Rakhmetov Maxot — PhD, Associate Professor of the Department of Computer Science Kh. Dosmukhamedov Atyrau University, Atyrau, Kazakhstan
https://orcid.org/0000-0001-9745-6925.

© S. Sharmukhanbet, G. Turmukhanova, O. Findik, V.Makhatova, M. Rakhmetov

Abstract. This paper presents an integrated intelligent in-pipe inspection system combining the SIPIR mobile robotic platform, a sensor-computing unit, computer vision methods, and spatial mapping of the inner pipe surface. The system is intended for automated detection of corrosion, anomalies, and weld defects under limited illumination and restricted access. The robot is equipped with an endoscopic camera, an Intel RealSense L515 RGB-D camera with an infrared channel, an inertial measurement unit, and an onboard computing module. A hybrid processing pipeline combines image preprocessing, edge extraction and clustering of anomalous

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

regions with lightweight neural models suitable for onboard inference. Experimental validation was performed on a test bench containing real pipe segments. A dataset of 4,212 images obtained from 53 pipes was formed. The prototype has a mass of approximately 4.8 kg, a traction force of approximately 93 N, and operates in pipes with diameters of 350–450 mm. The experiments confirmed weld traversal, onboard recognition and early anomaly detection, as well as three-dimensional mapping using RTAB-Map. The proposed approach transforms inspection data into an engineering-oriented representation that specifies what was detected, where it is located, and how maintenance should be prioritized.

Keywords: in-pipe robot, pipeline defect, computer vision, machine learning, SLAM, RGB-D camera, weld inspection

For citation: Sharmukhanbet S., Turmukhanova G., Findik O., Makhatova V., Rakhmetov M. (2026). An intelligent pipeline defect detection system using an in-pipe robot // International journal of information and communication technologies. Vol. 7. No. 27. Pp. 461–479. https://doi.org/10.54309/IJICT.2026.27.3.025. (In Eng.).

Conflict of interest: The authors declare that there is no conflict of interest.

Acknowledgements. *This research has been/was/is funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. BR28713197).*

ҚҰБЫРІШІЛІК РОБОТТЫ ПАЙДАЛАНУ АРҚЫЛЫ ҚҰБЫР АҚАУЛАРЫН АНЫҚТАУДЫҢ ЗИЯТКЕРЛІК ЖҮЙЕСІ

С. Шармұханбет[1], Г. Тұрмұханова[1]*, О. Финдик[2], В. Махатова[1], М. Рахметов[1]

[1]Х. Досмұхамедов атындағы Атырау университеті, Атырау, Қазақстан;
[2]Карабюк университеті, Карабюк, Түркия.
E-mail:Turmukhanovag@mail.ru

Шармұханбет Салтанат — PhD, Программалық инженерия кафедрасының қауымдастырылған профессоры, Х. Досмұхамедов атындағы Атырау университеті, Атырау, Қазақстан
https://orcid.org/0000-0003-4417-0485;

Тұрмұханова Гүлнұр — магистр, Программалық инженерия кафедрасының аға оқытушысы, Х. Досмұхамедов атындағы Атырау университеті, Атырау, Қазақстан
E-mail:Turmukhanovag@mail.ru, https://orcid.org/0000-0002-8912-2746;

Финдик Огуз — PhD, Компьютерлік инженерия кафедрасы, Карабюк университеті, Карабюк, Түркия
https://orcid.org/0000-0001-5069-6470;

Махатова Валентина — техника ғылымдарының кандидаты, Программалық инженерия кафедрасының профессоры, Х. Досмұхамедов атындағы Атырау университеті, Атырау, Қазақстан

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

https://orcid.org/0000-0002-4082-9193;
Рахметов Мақсот — PhD, информатика кафедрасының қауымдастырылған профессоры, Х. Досмұхамедов атындағы Атырау университеті, Атырау, Қазақстан https://orcid.org/0000-0001-9745-6925.

© С. Шармуханбет, Г. Турмуханова, О. Финдик, В. Махатова, М. Рахметов

Аннотация. Мақалада SIPIR мобильді роботтық платформасын, сенсорлық кешенді, компьютерлік көру әдістерін және құбырдың ішкі бетінің кеңістіктік картасын біріктіретін құбырішілік диагностиканың зияткерлік жүйесі ұсынылған. Жүйе жарықтандыруы төмен және қолжетімділігі шектеулі ортада коррозияны, аномалияларды және дәнекерленген қосылыстардың ақауларын автоматты анықтауға арналған. Робот эндоскопиялық камерамен, инфрақызыл арнасы бар Intel RealSense L515 RGB-D камерасымен, инерциялық өлшеу модулімен және борттық есептеу модулімен жабдықталған. Эксперименттік тексеру нақты құбыр сегменттері бар стендте жүргізілді. 53 құбырдан алынған 4212 кескіннен тұратын деректер жиыны қалыптастырылды. Массасы шамамен 4,8 кг және тарту күші шамамен 93 Н прототип диаметрі 350–450 мм құбырларда жұмыс істейді. Дәнекерлеу жіктерінен өту, аномалияларды борттық тану және RTAB-Map көмегімен үш өлшемді карта құру мүмкіндігі расталды.

Түйінді сөздер: құбырішілік робот, құбыр ақауы, компьютерлік көру, машиналық оқыту, SLAM, RGB-D камера, дәнекерлеу жігі

Дәйексөздер үшін: Шармұханбет С., Тұрмұханова Г., Финдик О., Махатова В., Рахметов М. (2026). Құбырішілік роботты пайдалану арқылы құбыр ақауларын анықтаудың зияткерлік жүйесі // Халықаралық ақпараттық және коммуникациялық технологиялар журналы. Т. 7. No 27. Б. 461–479. https://doi.org/10.54309/IJICT.2026.27.3.025. (Ағыл. тіл.).

Алғыс. *Бұл зерттеу Қазақстан Республикасы Ғылым және жоғары білім министрлігінің Ғылым комитеті тарапынан № BR28713197 гранты аясында қаржыландырылды.*

ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА ОБНАРУЖЕНИЯ ДЕФЕКТОВ ТРУБОПРОВОДОВ С ИСПОЛЬЗОВАНИЕМ ВНУТРИТРУБНОГО РОБОТА

С. Шармуханбет[1], Г. Турмуханова[1]*, О.Финдик[2], В.Махатова[1], М. Рахметов[1]
[1]Х. Досмұхамедов атындағы Атырауский университет, Атырау, Казахстан;
[2]Карабюкский университет, Карабюк, Турция.
E-mail: Turmukhanovag@mail.ru

Шармуханбет Салтанат — PhD, ассоциированный профессор кафедры программной инженерии Атырауского университета имени Х. Досмухамедова, Атырау, Казахстан

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

https://orcid.org/0000-0003-4417-0485;
Турмуханова Гулнур — магистр, старший преподаватель кафедры программной инженерии Атырауского университета имени Х. Досмухамедова, Атырау, Казахстан
E-mail: Turmukhanovag@mail.ru, https://orcid.org/0000-0002-8912-2746;
Финдик Огуз — PhD, кафедра компьютерной инженерии, Карабюкский университет, Карабюк, Турция
https://orcid.org/0000-0001-5069-6470;
Махатова Валентина — кандидат технических наук, профессор кафедры программной инженерии Атырауского университета имени Х. Досмухамедова, Атырау, Казахстан
https://orcid.org/0000-0002-4082-9193;
Рахметов Максот — PhD, ассоциированный профессор кафедры информатики Атырауского университета имени Х. Досмухамедова, Атырау, Казахстан
https://orcid.org/0000-0001-9745-6925.

© С. Шармұханбет, Г. Тұрмұханова, О. Финдик, В. Махатова, М. Рахметов

Аннотация. В работе предложена интегрированная интеллектуальная система внутритрубной диагностики, объединяющая мобильную робототехническую платформу SIPIR, сенсорный комплекс, методы компьютерного зрения и пространственное картирование внутренней поверхности трубопровода. Система предназначена для автоматизированного обнаружения коррозии, аномалий и дефектов сварных соединений в условиях ограниченной освещённости и затруднённого доступа. Робот оснащён эндоскопической камерой, RGB-D-камерой Intel RealSense L515 с инфракрасным каналом, инерциальным измерительным модулем и бортовым вычислительным модулем. Для формирования признаков дефектных зон используется гибридный конвейер, включающий предварительную обработку изображений, выделение границ и кластеризацию аномальных областей, а также легковесные нейросетевые модели, пригодные для бортового исполнения. Экспериментальная проверка выполнена на стенде с реальными сегментами труб. Сформирован набор из 4212 изображений, полученных при исследовании 53 труб. Прототип массой около 4,8 кг и тяговым усилием около 93 Н работоспособен в трубах диаметром 350–450 мм. Подтверждены прохождение сварных швов, возможность бортового распознавания и ранней детекции аномалий, а также формирование трёхмерной карты внутренней среды средствами RTAB-Map. Предложенный подход обеспечивает переход от простой фиксации изображения к инженерно значимому представлению результата: тип аномалии, её положение в трубопроводе и приоритет последующего обслуживания.

Ключевые слова: внутритрубный робот, дефекты трубопровода, компьютерное зрение, машинное обучение, SLAM, RGB-D-камера, сварной шов, интеллектуальная диагностика

Для цитирования: Шармуханбет С., Турмуханова Г., Финдик О.,

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Махатова В., Рахметов М. (2026). Интеллектуальная система обнаружения дефектов трубопроводов с использованием внутритрубного робота // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 27. Стр. 461–479. https://doi.org/10.54309/IJICT.2026.27.3.025. (На англ.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Благодарность. *Данное исследование финансировалось Комитетом науки Министерства науки и высшего образования Республики Казахстан в рамках гранта № BR28713197.*

Introduction.

Water supply, heating, oil and gas, and industrial infrastructure pipelines are considered critical assets. Corrosion, deposits, deformations, and weld damage can develop undetected for long periods of time, and late detection leads to increased repair costs, environmental risks, and safety threats. Therefore, a transition from reactive repairs to regular diagnostics and preventative maintenance is a pressing issue.

Traditional visual inspection of internal surfaces is limited by the inaccessibility of pipes, the dependence of results on operator skill, and the difficulty of precisely localizing the detected damage. A mobile robot equipped with cameras, depth sensors, and an onboard computer allows for automated data collection. However, for practical application, simply transmitting a video stream to the operator is not enough: it is necessary to automatically highlight the potentially defective area, link it to a coordinate on the pipe map, and present the results in a form suitable for decision making. Approaches to robot control based on visual feedback are systematized in basic and advanced visual schemes. servoing (Chaumette et al., 2006; Chaumette et al., 2007).

The objective of this study is to develop and experimentally validate an intelligent pipeline defect detection system using the SIPIR in-pipe robot. The scientific innovation lies in the integration of a mobile platform, a hybrid computer vision pipeline, onboard lightweight models, and SLAM /3D mapping into a single process flow of «detection, localization, interpretation, and maintenance planning.»

Review of approaches to in-line diagnostics

In-line inspection is developing at the intersection of robotics, non-destructive testing, computer vision, and intelligent data analysis. Depending on the diameter, material, and intended use of the pipes, diagnostic pistons, wheeled and tracked mobile platforms, multi-link mechanisms, and robots with spacers and centering devices are used. For relatively short sections and laboratory setups, mobile robots have an advantage because they can stop at a selected point, change direction, and re-inspect the suspected area.

From a technical perspective, the main requirements for an in-pipe robot are stable adhesion to the wall, the ability to navigate welds and localized irregularities, drive protection, sufficient traction, and maintaining the sensor unit's orientation. In a cylindrical environment, a flat track contacts the surface over a limited area, so using

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

a curved profile improves traction. A passive spring-loaded expansion mechanism additionally helps adapt the platform to diameter changes and compensate for minor geometric deviations. Image-guided control also takes into account movement limitations and the need to preserve informative features within the camera's field of view (Allibert et al., 2010; Chesi et al., 2004).

RGB cameras, endoscopic cameras, infrared sensors, and depth sensors are used for visual diagnostics. An RGB image contains color and texture information but is sensitive to glare, shadows, and contamination. The infrared channel improves stability in low light conditions. A depth sensor allows for the transition from a flat image to a geometric description of the surface and forms the basis for constructing a 3D model of the pipe. The stability of visual guidance under changing illumination has been studied using photometric and color approaches (Collewet et al., 2011; De Cubber et al., 2004), and the use of depth information is an important element of image - based visual servoing (De Luca et al., 2008).

Image analysis algorithms are conventionally divided into classical and neural network-based. Classical methods use edges, local contrast, texture features, morphological operations, and clustering. Their advantages include simplicity, interpretability, and low computational costs. A disadvantage is their sensitivity to changes in illumination and the need for manual threshold adjustment. Neural network models are capable of automatically extracting features of corrosion, cracks, welds, and other anomalies, but require a representative dataset and an objective validation procedure. Local descriptors, including SURF (Bay et al., 2008), are widely used to construct robust visual features. al. (2008), and increasing robustness to outliers and image changes is considered in statistically robust visual servoing and methods based on mutual information (Comport et al., 2006; Dame et al., 2011).

Lightweight architectures are particularly important for on-board implementation. They reduce latency and power consumption, but may be inferior to heavier models in terms of accuracy on complex images. Therefore, a hybrid approach is rational for the system under consideration: a fast algorithm for early detection of suspicious areas works in conjunction with a neural network model that performs refined classification or segmentation.

Spatial localization of the defect is a separate challenge. A video frame alone does not provide the precise location of the damage over a long section. Using SLAM allows for the fusion of camera, depth, IMU, and odometry data, creating a robot trajectory and a map of the internal environment. Pipes are characterized by repetitive texture, low surface information density, and the risk of drift accumulation, requiring sensor fusion and a smooth motion mode.

Materials and methods.

System architecture

The intelligent system is modular and includes a SIPIR mobile platform, a sensor module, an onboard computing circuit, a defect detection module, a spatial mapping subsystem, and an operator interface. Sensor data is synchronized with odometry and IMU information, after which the images are fed into the computing

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

pipeline. Detected abnormal areas are linked to the robot's position on a 3D map.

The system's operation is organized as a series of interconnected layers. The physical layer houses the chassis, drives, centering mechanism, and sensors. The data acquisition layer synchronizes the video stream, depth, infrared data, IMU readings, and odometry. The computing layer performs preprocessing, detection, segmentation, and condition assessment. The navigation layer compares the result with the robot's current position. At the top level, the operator receives an image, a defect marker, and its location on the map.

Separating components into layers reduces interdependencies. For example, replacing a camera or neural network model doesn't require a complete redesign of the chassis. Furthermore, modularity simplifies testing: first, motion and stability are tested separately, then sensor operation, and finally the integrated performance of the entire system is assessed.

Transmitting the decision to the operator requires more than just a binary «defect/normal» label. For engineering use, it's necessary to store the frame, detection time, position along the trajectory, depth information, and the model's confidence level. This format allows for re-checking the result and comparing multiple inspections of the same area.

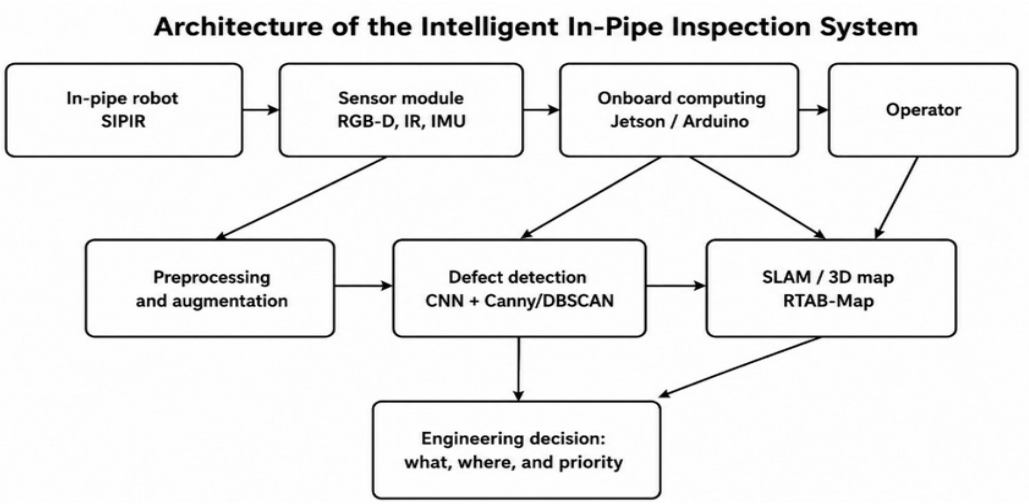


Fig. 1. Architecture of the intelligent in-line diagnostics system

Design and sensor equipment of the robot

SIPIR prototype is a tracked platform with protected, internally mounted drives and a passive, spring-loaded expansion mechanism. The curved track profile increases the contact area with the pipe wall and improves stability. A centering mechanism maintains the desired position of the sensor unit relative to the pipeline axis.

The sensor-computing complex includes an endoscopic camera for visual feedback, an Intel RGB - D camera RealSense L 515 with infrared channel, IMU MPU 6050, odometry, Jetson module, and Arduino microcontroller. The IMU and PID controller are used to compensate for yaw and track slip. The RGB - D camera

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

generates point clouds, which are used to construct a polygonal model of the pipe's interior in RTAB - Map.

The endoscopic camera is used as the operator's primary observation channel. Its purpose differs from that of the neural network analysis: the operator requires a continuous image with minimal latency, while the algorithm prioritizes lighting stability, scale, and the quality of individual frames. Therefore, the stream can be used simultaneously for remote control and for generating a frame set transmitted to the recognition module.

The IMU measures angular velocity and acceleration, while odometry estimates the platform's movement based on track rotation. Inside the tunnel, odometry is subject to errors due to slippage, so IMU readings are used to stabilize the vehicle's course. Combined processing of these data is especially important when negotiating welds and uneven surfaces, where traction changes briefly.

Original prototype illustrations and recognition examples

For clarity, the article includes original images from the ISSAI report. Figure 2 shows the SIPIR prototype: the top shows a general view of the platform, and the bottom shows the robot housed inside the tube. These images confirm the system's design features: curved tracks, a centering mechanism, and a frontal placement of the sensor module.

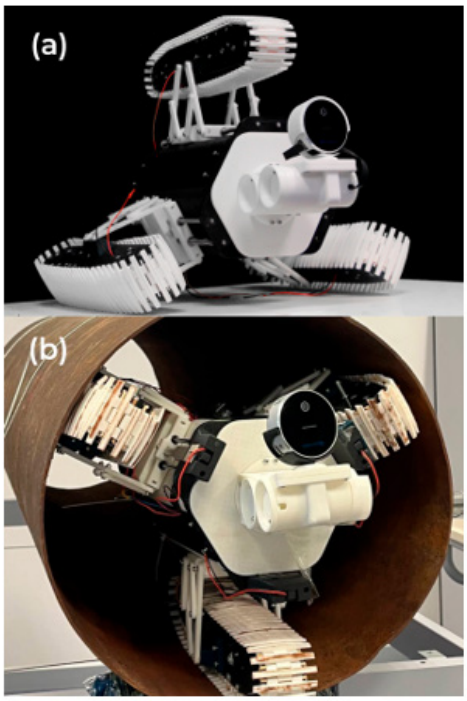


Fig. 2. Original images of the SIPIR prototype from the ISSAI report: (a) general view; (b) robot inside the pipe segment

Figure 3 shows examples of detecting abnormal areas in images of welded

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

joints and surface areas. The detection contour shows potentially defective areas, which can be used either as the final result of early detection or as input to a subsequent classification or segmentation module.



Fig. 3. Examples of highlighting abnormal areas and defective zones on surface and weld seam images

Figure 4 illustrates two weld seam recognition modes presented in the report: segmentation using DeepLabV 3+ and localization using YOLOv 8. Segmentation more accurately captures the seam shape, while detection defines a bounding box and is generally easier to quickly implement in a workflow. These images are useful for explaining the types of output generated by the intelligent subsystem.

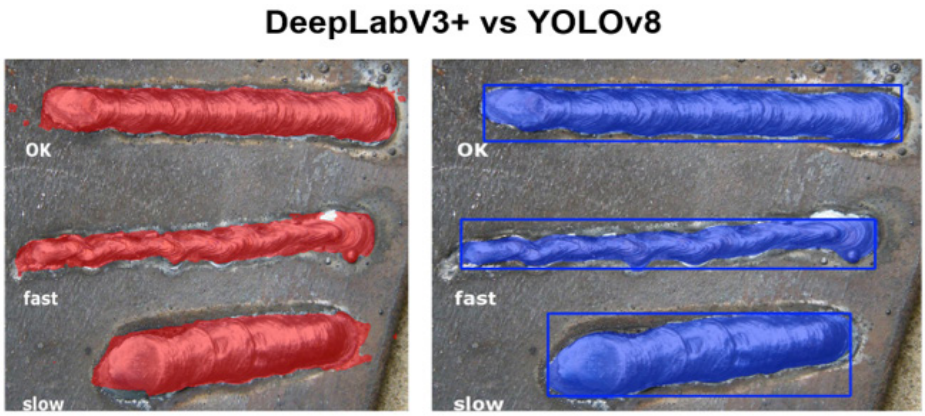


Fig. 4. An example of visual comparison of the results of DeepLabV 3+ and YOLOv 8 for weld seam recognition

Figure 5 shows examples of corrosion classification in internal pipe images. These visual examples demonstrate that the system is able to distinguish between frames containing signs of corrosion and those without any visible corrosion. However, to draw a final quantitative conclusion on the classification quality, it is necessary to provide additional numerical metric values on an independent test set.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

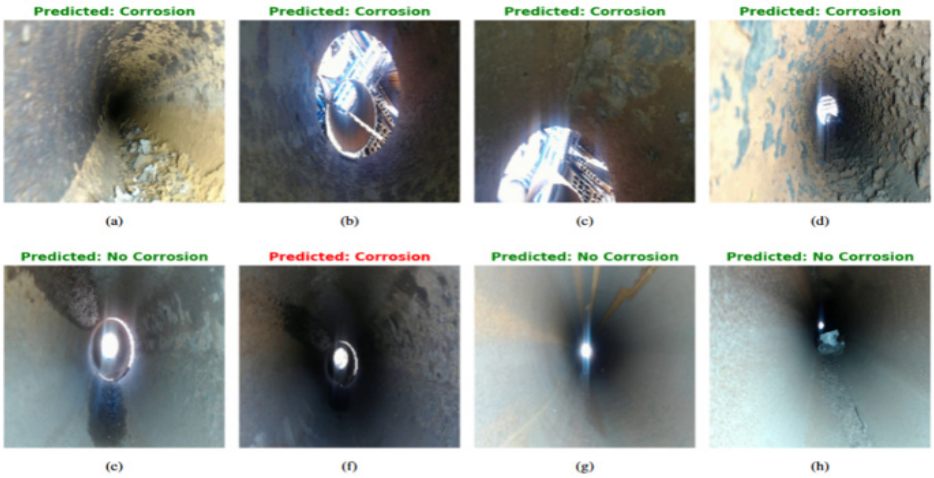


Fig. 5. Examples of classification of images of the internal surfaces of pipes: frames with the prediction " Corrosion " and " No Corrosion»

Jetson onboard computing module runs the computer vision algorithms, while the Arduino handles low-level actuator control and some sensor readings. This distribution of functions reduces the load on the main processor and allows critical motion loops to be isolated from resource-intensive image analysis operations.

Table 1. Main characteristics of the experimental complex

Parameter	Meaning / Characteristics
Weight prototype	about 4.8 kg
Traction effort	about 93 N
Worker range diameter pipes	350–450 mm
Testing segments	cast iron pipes with a diameter of 350–400 mm and a length of about 1.5 m
Visual sensor	endoscopic camera
RGB-D sensor	Intel RealSense L515, RGB + Depth + IR
Navigation data	IMU, odometry
Onboard calculations	Jetson and Arduino
3D mapping	RTAB-Map

Dataset and preprocessing

To train and validate the machine vision methods, the video system collected a dataset of 4,212 images obtained during inspection of 53 real pipes. The source report does not provide an exact distribution of the images by class, so this article does not include the unconfirmed values for the training, validation, and test sample sizes.

Pre-processing includes brightness and contrast correction, noise reduction, scaling, region-of-interest extraction, and augmentation. These operations are necessary due to the glare, shadows, low light, repetitive texture, and contamination typical of the in-pipe environment. The L 515 camera's infrared channel is used to maintain system performance in low-light conditions. Modern robotic systems also

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

utilize automatic and predictive exposure control methods based on deep learning to compensate for complex lighting conditions (Deng et al., 2024).

When creating a dataset, it's crucial to consider the variability of objects. Images of different pipes may differ in material, contamination level, surface color, illumination, weld geometry, and corrosion severity. Randomly distributing nearly identical frames of a single pipe between the training and test sets will overestimate the final accuracy. Therefore, for subsequent validation, it is recommended to separate the data at the pipe level rather than individual images.

The markings should include, at a minimum, the pipe identifier, section type, the presence or absence of anomalies, and the coordinates of the suspicious area. A pixel mask is additionally required for segmentation tasks. If multiple classes are used, uniform criteria for distinguishing between corrosion, weld defects, deposits, and background surface features must be established.

Augmentations should be approached with caution. Rotations, brightness and contrast changes, and the addition of moderate noise improve the stability of the model. However, excessive color or geometry changes can create physically unrealistic samples. The in-pipe environment has a pronounced axial structure, so the admissibility of each operation must be assessed taking into account the actual camera orientation.

Preprocessing may include brightness equalization, local contrast enhancement, impulse and Gaussian noise reduction, size normalization, and region-of-interest extraction. To assess robustness, it is advisable to create separate test subsets: normal illumination, low illumination, glare, contamination, and partial object occlusion.

Algorithm for detection and localization of defects

The computational pipeline implements a hybrid strategy. In the early stages, edge detection and clustering methods are used, in particular Canny and DBSCAN, to quickly identify potentially anomalous regions. The next stage utilizes lightweight convolutional networks suitable for on-board execution. The report mentions MobileNetV 2, as well as DeepLabV 3 and YOLOv 8 segmentation and detection models for weld seam analysis.

The algorithm's task is to isolate the weld joint, localize the potentially defective zone, and generate a conclusion about the object's condition. Interpretation is ensured by a visual representation of the mask or bounding box and the ability to compare the result with the weld geometry. The defect coordinates are combined with the robot's trajectory and a 3D model of the internal surface.

During the early detection stage, the image is converted into a representation suitable for highlighting local brightness variations and edges. The Canny operator generates a contour map, after which DBSCAN combines adjacent points into clusters. Small isolated groups can be discarded as noise, while stable clusters are passed to the next module as anomaly candidates.

The neural network component can solve three interconnected tasks. Classification determines the state of a frame or region of interest. Detection generates a bounding box around the defect. Segmentation identifies the precise shape of a weld or anomaly at the pixel level. Segmentation is the most informative for engineering

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

analysis, but it requires more labor-intensive labeling.

MobileNetV 2 is suitable for feature extraction on limited computing resources. DeepLabV 3 can be used for weld segmentation, and YOLOv 8 can be used for localizing individual defect zones. The final architecture should be selected based on a comparison of accuracy, latency, and memory consumption on the actual onboard module.

After obtaining a mask or frame, the center of the anomalous area is calculated. The resulting timestamp is compared with the robot's trajectory, generated using odometry, IMU, and depth data. The timestamp is then transferred to the 3D model. When the same area is detected repeatedly, similar timestamps must be merged to prevent the same defect from being detected multiple times.

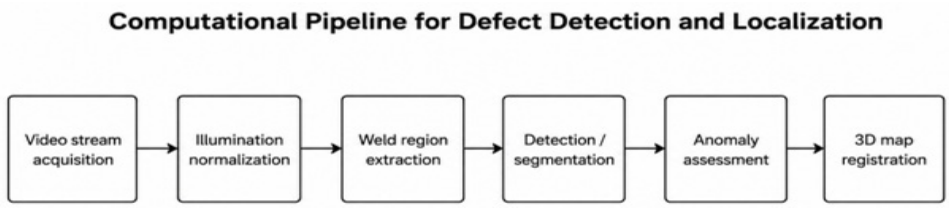


Fig. 6. Sequence of data processing when detecting defects

For a binary assessment of the quality of a welded joint, the model generates the probability that an image or a selected area belongs to a defective class:

$$\hat{y} = 1 \text{ if } p(x) \geq \tau ; \hat{y} = 0 \text{ if } p(x) < \tau ,$$

where p (x) is the probability estimate of a defective state, and τ is the decision threshold. Precision, Recall, F1 - score, IoU, and mAP should be used to evaluate the quality of detection and segmentation. However, the numerical values of these metrics are missing from the provided report, so they are not included in the results section and should be added after revalidating the model on a fixed test set.

Quality assessment criteria

Accuracy, Precision, Recall, and F1 - score are used to quantitatively evaluate classification. Accuracy alone is insufficient in industrial diagnostics, as a defect class is typically less frequently represented than normal samples. High recall is important for reducing the number of missed defects, and high positive prediction accuracy reduces the number of false alarms.

For segmentation, IoU and the Dice coefficient are used, characterizing the match between the predicted and reference regions. For detection, the average accuracy of the mAP is used at given IoU thresholds. In addition to recognition quality, it is necessary to measure the average frame processing time, frame rate, memory capacity, and stability under changing lighting conditions.

The system experiment should include not only algorithm metrics, but also robotic platform performance: weld passability, slippage, deviation from the pipe axis,

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

communication stability, video channel latency, and the accuracy of spatial referencing of the detected defect.

Note on the presence of metrics in the original report

A re-examination of the original report revealed references to Accuracy, Precision, Recall, F1 - score, IoU, Dice, and mAP as recommended evaluation criteria, but no specific numerical results for the models were provided in the report text. For this reason, the fictitious accuracy table is not included in this article. Instead, the table below details the models mentioned in the report and what information is documented about them.

Table 2. Information about the models and the presence of numerical metrics in the original report

Model / Method	Purpose	What confirmed report	Numerical metrics in the report
MobileNetV2	lightweight model for on-board recognition	The implementation of lightweight models and the use of a MobileNetV 2-like approach are indicated	Not given
DeepLab V3+	segmentation areas welded seam	There is visual examples recognition	Not given
YOLOv8	detection/localization of the weld area	There is visual examples recognition	Not given
Canny + DBSCAN	early detection and identification of abnormal zones	described as part of a hybrid pipeline	Not applicable / not applicable given
RTAB-Map	SLAM and 3D mapping	confirmed 3D mesh generation	Not given

Information model of inspection results

For practical use, results must be stored in a structured format. A minimal event record includes the survey ID, pipe ID, timestamp, frame number, coordinate along the trajectory, camera orientation, suspected anomaly type, confidence level, a link to the original image, and a depth map fragment. This structure ensures traceability and allows for re-analysis of disputed cases.

Within a single inspection, several adjacent frames may contain the same defect. Therefore, after frame-by-frame processing, event aggregation is required. Marks are combined based on coordinate proximity and similarity of visual features. The frame with the highest confidence, the coordinate range, and the number of confirmatory observations are saved for the final event.

The history of repeated inspections should be linked to a single pipeline section. This allows for the assessment of trends, such as increases in corrosion area, changes in weld geometry, or the emergence of new anomalies. Thus, the system can be used not only for one-time inspections but also for condition monitoring.

In the operator interface, it's advisable to separate automatic assessment and expert decision-making. The automatic model generates a proposal, and the expert

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

confirms or corrects it. Saving both labels is important for subsequent error analysis and further training.

Reliability and safety of operation

System reliability is determined not only by the accuracy of the neural network but also by the robustness of all components. Video signal loss, depth sensor error, track slippage, or computer overload can lead to missed sections. Therefore, the software must monitor the data feed rate, the temperature of the computer module, the connection quality, and the consistency of the odometry with the IMU.

If a critical condition is detected, the robot must enter a safe mode: stop, save the latest data, and transmit a message to the operator. To prevent the platform from being lost in the pipe, an emergency retrieval system, a distance limitation, and battery monitoring must be provided.

In-line inspection is often performed at hazardous facilities. Prior to implementation, the suitability of the electronics, enclosure sealing, and electrical safety must be assessed, taking into account the specific environment. The developed laboratory prototype validates the functional concept but is not a certified industrial device.

From the perspective of the intelligent module, a conservative decision-making mode is considered safe. If confidence is insufficient, the system should not issue a definitive conclusion. A more appropriate approach is to mark the area as requiring expert review and retain the full set of initial data.

Experimental setup

Tests were performed on a laboratory rig with segments of real cast iron pipes with a diameter of 350–400 mm and a length of approximately 1.5 m. The rig was used to test the stability of movement, weld seams, sensor operation, and the generation of a 3D model. An Ultimaker 3D printer was used to manufacture the prototype components. S 5. The OptiTrack system was used to independently evaluate latencies and movement accuracy in remotely controlled experiments.

The experiment included robot movement along linear segments, traversing localized irregularities and welds, acquiring RGB, depth, and infrared data, processing the onboard video stream, and constructing a 3D mesh. The endoscopic camera provided visual feedback with a latency of approximately 0.2 seconds. In the control loop based on the operator's head position, the average camera response time was 0.15 ± 0.25 seconds.

Before each test, the condition of the tracks, drives, sensor module mount, and video signal quality were checked. The robot was then positioned at the starting point of the pipe, after which odometry and orientation were recorded to zero. During movement, the video stream, depth, IMU readings, and control commands were recorded.

Rig tests allow for reproducible evaluation of basic functions, but do not fully simulate industrial conditions. A real pipeline may contain water, oil products, condensate, loose deposits, sharp bends, and longer sections. Therefore, rig results should be considered as proof of prototype performance and a basis for further field

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

testing.

To study the influence of illumination, it is advisable to conduct series with standard illumination, reduced brightness, and the use of an infrared channel. To assess the stability of recognition, it is necessary to preserve the same areas in different modes and compare the consistency of the results. This protocol is consistent with the need to separately assess the resilience of the visual system to changes in illumination (De Cubber et al., 2004; Collewet et al., 2011).

Results and discussion.

The prototype, weighing approximately 4.8 kg, generates a tractive force of approximately 93 N and is designed for use in pipes with diameters of 350–450 mm. The platform's ability to navigate welds and maintain control was confirmed on the test bench. The integrated sensor system provides simultaneous acquisition of visual data, depth, and motion parameters.

The resulting dataset includes 4,212 images from 53 pipes. Its practical value lies in the use of data obtained from real objects, not just synthetic or laboratory-produced samples. The feasibility of on-board recognition and early detection of anomalies was demonstrated. The system's response time during bench testing was less than one second, allowing the developed pipeline to be considered as the basis for operational operator support.

RealSense Applications L 515 and RTAB - Map allowed us to obtain point clouds and construct a 3D polygonal mesh of the pipe's interior. This fundamentally distinguishes the system from conventional video diagnostics: the defect is not only displayed on the frame but can also be linked to a spatial coordinate. This linking simplifies re-inspection, repair planning, and damage monitoring.

The obtained thrust and operating diameter parameters demonstrate that the platform has sufficient capacity to move its own weight and the sensor system within the laboratory area. The internal placement of the motors reduces the risk of mechanical damage and facilitates protection from the external environment. The curved tracks and spring-loaded expansion joints ensure more uniform contact with the surface.

Having two observation channels-visual and depth-increases the diagnostic information yield. The RGB frame is used for texture and color analysis, while depth is used for geometry reconstruction. In low-light conditions, the infrared channel can support surface structure detection, although the recognition performance of specific defect classes must be confirmed through separate testing.

Building a 3D mesh in RTAB - Map confirms the ability to spatially represent the interior environment. However, the cylindrical and repeating surface creates a risk of drift. To reduce errors, it is recommended to use smooth trajectories, integrate the IMU, and periodically perform snapping to stable geometric features, such as welds.

A response time of less than one second applies to the integrated early detection and control circuitry. This value demonstrates the fundamental feasibility of near-real-time operation, but does not replace a detailed latency profile. In the final version of the experiment, it is advisable to separately measure the frame capture, preprocessing, inference, result transmission, and operator display times.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

The resulting dataset of 4,212 images is a significant result, but its size alone does not guarantee the generalization ability of the model. The diversity of defects, class balance, labeling quality, and independence of the test set are crucial. Therefore, before publishing the final metrics, it is necessary to document the data composition and eliminate overlap between closely related frames between training and testing.

Table 3. Summary results of experimental testing

Verifiable component	Received result	Engineering meaning
Mobile platform	Works in the range of Ø350–450 mm; passing welded seams	Access to the internal surface without disassembling the pipeline
System computer vision	On-board recognition and early detection of anomalies	Operational warning operator
Kit data	4212 images from 53 pipes	A basis for training and subsequent validation of models
RGB-D and SLAM	Obtaining point clouds and 3D meshes	Spatial localization defects
Remote control	Delay video calls about 0.2 s	Support safe remote management
Reaction systems	Less than 1 sec in bench tests	Possibility of application in near real-time mode

Comparison of components and tasks to be solved

A mechanical platform solves the problem of delivering sensors to hard-to-reach areas. Its characteristics determine which areas can be realistically inspected. The sensor system generates an observable representation of the surface. Computer vision translates the image into a set of features and defect hypotheses. SLAM links the hypothesis to a spatial coordinate, and the operator interface converts the technical data into a maintenance solution.

This decomposition allows us to identify the sources of errors. If a defect isn't visible, the problem lies in the camera's trajectory or orientation. If the area is visible but not detected, model and data analysis are required. If a defect is found but incorrectly localized, odometry and SLAM must be improved. This facilitates system diagnostics and upgrade planning.

An integrated approach is especially important for pipelines where repeated access to a section is difficult. A simple video recording requires manual review and does not guarantee repeatable coordinates. Automatically tagging the site with a map reduces manual work and provides a basis for comparing inspections.

Practical application scheme

Before beginning an inspection, the operator specifies the pipeline ID, diameter, and starting point. The robot is inserted into the pipe and moves at a constant or adaptive speed. The system stores synchronized sensor data and flags suspicious areas. If a confidence threshold is exceeded, the operator receives a visual warning.

After completing the pass, a report is generated containing a list of events, frames, the type of suspected anomaly, the position on the trajectory, and a 3D map fragment. The engineer can confirm, reject, or modify the automatic flag. Confirmed

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

examples should be returned to the database for further model training.

For industrial implementation, the system must support logging, model version control, source data storage, and the ability to replay the solution. This is necessary for traceability and expert verification of diagnostic results.

A limitation of the study is the absence of a complete machine learning protocol in the report: class composition, sample splitting procedure, training parameters, and numerical values for Accuracy, Precision, Recall, F1 - score, IoU, and mAP. Therefore, the obtained results confirm the operability of the integrated system and the possibility of early detection, but do not allow for a rigorous quantitative comparative evaluation of neural network architectures. The next step should be validation on an independent test set and field testing on long pipelines with bends, contamination, and varying degrees of corrosion.

An additional limitation is the relatively short length of the test segments. Long sections increase the accumulation of odometry and SLAM errors, increasing communication and power requirements. Furthermore, the report does not include test results for pipes containing liquid, heavily contaminated, or sharp bends.

Recommendations for further research

To obtain publication- relevant quantitative results, it is recommended to develop a fixed data partitioning scheme, conduct relabeling by at least two experts, and assess the labeling consistency. After this, MobileNetV 2, YOLOv 8, and DeepLabV 3 should be compared in terms of precision, recall, F1 - score, IoU, mAP, inference time, and memory consumption.

Field tests should include pipes of varying lengths and diameters, sections with bends, moisture, contamination, and non-uniform illumination. To assess spatial accuracy, it is necessary to place control markers with known coordinates and compare them with the coordinates calculated by SLAM.

Active learning offers a promising approach. The model identifies the most uncertain images, an expert labels only these examples, and then the system undergoes further training. This approach allows for more efficient dataset expansion and reduces the amount of manual labeling.

To increase reliability, an ensemble combining a classical algorithm and a neural network solution can be used. Consistent results increase confidence, while discrepancies become grounds for expert verification.

Table 4. Extended validation program for an intelligent system

Direction checks	Necessary data	Expected result
Recognition	Independent test sample, expert labeling	Precision, Recall, F1, IoU, mAP
Onboard performance	Processing time and memory usage	Real-time confirmation of work
Navigation	Control coordinates inside pipes	Error spatial bindings
Mechanical cross-country ability	Series of passes over seams and uneven areas	Probability successful passages
Field conditions	Humidity, dirt, bending, different lighting conditions	Grade stability systems

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

The procedure for expert confirmation of defects is not described. For an objective assessment, it is necessary to compare the system's visual assessment with a reference inspection performed by a qualified specialist or an independent non-destructive testing method. Without such verification, it is more accurate to speak of anomaly detection rather than full confirmation of the defect type and severity.

Another area of development is uncertainty assessment. The model should report not only the class but also the confidence level. Low-confidence cases should be automatically communicated to the operator without making a hard decision, reducing the risk of erroneous rejection or missed damage.

Conclusion.

An intelligent pipeline defect detection system based on the SIPIR in-pipe robot has been developed and experimentally evaluated. The proposed system integrates a protected tracked robotic platform, RGB-D and infrared sensing, inertial sensors, onboard computing, a hybrid computer vision pipeline, and SLAM-based 3D mapping into a unified inspection framework.

The developed prototype, with a mass of approximately 4.8 kg, demonstrated stable operation in pipes with diameters of 350–450 mm. Experimental tests confirmed the robot's ability to pass weld joints, acquire synchronized visual and depth information, and perform early anomaly detection with a response time of less than one second. A dataset consisting of 4,212 images collected from 53 real pipes was formed and used as the basis for training and further validation of the computer vision models.

The integration of visual inspection with spatial mapping makes it possible not only to detect potentially defective areas but also to associate them with their location inside the pipeline. This feature increases the practical value of the system, since inspection results can be transformed into structured information suitable for repeated examination, maintenance planning, and subsequent technical decision-making.

The developed approach also demonstrates the advantages of combining lightweight onboard computer vision methods with depth sensing and SLAM. Such integration reduces the dependence of the inspection process on continuous manual observation and creates a basis for more autonomous and reproducible pipeline diagnostics. At the same time, the reliability of the final diagnostic decision depends on the quality of the image dataset, the stability of spatial localization, and the accuracy of the recognition algorithms.

The current study confirms the functional feasibility of the proposed robotic inspection concept under laboratory conditions. However, further work should include quantitative validation using Precision, Recall, F1-score, IoU, mAP, and inference-time measurements, as well as a detailed description of the training, validation, and test subsets. Additional field experiments are also required for pipes with different geometries, contamination levels, illumination conditions, and surface defects.

Thus, the proposed development should be considered not as an individual recognition algorithm, but as an integrated robotic inspection platform in which mechanical design, sensing, computer vision, onboard processing, and spatial mapping operate as a single system. This integrated architecture provides a promising basis for

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

preventive diagnostics, condition monitoring, and maintenance planning of pipeline infrastructure.

REFERENCES

Allibert, G., Courtial, E., & Chaumette, F. (2010). Predictive control for constrained image-based visual servoing // IEEE Transactions on Robotics. — Vol. 26. — No. 5. — Pp. 933–939. DOI: https://doi.org/10.1109/TRO.2010.2056590

Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-Up Robust Features (SURF) // Computer Vision and Image Understanding. — Vol. 110. — No. 3. — Pp. 346–359. DOI: https://doi.org/10.1016/j.cviu.2007.09.014

Chaumette, F., & Hutchinson, S. (2006). Visual servo control. I. Basic approaches // IEEE Robotics & Automation Magazine. — Vol. 13. — No. 4. — Pp. 82–90. DOI: https://doi.org/10.1109/MRA.2006.250573

Chaumette, F., & Hutchinson, S. (2007). Visual servo control. II. Advanced approaches // IEEE Robotics & Automation Magazine. — Vol. 14. — No. 1. — Pp. 109–118. DOI: https://doi.org/10.1109/MRA.2007.339609

Chesi, G., Hashimoto, K., Prattichizzo, D., & Vicino, A. (2004). Keeping features in the field of view in eye-in-hand visual servoing: A switching approach // IEEE Transactions on Robotics. — Vol. 20. — No. 5. — Pp. 908–914. DOI: https://doi.org/10.1109/TRO.2004.829456

Collewet, C., & Marchand, E. (2011). Photometric visual servoing // IEEE Transactions on Robotics. — Vol. 27. — No. 4. — Pp. 828–834. DOI: https://doi.org/10.1109/TRO.2011.2112593

Comport, A., Marchand, E., & Chaumette, F. (2006). Statistically robust 2D visual servoing // IEEE Transactions on Robotics. — Vol. 22. — No. 2. — Pp. 415–421. DOI: https://doi.org/10.1109/TRO.2006.870666

De Cubber, G., Berrabah, S. A., & Sahli, H. (2004). Color-based visual servoing under varying illumination conditions // Robotics and Autonomous Systems. — Vol. 47. — No. 4. — Pp. 225–249. DOI: https://doi.org/10.1016/j.robot.2004.03.015

De Luca, A., Oriolo, G., & Giordano, P. R. (2008). Feature depth observation for image-based visual servoing: Theory and experiments // The International Journal of Robotics Research. — Vol. 27. — No. 10. — Pp. 1093–1116. DOI: https://doi.org/10.1177/0278364908096706

Deng, W., Liu, Q., Pham, D. T., Hu, J., Lam, K.-M., Wang, Y., & Zhou, Z. (2024). Predictive exposure control for vision-based robotic disassembly using deep learning and predictive learning // Robotics and Computer-Integrated Manufacturing. — Vol. 85. — Article 102619. DOI: https://doi.org/10.1016/j.rcim.2023.102619

Dame, A., & Marchand, E. (2011). Mutual information-based visual servoing // IEEE Transactions on Robotics. — Vol. 27. — No. 5. — Pp. 958–969. DOI: https://doi.org/10.1109/TRO.2011.2147090

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INFORMATION SECURITY AND COMMUNICATIONTECHNOLOGIES
АҚПАРАТТЫҚ ҚАУІПСІЗДІК ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАРҒА АРНАЛҒАН
ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ И КОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 480–494
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.026
МРНТИ 28.23.15

MACHINE LEARNING FOR POLLUTION CLASSIFICATION IN ROCKET BOOSTER IMPACT AREAS

S.Z. Daruish[1]*, A.U. Utegenova[2], A.U. Kalizhanova[3], Zh.Zh. Zhumabekova[4]
[1]Al-Farabi Kazakh National University, Almaty, Kazakhstan;
[2]Kazakh National Research Technical University named after K.I. Satbayev, Almaty, Kazakhstan;
[3]Institute of Information and Computing Technologies, Almaty, Kazakhstan;
[4]Branch of the Republican State Enterprise «Infracos», Almaty, Kazakhstan.
E-mail: daruish_serik@live.kaznu.kz

Serik Z. Daruish — Doctoral student of the Faculty of Information Systems, Al-Farabi Kazakh National University, Almaty, Kazakhstan
E-mail: daruish_serik@live.kaznu.kz, https://orcid.org/0009-0006-1878-4212;
Anar U. Utegenova — PhD, Associate Professor, Head of the Software Engineering Department, Kazakh National Research Technical University named after K.I. Satbayev, Almaty, Kazakhstan
https://orcid.org/0000-0002-3875-5164;
Aliya U. Kalizhanova — Candidate of Physical and Mathematical Sciences, Professor, Institute of Information and Computing Technologies, Almaty, Kazakhstan
https://orcid.org/0000-0002-5979-9756;
Zhazira Zh. Zhumabekova — PhD, Head of the Scientific and Technical Information Department, Branch of the Republican State Enterprise «Infracos», Almaty, Kazakhstan
https://orcid.org/0000-0003-1854-798X.

© S.Z. Daruish, A.U. Utegenova, A.U. Kalizhanova, Zh.Zh. Zhumabekova

Abstract. Environmental monitoring of rocket booster impact areas requires assessing soil contamination with unsymmetrical dimethylhydrazine (UDMH) and N-nitrosodimethylamine (NDMA), which are costly to determine. We tested whether such contamination can be predicted from indirect indicators: sample position relative

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

to the impact centre, sampling depth, pH, nitrates, nitrites and petroleum hydrocarbons. An open dataset from the Soyuz-FG launch vehicle crash site in Kazakhstan was used: 53 soil samples with measured UDMH and NDMA, 63 records for 2018 and 2019. UDMH and NDMA were excluded from the features. Logistic regression, k-nearest neighbours, random forest and gradient boosting were compared with a distance rule under repeated, spatial and temporal cross-validation. Logistic regression and random forest reached ROC-AUC 0.75–0.76 versus 0.59 for the distance rule, and 0.85–0.87 on the 2018 samples with fully measured chemistry. Nitrate content and sampling depth were the most informative features. The model can select samples requiring costly analysis in GIS-based monitoring.

Keywords: machine learning, rocket booster impact areas, unsymmetrical dimethylhydrazine, soil contamination classification, spatial cross-validation, geographic information systems

For citation: Daruish S.Z., Utegenova A.U., Kalizhanova A.U., Zhumabekova Zh.Zh. (2026). Machine learning for pollution classification in rocket booster impact areas // International Journal of Information and Communication Technologies. Vol. 7. No. 27. Pp. 480–494. https://doi.org/10.54309/IJICT.2026.27.3.026. (In Russ.).

Conflict of interest: The authors declare that there is no conflict of interest.

Acknowledgements: *This research was carried out with the support of the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan under Grant No. AP23488291, "Development of a Multifunctional Resource for Environmental Certification of Rocket Carrier Separated-Stage Impact Areas Using Adaptive Representation of Interactive GIS."*

ЗЫМЫРАН ТАСЫМАЛДАУШЫЛАРДЫҢ ҚҰЛАУ АЙМАҚТАРЫНЫҢ ЛАСТАНУЫН МАШИНАЛЫҚ ОҚЫТУМЕН ЖІКТЕУ

С.З. Дәруіш[1]*, А.У. Утегенова[2], А.У. Калижанова[3], Ж.Ж. Жумабекова[4]

[1]Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан;
[2]Қ.И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті, Алматы, Қазақстан;
[3]Ақпараттық және есептеуіш технологиялар институты, Алматы, Қазақстан;
[4]«Инфракос» республикалық мемлекеттік кәсіпорнының филиалы, Алматы, Қазақстан.
E-mail: daruish_serik@live.kaznu.kz

Дәруіш Серік Заңғарұлы — «Ақпараттық жүйелер» факультетінің докторанты, Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан
E-mail: daruish_serik@live.kaznu.kz, https://orcid.org/0009-0006-1878-4212;
Утегенова Анара Урантаевна — PhD, қауымдастырылған профессор, «Бағдарламалық инженерия» кафедрасының меңгерушісі, Қ.И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті, Алматы, Қазақстан

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

https://orcid.org/0000-0002-3875-5164;
Калижанова Алия Уалиевна — физика-математика ғылымдарының кандидаты, профессор, Ақпараттық және есептеуіш технологиялар институты, Алматы, Қазақстан
https://orcid.org/0000-0002-5979-9756;
Жумабекова Жазира Жаксиликовна — PhD, «Инфракос» РМК филиалының ғылыми-техникалық ақпарат бөлімінің бастығы, Алматы, Қазақстан
https://orcid.org/0000-0003-1854-798X.

© С.З. Дәруіш, А.У. Утегенова, А.У. Калижанова, Ж.Ж. Жумабекова

Аннотация. Зымыран тасымалдаушылардың бөлінетін бөліктері құлайтын аудандардың экологиялық мониторингі топырақтың симметриясыз диметилгидразинмен (НДМГ) және N-нитрозодиметиламинмен (NDMA) ластануын бағалауды талап етеді, ал бұл көрсеткіштерді анықтау қымбат әрі еңбекті көп қажет етеді. Жұмыстың мақсаты – мұндай ластануды жанама көрсеткіштер бойынша болжауға болатынын тексеру: сынаманың құлау орталығына қатысты орналасуы, іріктеу тереңдігі, pH, нитраттар, нитриттер мен мұнай өнімдерінің мөлшері. Қазақстандағы «Союз-ФГ» зымыран тасымалдаушысы апат орнын зерттеудің ашық деректер жиынтығы пайдаланылды: НДМГ мен NDMA өлшенген 53 топырақ сынамасы, 2018 және 2019 жылдардағы 63 жазба. НДМГ мен NDMA белгілер қатарынан алынып тасталды. Логистикалық регрессия, k жақын көршілер әдісі, кездейсоқ орман және градиенттік бустинг қашықтық ережесімен қайталанатын, кеңістіктік және уақыттық кросс-валидация арқылы салыстырылды. Логистикалық регрессия мен кездейсоқ орман ROC-AUC 0,75–0,76 көрсетті, қашықтық ережесі – 0,59, ал химиясы толық өлшенген 2018 жылғы сынамаларда – 0,85–0,87. Ең ақпаратты белгілер – нитраттар мөлшері мен іріктеу тереңдігі. Модель мониторингтің геоақпараттық жүйелерінде қымбат талдауды қажет ететін сынамаларды іріктеу үшін қолданылуы мүмкін.

Түйінді сөздер: машиналық оқыту, зымыран тасымалдаушылардың құлау аудандары, симметриясыз диметилгидразин, топырақ ластануын жіктеу, кеңістіктік кросс-валидация, геоақпараттық жүйелер

Дәйексөздер үшін: Дәруіш С.З., Утегенова А.У., Калижанова А.У., Жумабекова Ж.Ж. (2026). Зымыран тасымалдаушылардың құлау аймақтарының ластануын машиналық оқытумен жіктеу // Халықаралық ақпараттық және коммуникациялық технологиялар журналы. Т. 7. No 27. Б. 480–494. https://doi.org/10.54309/IJICT.2026.27.3.026. (Орыс. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

Алғыс: *Зерттеу Қазақстан Республикасы Ғылым және жоғары білім министрлігі Ғылым комитетінің № AP23488291 «Зымыран-тасығыштардың бөлініп қалған бөліктерінің құлау аудандарын интерактивті геоақпараттық жүйелерді (ГАЖ) бейімдеп ұсыну әдісі арқылы экологиялық паспорттау үшін*

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

көпфункционалды ресурсты әзірлеу» гранты аясында жүргізілді.

МАШИННОЕ ОБУЧЕНИЕ ДЛЯ КЛАССИФИКАЦИИ ЗАГРЯЗНЕНИЯ РАЙОНОВ ПАДЕНИЯ РАКЕТ-НОСИТЕЛЕЙ

С.З. Дәруіш[1]*, А.У. Утегенова[2], А.У. Калижанова[3], Ж.Ж. Жумабекова[4]
[1]Казахский национальный университет имени аль-Фараби, Алматы, Казахстан;
[2]Казахский национальный исследовательский технический университет имени К.И. Сатпаева, Алматы, Казахстан;
[3]Институт информационных и вычислительных технологий КН МНВО РК, Алматы, Казахстан;
[4]Республиканское государственное предприятие «Инфракос», Алматы, Казахстан.
E-mail: daruish_serik@live.kaznu.kz

Дәруіш Серік Заңғарұлы — докторант факультета «Информационные системы», Казахский национальный университет имени аль-Фараби, Алматы, Казахстан
E-mail: daruish_serik@live.kaznu.kz, https://orcid.org/0009-0006-1878-4212;
Утегенова Анара Урантаевна — PhD, ассоциированный профессор, зав. кафедрой программная инженерия, Казахский национальный исследовательский технический университет имени К.И. Сатпаева, Алматы, Казахстан
https://orcid.org/0000-0002-3875-5164;
Калижанова Алия Уалиевна — к.ф.-м.н., профессор, Институт информационных и вычислительных технологий, Алматы, Казахстан
https://orcid.org/000-0002-5979-9756;
Жумабекова Жазира Жаксиликовна — PhD, начальник отдела научно-технической информации, филиал РГП «Инфракос», Алматы, Казахстан
https://orcid.org/0000-0003-1854-798X.

© С.З. Дәруіш, А.У. Утегенова, А.У. Калижанова, Ж.Ж. Жумабекова

Аннотация. Экологический мониторинг районов падения отделяющихся частей ракет-носителей требует оценки загрязнения почв несимметричным диметилгидразином (НДМГ) и N-нитрозодиметиламином (NDMA), определение которых трудоёмко и дорого. Цель работы состояла в проверке того, можно ли предсказать факт такого загрязнения по косвенным показателям: положению пробы относительно центра падения, глубине отбора, pH, содержанию нитратов, нитритов и нефтепродуктов. Использован открытый датасет обследования места аварии ракеты-носителя «Союз-ФГ» в Казахстане: 53 пробы почвы с измеренными НДМГ и NDMA, 63 записи за 2018 и 2019 гг. Сами НДМГ и NDMA из признаков исключены. Логистическая регрессия, метод k ближайших соседей, случайный лес и градиентный бустинг сравнены с правилом по расстоянию при

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

повторной, пространственной и временной кросс-валидации. Логистическая регрессия и случайный лес достигли ROC-AUC 0,75–0,76 против 0,59 у правила по расстоянию, а на пробах 2018 г. с полностью измеренной химией 0,85–0,87. Наиболее информативны содержание нитратов и глубина отбора. Модель применима для отбора проб, требующих дорогостоящего анализа, в геоинформационных системах мониторинга.

Ключевые слова: машинное обучение, районы падения ракет-носителей, несимметричный диметилгидразин, классификация загрязнения почв, пространственная кросс-валидация, геоинформационные системы

Для цитирования: Дәруіш С.З., Утегенова А.У., Калижанова А.У., Жумабекова Ж.Ж. (2026). Машинное обучение для классификации загрязнения районов падения ракет-носителей // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 27. Стр. 480–494. https://doi.org/10.54309/IJICT.2026.27.3.026. (На русс.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Благодарности: *Исследования проведены при поддержке Комитета науки Министерства науки и высшего образования Республики Казахстан в рамках гранта № AP23488291 «Разработка многофункционального ресурса экологической паспортизации районов падения отделяющихся частей ракет-носителей методом адаптивного представления интерактивных ГИС».*

Введение.

Экологический мониторинг территорий, подверженных воздействию ракетно-космической деятельности, - актуальная задача обеспечения экологической безопасности. Космодром Байконур десятилетиями используется для запусков ракет-носителей, отработавшие ступени которых падают на специально выделенные участки - районы падения. На территории Республики Казахстан размещено 53 района падения общей площадью 4,94 млн га. Их значительная площадь и рассредоточенность делают систематический экологический контроль чрезвычайно трудоемким.

Падение отработавших ступеней сопровождается загрязнением окружающей среды высокотоксичным ракетным топливом - несимметричным диметилгидразином (НДМГ, UDMH) и продуктами его трансформации, ключевую роль среди которых играет N-нитрозодиметиламин (NDMA) - стабильный высокотоксичный канцерогенный продукт с высокой персистентностью (Королева и др., 2023; Milyushkin et al., 2023). Восстановление территорий требует методов биоремедиации (Bekeshev et al., 2024), что делает актуальной задачу точной и оперативной оценки степени загрязнения.

Оценка загрязнения основана на отборе проб и лабораторном определении НДМГ и NDMA. Эти анализы требуют специализированного оборудования, дороги и выполняются с задержкой, тогда как pH, содержание нитратов, нитритов и нефтепродуктов определяются в обычной лаборатории, а положение про-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

бы относительно места падения и глубина отбора известны сразу. Применение геоинформационных технологий для экологической оценки территорий рассмотрено в работах (Павлова и др., 2024; Кузнецова и др., 2024; Сегал и др., 2023). Методы машинного обучения позволяют связать совокупность косвенных показателей с фактом загрязнения без явного задания пороговых правил, их теоретические основы изложены в работах (Cover et al., 1967; Breiman, 2001; Friedman, 2001; Hastie et al., 2009). При этом для данных с пространственной структурой обычная случайная кросс-валидация завышает качество моделей, поскольку соседние пробы попадают одновременно в обучение и проверку, и рекомендуются блочные схемы проверки (Roberts et al., 2017), а при редких положительных примерах доля правильных ответов теряет смысл и предпочтительны ROC-AUC и площадь под кривой точность-полнота (Saito, 2015).

Работы по машинному обучению для районов падения немногочисленны, а в опубликованных сравнениях алгоритмов классы загрязнения нередко задаются по тем же показателям, которые подаются моделям на вход. В таком случае модель воспроизводит пороговое правило, и высокая точность ничего не говорит о её прогностической способности. Открытым остаётся вопрос, несут ли косвенные показатели и положение пробы достаточно информации, чтобы предсказать загрязнение НДМГ и NDMA без их прямого определения.

Объектом исследования являются почвы места падения отделяющихся частей ракеты-носителя, предметом - методы классификации загрязнения по косвенным показателям. И Цель исследования - проверить, можно ли по положению пробы, глубине отбора и общим химическим показателям почвы предсказать факт загрязнения НДМГ и NDMA с качеством, превышающим простое правило по расстоянию от места падения, и оценить, какие показатели для этого наиболее информативны. Гипотеза исследования: косвенные показатели содержат сигнал о загрязнении, и модели машинного обучения ранжируют пробы по риску загрязнения лучше правила по расстоянию при проверке, учитывающей пространственную структуру данных.

Материалы и методы.

Четыре алгоритма классификации представляют различные семейства методов: линейные модели (логистическая регрессия), метрические непараметрические (k ближайших соседей), ансамбли на основе бэггинга (случайный лес) и бустинга (градиентный бустинг). Их сравнение с базовым правилом по расстоянию показывает, добавляют ли косвенные химические показатели что-либо к положению пробы.

Исходные данные - открытый датасет обследования места аварии ракеты-носителя «Союз-ФГ» в Казахстане (Semenkov et al., 2024; Bekeshev et al., 2024). Он содержит 227 проб почвы, отобранных на расстояниях от 0 до 50 м от центра падения баков с тремя слоями по глубине, с результатами определения pH, нитратов, нитритов и нефтепродуктов двумя лабораториями в осеннюю и весеннюю съёмки, а также НДМГ и NDMA в 2018 и 2019 гг. Существенно, что НДМГ и NDMA определены не во всех пробах: пометка «nd» в датасете означает,

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

что анализ не выполнялся, а не что вещество не обнаружено. Поэтому в разметке использованы только пробы с фактическими измерениями: 26 записей 2018 г. и 37 записей 2019 г., всего 63 записи по 53 пробам, из которых 10 проб измерены в оба года. В предыдущей редакции работы значения «nd» трактовались как значения ниже предела обнаружения, что завышало число незагрязнённых проб, и этот недостаток здесь устранён.

Классы загрязнения заданы по пороговым значениям НДМГ и NDMA, принятым в санитарно-гигиенической практике: «низкий» при НДМГ < 0,05 и NDMA < 0,05 мг/кг, «средний» при 0,05–0,5 мг/кг хотя бы по одному веществу, «высокий» при значении выше 0,5 мг/кг. Среди 63 записей 44 относятся к классу «низкий», 13 к «среднему» и 6 к «высокому». Поскольку шесть записей класса «высокий» не позволяют обучить отдельный класс, в качестве целевой переменной принят бинарный признак «загрязнено» (НДМГ или NDMA не ниже 0,05 мг/кг), положительных записей 19 (30 %). Такое объединение сохраняет практический смысл задачи: выявить пробы, для которых загрязнение подтверждено.

Признаки не включают НДМГ и NDMA. Использованы семь косвенных показателей: расстояние от центра падения (для 15 проб, где оно не записано, восстановлено по координатам), глубина отбора (слой 1–3), год отбора, pH, содержание нитратов, нитритов и нефтепродуктов того же года измерения, а при их отсутствии - данные съёмки того же сезона. Расстояние и концентрации логарифмированы, значения ниже предела обнаружения заменены половиной предела. Для записей 2019 г. химические показатели известны лишь частично, поэтому пропуски заполнялись медианой с добавлением индикатора пропуска. Заполнение пропусков и стандартизация выполнялись внутри конвейера обучения на каждом обучающем фолде отдельно, чтобы исключить утечку информации из проверочной части. Один и тот же набор признаков использован во всех моделях, при анализе важности и в кластерном анализе. Дополнительно рассмотрен сокращённый набор только из положения пробы (расстояние, глубина, год), который показывает вклад химических показателей.

Для классификации применялись четыре алгоритма. Логистическая регрессия (Logistic Regression) служит базовой линейной моделью и оценивает вероятность класса через логистическую функцию. Метод k ближайших соседей (k-NN) относит объект к классу по k ближайшим обучающим примерам с весами, обратными расстоянию (Cover et al., 1967). Случайный лес (Random Forest) (Breiman, 2001) - ансамбль решающих деревьев на бутстрэп-выборках со случайным подмножеством признаков в узле, качество разбиения оценивается индексом неоднородности Джини (1):

$$G = 1 - \Sigma_k \, p_k^2 \qquad (1)$$

где p_k - доля объектов класса k в узле. Градиентный бустинг (Gradient Boosting) (Friedman, 2001) строит ансамбль последовательно, аппроксимируя антиградиент функции потерь, использована реализация на гистограммах. Дис-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

баланс классов учтён весами классов, обратно пропорциональными их частоте, во всех моделях, где это предусмотрено. Важность признаков оценивалась перестановочным методом на отложенных фолдах: признак случайно перемешивался в проверочной части, и фиксировалось снижение ROC-AUC, что не зависит от способа построения дерева и применимо к любой модели (Hastie et al., 2009).

Гиперпараметры заданы заранее и не подбирались по проверочным данным: k = 5 для k-NN, 500 деревьев и минимум два объекта в листе для случайного леса, глубина 3, шаг обучения 0,05 и 150 итераций для градиентного бустинга, параметр регуляризации C = 1 для логистической регрессии. Базовое правило ранжирует пробы по близости к центру падения, а в бинарной форме относит к загрязнённым пробы в пределах 0,5 м от центра. Реализация выполнена на Python 3.9 с использованием scikit-learn 1.6 (Pedregosa et al., 2011), генератор случайных чисел фиксирован.

Кластерный анализ методом k-средних (MacQueen, 1967) выполнен по тем же косвенным признакам без НДМГ и NDMA, чтобы проверить, группируются ли пробы по общим химическим показателям и положению так, что группы совпадают с фактом загрязнения. Число кластеров варьировалось от 2 до 6, оценивались инерция (метод локтя) и коэффициент силуэта (Rousseeuw, 1987; Chicco et al., 2025), вычисляемый для каждого объекта по формуле (2):

$$s(i) = (b(i) - a(i)) / \max(a(i), b(i)) \quad (2)$$

где a(i) - среднее расстояние до объектов своего кластера, b(i) - минимальное среднее расстояние до объектов ближайшего кластера. Согласие кластеров с фактом загрязнения оценивалось таблицей сопряжённости и скорректированным индексом Рэнда (ARI). Кластеризация выполнена для 63 размеченных записей и отдельно для 111 проб с полной осенней съёмкой одной лаборатории.

Качество классификации оценивалось по ROC-AUC и площади под кривой точность-полнота (PR-AUC), которые не зависят от порога и устойчивы к дисбалансу (Saito et al., 2015), случайный уровень PR-AUC равен доле положительных записей. Дополнительно при пороге 0,5 рассчитывались сбалансированная точность (3) и полнота класса «загрязнено»:

$$BA = (TPR + TNR) / 2 \quad (3)$$

где TPR - доля верно выявленных загрязнённых записей, TNR - доля верно выявленных незагрязнённых. Основной протокол проверки - повторная стратифицированная кросс-валидация по группам: 5 фолдов, 20 повторов с разными разбиениями, при этом записи одной пробы за 2018 и 2019 гг. всегда попадают в один фолд. По повторам приводятся среднее и интервал между 2,5-м и 97,5-м процентилями. Пространственная проверка выполнена по схеме «одна группа расстояния целиком в проверке» для четырёх групп: 0–0,5 м, 2,5–5 м, 7,5–15 м и 20 м и дальше (Roberts et al., 2017). Временная проверка - обучение на записях

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

2018 г. и проверка на записях 2019 г. Отдельно оценена подвыборка 2018 г. с полностью измеренной химией (26 записей, 10 загрязнённых), где заполнение пропусков не требовалось.

Результаты и обсуждение.

Результаты повторной кросс-валидации представлены в таблице 1. Правило по расстоянию ранжирует записи лишь немного лучше случайного (ROC-AUC 0,59), модели только по положению пробы дают 0,70–0,72, а добавление химических показателей поднимает ROC-AUC логистической регрессии до 0,76 и случайного леса до 0,75 при PR-AUC 0,65–0,68 против случайного уровня 0,30. Интервалы по повторам у этих двух моделей не пересекаются с результатом правила. Градиентный бустинг на 63 записях переобучается и не превосходит случайный уровень, что согласуется с малым объёмом выборки.

Таблица 1. Качество классификации при повторной стратифицированной кросс-валидации (5 фолдов × 20 повторов)

Признаки	Модель	ROC-AUC	PR-AUC	Сбалансированная точность
Положение	Логистическая регрессия	0,71 (0,66–0,73)	0,54 (0,46–0,59)	0,61
Положение	k ближайших соседей	0,70 (0,67–0,74)	0,51 (0,44–0,57)	0,70
Положение	Случайный лес	0,72 (0,64–0,77)	0,59 (0,49–0,65)	0,64
Положение	Градиентный бустинг	0,55 (0,45–0,67)	0,42 (0,29–0,54)	0,53
Положение + химия	Логистическая регрессия	0,76 (0,71–0,79)	0,65 (0,60–0,69)	0,68
Положение + химия	k ближайших соседей	0,73 (0,67–0,77)	0,54 (0,44–0,61)	0,62
Положение + химия	Случайный лес	0,75 (0,66–0,81)	0,68 (0,59–0,74)	0,67
Положение + химия	Градиентный бустинг	0,54 (0,46–0,64)	0,40 (0,31–0,48)	0,53
Правило по расстоянию	-	0,59	0,36	0,57

Кривые точность-полнота (рисунок 1) показывают, что преимущество моделей с химическими показателями сохраняется во всём диапазоне полноты: при полноте 0,5 точность случайного леса и логистической регрессии составляет 0,55-0,65, правила по расстоянию 0,37. Логистическая регрессия и случайный лес дают близкие результаты, и для внедрения предпочтительна линейная модель как более простая и устойчивая при малом числе наблюдений (Hastie et al., 2009).

Проверки устойчивости сведены в таблицу 2. При пространственной проверке, когда целая группа расстояния уходит в проверочную часть, логистическая регрессия сохраняет ROC-AUC 0,72 и PR-AUC 0,61, метод k ближайших соседей 0,73 и 0,48, случайный лес 0,68 и 0,48. На подвыборке 2018 г. с полностью измеренной химией качество заметно выше: логистическая регрессия 0,87 (0,82–0,91), случайный лес 0,85 (0,80-0,89) против 0,76 у правила по расстоянию

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

и 0,77 у модели только по положению. Перенос во времени не удался: модели, обученные на записях 2018 г., ранжируют записи 2019 г. не лучше случайного (ROC-AUC 0,36–0,64). В 2019 г. пробы отбирались в других точках, химические показатели для них определены лишь частично, а загрязнение выявлено в основном по NDMA, поэтому связь между косвенными показателями и загрязнением изменилась.

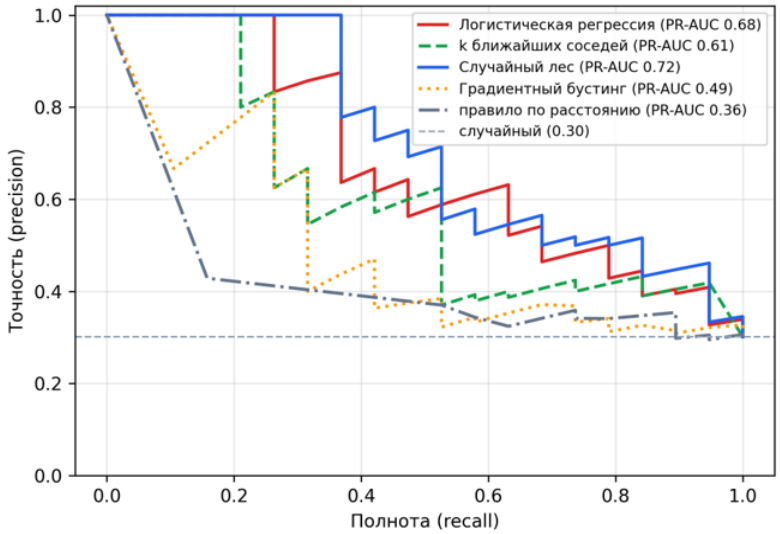


Рис. 1. Кривые точность-полнота моделей и правила по расстоянию (положение + химия, оценки вне фолда)

Таблица 2. Проверки устойчивости: пространственная, временная и подвыборка 2018 г.

Проверка	Модель	ROC-AUC	PR-AUC
Пространственная (по группам расстояния)	Логистическая регрессия	0,72	0,61
Пространственная (по группам расстояния)	k ближайших соседей	0,73	0,48
Пространственная (по группам расстояния)	Случайный лес	0,68	0,48
Временная (обучение 2018, проверка 2019)	Логистическая регрессия	0,64	0,34
Временная (обучение 2018, проверка 2019)	Случайный лес	0,43	0,32
Временная (обучение 2018, проверка 2019)	Правило по расстоянию	0,42	0,22
Подвыборка 2018 г., полная химия (n = 26)	Логистическая регрессия	0,87 (0,82–0,91)	0,82 (0,76–0,88)
Подвыборка 2018 г., полная химия (n = 26)	Случайный лес	0,85 (0,80–0,89)	0,85 (0,80–0,91)
Подвыборка 2018 г., полная химия (n = 26)	Только положение, логистическая регрессия	0,77 (0,71–0,81)	0,72 (0,64–0,78)
Подвыборка 2018 г., полная химия (n = 26)	Правило по расстоянию	0,76	0,63

Перестановочная важность признаков случайного леса (рисунок 2) показывает, что наибольший вклад вносят содержание нитратов (снижение ROC-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

AUC на 0,13 при перемешивании) и глубина отбора (0,12), далее расстояние от центра (0,04) и нефтепродукты (0,04). Вклад pH, нитритов и года отбора не отличим от нуля. Половина суммарной важности приходится на два показателя, что допускает упрощённую схему предварительной оценки по глубине и содержанию нитратов.

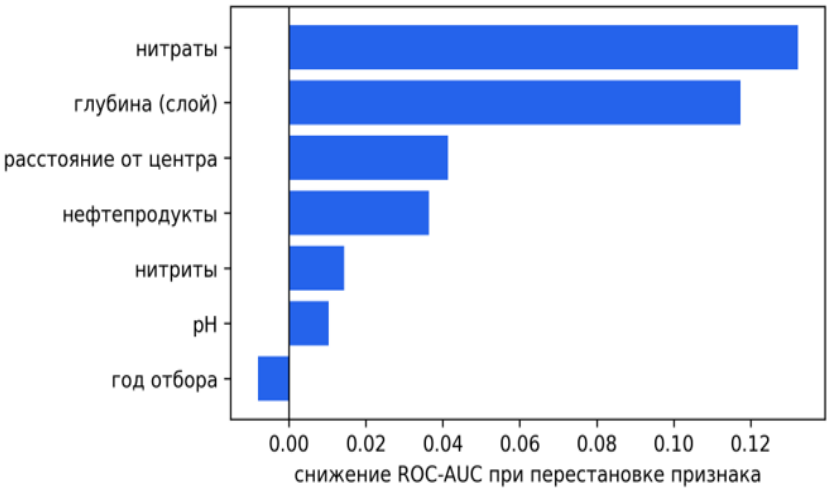


Рис. 2. Перестановочная важность признаков случайного леса на отложенных фолдах

Ведущая роль нитратов согласуется с представлениями о трансформации НДМГ в почве: окисление НДМГ и его продуктов сопровождается образованием азотсодержащих соединений, и повышенное содержание нитратов служит следом попадания топлива (Королева и др., 2023; Milyushkin et al., 2023), а загрязнение снегового покрова и почв азотсодержащими веществами при пусках описано в работе (Sharapova et al., 2020). Роль глубины отражает поверхностный характер загрязнения: в верхнем слое доля загрязнённых записей достигает 62 % у центра падения, во втором слое загрязнение отмечено только у центра, в третьем не выявлено. Приоритет контроля продуктов трансформации при эколого-гигиеническом мониторинге отмечен и в работе (Зяблицкая и др., 2023).

Кластерный анализ по косвенным признакам без НДМГ и NDMA (рисунок 3) не выделяет загрязнённые пробы как отдельную группу. Для 63 размеченных записей наибольший коэффициент силуэта 0,58 достигается при двух кластерах, но второй кластер состоит из двух проб с содержанием нитратов около 3000 мг/кг, а остальные записи, включая 17 из 19 загрязнённых, попадают в один кластер, скорректированный индекс Рэнда с фактом загрязнения равен 0,08. Для 111 проб с полной осенней съёмкой силуэт не превышает 0,30, кластеры отражают глубину и удалённость проб, а не загрязнение.

Таким образом, без учителя структура косвенных показателей воспроизводит схему отбора, а не факт загрязнения, и высокие значения силуэта при двух кластерах в подобных данных возникают из-за нескольких аномальных проб. В практике мониторинга кластеризация пригодна для выявления таких аномалий,

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

тогда как зонирование по риску загрязнения требует обучения с учителем.

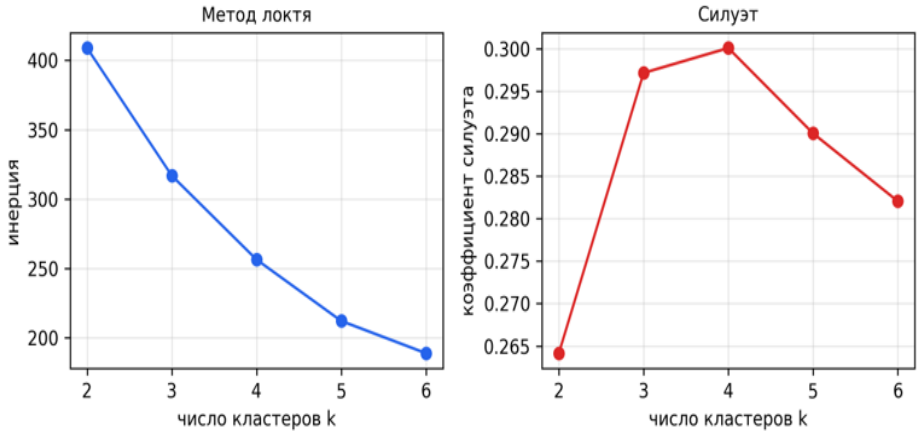


Рис. 3. Метод локтя и коэффициент силуэта для кластеризации 111 проб по косвенным признакам осенней съёмки

Схема зон загрязнения по расстоянию от центра и глубине отбора приведена на рисунке 4. Слева показана фактическая доля загрязнённых записей, справа средняя оценка риска случайного леса, полученная вне фолда, то есть для каждой записи моделью, которая её не видела. Модель воспроизводит основные черты: наибольший риск в верхнем слое у центра падения (0,58 при факте 0,62), заметный риск в верхнем слое на удалении 7,5–15 м (0,41 при факте 0,45) и низкий риск в нижних слоях. Расхождения проявляются во втором и третьем слоях на удалении, где модель даёт 0,2–0,3 при фактическом нуле, но число записей там мало. Координаты проб в датасете записаны с точностью около 3 м, что грубее шага отбора, поэтому схема построена по группам расстояния, а не по координатам.

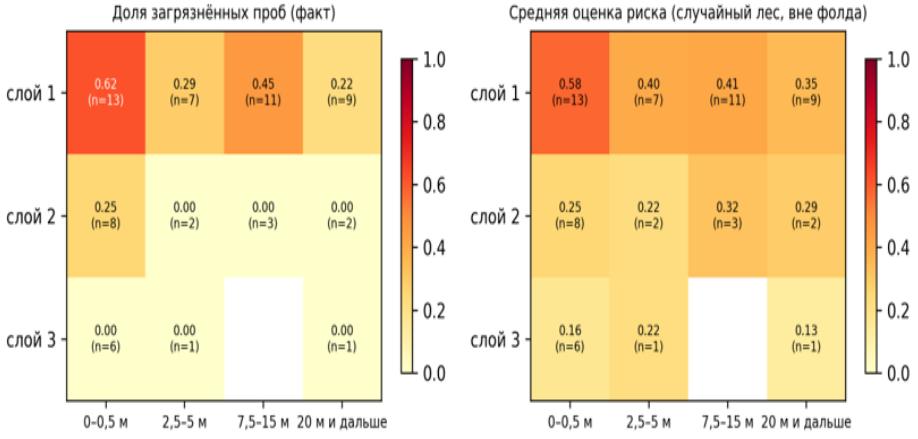


Рис. 4. Схема зон загрязнения по расстоянию от центра падения и глубине: факт и оценка риска вне фолда

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

С практической точки зрения модель по косвенным показателям применима как инструмент отбора проб для дорогостоящего анализа: при пороге, отвечающем полноте 0,8, в проверку на НДМГ и NDMA направляется около половины проб вместо всех, и при этом пропускается пятая часть загрязнённых. В геоинформационной системе мониторинга такая оценка риска может рассчитываться по данным съёмки сразу после отбора и отображаться по схеме отбора, а окончательное решение о зоне загрязнения принимается по результатам прямого анализа. Логистическая регрессия предпочтительна из-за простоты и интерпретируемости коэффициентов.

Ограничения исследования связаны с объёмом данных. Размеченных записей 63, загрязнённых 19, поэтому интервалы оценок широки, а различия между логистической регрессией, случайным лесом и методом ближайших соседей статистически не разделяются. Для записей 2019 г. химические показатели известны лишь частично, что снижает общий результат относительно подвыборки 2018 г. Модель не переносится между годами отбора, и для нового обследования её следует переобучать по мере поступления прямых анализов. В датасете отсутствуют метеорологические данные (температура, осадки, ветер), которые влияют на миграцию и трансформацию загрязнителей, их учёт требует иных наблюдений. Синтетическое дополнение выборки (SMOTE) при 19 положительных записях не применялось, поскольку интерполяция между немногими примерами не добавляет информации, дисбаланс учтён весами классов.

Заключение.

В работе проверено, можно ли предсказать загрязнение почв НДМГ и NDMA на месте падения ракеты-носителя по косвенным показателям, не используя сами эти вещества в качестве признаков. На 63 записях по 53 пробам с фактическими измерениями логистическая регрессия и случайный лес ранжируют пробы по риску загрязнения с ROC-AUC 0,75–0,76 при повторной кросс-валидации и 0,68–0,72 при пространственной проверке, тогда как правило по расстоянию даёт 0,59. На пробах с полностью измеренной химией качество достигает 0,85–0,87. Гипотеза о наличии сигнала в косвенных показателях подтверждена, наиболее информативны содержание нитратов и глубина отбора. Кластеризация без учителя загрязнение не воспроизводит, а модель не переносится между годами отбора.

Практическая значимость состоит в возможности отбирать пробы для дорогостоящего определения НДМГ и NDMA по результатам общего анализа и положению пробы, что сокращает объём прямых анализов при контролируемых потерях полноты, и в схеме зонирования по расстоянию и глубине, которая может встраиваться в геоинформационные системы экологического мониторинга районов падения. Дальнейшая работа связана с расширением выборки за счёт новых обследований с полным набором показателей и проверкой переносимости модели между местами падения.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

REFERENCES

Bekeshev Y., Semenkov I., Stepanova Y., Karpachevskiy A., Lednev S., Klink G., Yerzhanov Y., Bapyshev A. & Koroleva T. (2024). Data on the temporal changes in soil properties at the emergency crash site of the launch vehicle 'Soyuz-FG' in Kazakhstan // Data in Brief. Article 110646. DOI: https://doi.org/10.1016/j.dib.2024.110646 [In Eng].

Bekeshev Y., Zhumabekova Zh. & Moldabekov M. (2024). Detoxification of contaminated soils from hydrocarbon rocket fuel used in «Soyuz» launch vehicles using the bioremediation method // Journal of Space Safety Engineering. — Pp. 605–613. DOI: https://doi.org/10.1016/j.jsse.2024.07.004 [In Eng].

Breiman L. (2001). Random Forests // Machine Learning. — Pp. 5–32. DOI: https://doi.org/10.1023/A:1010933404324 [In Eng].

Chicco D., Campagner A., Spagnolo A., Ciucci D. & Jurman G. (2025). The Silhouette coefficient and the Davies-Bouldin index are more informative than Dunn index, Calinski-Harabasz index, Shannon entropy, and Gap statistic for unsupervised clustering internal evaluation of two convex clusters // PeerJ Computer Science. DOI: https://doi.org/10.7717/peerj-cs.3309 [In Eng].

Cover T. & Hart P. (1967). Nearest neighbor pattern classification // IEEE Transactions on Information Theory. — Pp. 21–27. DOI: https://doi.org/10.1109/TIT.1967.1053964 [In Eng].

Friedman J.H. (2001). Greedy function approximation: a gradient boosting machine // The Annals of Statistics. — Pp. 1189–1232. DOI: https://doi.org/10.1214/aos/1013203451 [In Eng].

Hastie T., Tibshirani R. & Friedman J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction // 2nd ed. - New York: Springer. — Pp. 745. DOI: https://doi.org/10.1007/978-0-387-84858-7 [In Eng].

Koroleva T.V., Semenkov I.N., Sharapova A.V., Krechetov P.P. & Lednev S.A. (2021). Ecological consequences of space rocket accidents in Kazakhstan between 1999 and 2018 // Environmental Pollution. Article 115711. DOI: https://doi.org/10.1016/j.envpol.2020.115711 [In Eng].

Koroleva T.V., Semenkov I.N., Lednev S.A. & Soldatova O.S. (2023). Nesimmetrichnyy dimetilgidrazin (NDMG) i produkty yego transformatsii v pochvakh: istochniki postupleniya, diagnostika, povedeniye, toksichnost' i remediatsiya zagryaznennykh territoriy (obzor) [Unsymmetrical dimethylhydrazine (UDMH) and its transformation products in soils: sources, diagnostics, behavior, toxicity and remediation of contaminated areas (review)] // Pochvovedeniye. — Pp. 240–258. DOI: https://doi.org/10.1134/S1064229322602001 [In Russ.].

Kuznetsova M.N. & Melikhova A.V. (2024). Ispol'zovaniye GIS-tekhnologiy pri ekologicheskoy otsenke territoriy [Use of GIS technologies in environmental assessment of territories] // Mezhdunarodnyy zhurnal gumanitarnykh i yestestvennykh nauk. — Pp. 44–47. [In Russ.].

MacQueen J. (1967). Some methods for classification and analysis of multivariate observations // Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. — Pp. 281–297. [In Eng].

Milyushkin A.L. & Karnaeva A.E. (2023). Unsymmetrical dimethylhydrazine transformation products: A review // Science of The Total Environment. Article 164367. DOI: https://doi.org/10.1016/j.scitotenv.2023.164367 [In Eng].

Pavlova A.V., Gladskoy I.B. & Telyatnikov I.S. (2024). Tsifrovaya model' mestnosti regiona v primenenii k problemam ekologicheskogo monitoringa [Digital terrain model of the region in application to environmental monitoring problems] // Izvestiya Tul'skogo gosudarstvennogo universiteta. Nauki o zemle. — Pp. 68–77. [In Russ].

Pedregosa F., Varoquaux G., Gramfort A. et al. (2011). Scikit-learn: Machine learning in Python // Journal of Machine Learning Research. — Pp. 2825–2830. [In Eng].

Roberts D.R., Bahn V., Ciuti S., Boyce M.S., Elith J., Guillera-Arroita G., Hauenstein S., Lahoz-Monfort J.J., Schröder B., Thuiller W., Warton D.I., Wintle B.A., Hartig F. & Dormann C.F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure // Ecography. — Pp. 913–929. DOI: https://doi.org/10.1111/ecog.02881 [In Eng].

Rousseeuw P.J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis // Journal of Computational and Applied Mathematics. — Pp. 53–65. DOI: https://doi.org/10.1016/0377-0427(87)90125-7 [In Eng].

Saito T. & Rehmsmeier M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets // PLoS ONE. - e0118432. DOI: https://doi.org/10.1371/journal.pone.0118432 [In Eng].

Segal A.M., Arustamov E.A. & Balakin V.A. (2023). Otsenka ekologicheskogo sostoyaniya territorii i monitoring okruzhayushchey sredy s ispol'zovaniyem GIS-tekhnologiy [Assessment of the environmental state of the territory and environmental monitoring using GIS technologies] // Geodeziya i kartografiya. — Pp. 45–52. [In

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Russ.].

Semenkov I., Bekeshev Y., Stepanova Y. & Koroleva T. (2024). Results of chemical analysis of soil properties at the emergency crash site of the launch vehicle 'Soyuz-FG' in Kazakhstan // Mendeley Data- V1. DOI: https://doi.org/10.17632/47twnk5dsh.1 [In Eng].

Semenkov I.N. & Koroleva T.V. (2022). Review on the environmental impact of emissions from space launches: a case study for areas affected by the Russian space programme // Environmental Science and Pollution Research. — Pp. 89807–89822. DOI: https://doi.org/10.1007/s11356-022-23888-8 [In Eng].

Sharapova A.V., Semenkov I.N., Koroleva T.V., Krechetov P.P., Lednev S.A. & Smolenkov A.D. (2020). Snow pollution by nitrogen-containing substances as a consequence of rocket launches from the Baikonur Cosmodrome // Science of the Total Environment. Article 136072. DOI: https://doi.org/10.1016/j.scitotenv.2019.136072 [In Eng].

Zyablitskaya A.N., Novikova I.I. & Shchuchinov L.V. (2023). Rezul'taty provedeniya ekologo-gigiyenicheskogo monitoringa v rayonakh padeniya otdelyayushchikhsya chastey raket-nositeley [Results of ecological and hygienic monitoring in the areas of falling rocket booster parts] // Meditsina truda i ekologiya cheloveka. — Pp. 82–94. [In Russ.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 495–511
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.027
IRSTI 20.15.05; 28.23.15

MACHINE LEARNING-BASED CYBERATTACK DETECTION IN OT-NETWORKS: A COMPARATIVE ANALYSIS OF NETWORK ANOMALY CLASSIFICATION METHODS

S.N. Issabayeva[1]*, A.K. Rysbayeva[2], Zh.B. Ibrayeva[3], N.R. Mussabekov[4], N.Sh. Akimzhan[5]

[1]Temirbek Zhurgenov Kazakh National Academy of Arts, Almaty, Kazakhstan;
[2]LLP «International Educational Corporation», Kazakh Leading Academy of Architecture and Civil Engineering, Almaty, Kazakhstan;
[3]SDU University, Almaty, Kazakhstan;
[4]Satbayev University, Almaty, Kazakhstan;
[5]Abai Kazakh National Pedagogical University, Almaty, Kazakhstan.
E-mail: suluakma@gmail.com

Sulu N. Issabayeva — candidate of pedagogical sciences, associate professor of the Department of History of Kazakhstan and Social Sciences, Temirbek Zhurgenov Kazakh National Academy of Arts, Almaty, Kazakhstan
E-mail: suluakma@gmail.com, https://orcid.org/0000-0002-5071-0110;
Aiman K. Rysbayeva — PhD, associate professor, Researcher and Lecturer of the Department of Technical and Natural Sciences, LLP «International Educational Corporation», Kazakh Leading Academy of Architecture and Civil Engineering, Almaty, Kazakhstan
https://orcid.org/0000-0001-8135-4596;
Zhanar B. Ibrayeva — PhD, associate professor of the Department of Natural Sciences Pedagogy, SDU university, Almaty, Kazakhstan
https://orcid.org/0000-0003-3196-696X;
Nazarbek R. Mussabekov — PhD, associate professor of the Department of Automation and Control, Satbayev University, Almaty, Kazakhstan
https://orcid.org/0000-0003-2334-4730;
Nagima Sh. Akimzhan — PhD, Senior lecturer of the Department of informatics and informatization of education, Abai Kazakh National Pedagogical University, Almaty, Kazakhstan
https://orcid.org/0009-0002-9939-8145.

© S.N. Issabayeva, A.K. Rysbayeva, Zh.B. Ibrayeva, N.R. Mussabekov, N. Sh. Akimzhan

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Abstract. The growing number of cyberattacks targeting industrial networks increases the need for timely detection of anomalous activity in operational technology (OT) environments. This study aims to comparatively evaluate machine learning methods for cyberattack detection and network anomaly classification in OT networks. The experimental methodology includes network traffic preprocessing, informative feature space construction, and training of Random Forest, XGBoost, Support Vector Machine, and multilayer neural network models. Algorithm performance is evaluated using Accuracy, Precision, Recall, F1-score, and ROC-AUC, as well as their ability to identify different categories of attacks. XGBoost demonstrated the best performance, achieving an Accuracy of 98.4%, an F1-score of 98.1%, and a ROC-AUC of 0.994. The comparative analysis indicates that ensemble methods can improve network anomaly detection accuracy while maintaining a low false-positive rate. The obtained results can be applied to the development of intelligent intrusion detection systems for industrial OT infrastructures.

Keywords: OT-networks, cyberattacks, machine learning, anomaly detection, network traffic, classification, information security

For citation: Issabayeva S.N., Rysbayeva A.K., Ibrayeva Zh.B., Mussabekov N.R., Akimzhan N.Sh. (2026). Machine learning-based cyberattack detection in ot-networks: a comparative analysis of network anomaly classification methods // International journal of information and communication technologies. Vol. 7. No. 27. Pp. 495–511. https://doi.org/10.54309/IJICT.2026.27.3.027. (In Eng.).

Conflict of interest: The authors declare that there is no conflict of interest.

МАШИНАЛЫҚ ОҚЫТУ НЕГІЗІНДЕ OT-ЖЕЛІЛЕРІНДЕГІ КИБЕРШАБУЫЛДАРДЫ АНЫҚТАУ: ЖЕЛІЛІК АНОМАЛИЯЛАРДЫ ЖІКТЕУ ӘДІСТЕРІНІҢ САЛЫСТЫРМАЛЫ ТАЛДАУЫ

С.Н. Исабаева[1]*, А.К. Рысбаева[2], Ж.Б. Ибраева[3], Н.Р. Муссабеков[4], Н.Ш. Акимжан[5]

[1]Темірбек Жүргенов атындағы Қазақ ұлттық өнер академиясы, Алматы, Қазақстан;

[2]«Халықаралық білім беру корпорациясы» ЖШС, Қазақ бас сәулет-құрылыс академиясы, Алматы, Қазақстан;

[3]SDU University, Алматы, Қазақстан;

[4]Қ.И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті, Алматы, Қазақстан;

[5]Абай атындағы Қазақ ұлттық педагогикалық университеті, Алматы, Қазақстан.

E-mail: suluakma@gmail.com

Исабаева Сулу Нагашыбаевна — педагогика ғылымдарының кандидаты, "Қазақстан тарихы және әлеуметтік ғылымдар" кафедрасының қауымдастырылған профессоры, Темірбек Жүргенов атындағы Қазақ ұлттық өнер академиясы, Алматы, Қазақстан

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

E-mail: suluakma@gmail.com, https://orcid.org/0000-0002-5071-0110;
Рысбаева Айман Калиевна — PhD, "Техникалық және жаратылыстану ғылымдары" кафедрасының қауымдастырылған профессоры, «Халықаралық білім беру корпорациясы» ЖШС, Қазақ бас сәулет-құрылыс академиясы, Алматы, Қазақстан
https://orcid.org/0000-0001-8135-4596;
Ибраева Жанар Базарбековна — PhD, "Жаратылыстану ғылымдарының педагогикасы" кафедрасының қауымдастырылған профессоры, SDU university, Алматы, Қазақстан
https://orcid.org/0000-0003-3196-696X;
Муссабеков Назарбек Расулбекович — PhD, "Автоматтандыру және басқару" кафедрасының қауымдастырылған профессоры, Қ.И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті, Алматы, Қазақстан
https://orcid.org/0000-0003-2334-4730;
Акимжан Нагима Шопанкызы — PhD, "Информатика және білімді ақпараттандыру" кафедрасының аға оқытушысы, Абай атындағы Қазақ ұлттық педагогикалық университеті, Алматы, Қазақстан
https://orcid.org/0009-0002-9939-8145.

© С.Н. Исабаева, А.К. Рысбаева, Ж.Б. Ибраева, Н.Р. Муссабеков, Н.Ш. Акимжан

Аннотация. Өнеркәсіптік желілерге бағытталған кибершабуылдар санының артуы операциялық технологиялар (ОТ) ортасындағы аномальды белсенділікті уақтылы анықтауға қойылатын талаптарды күшейтуде. Зерттеудің мақсаты – ОТ желілеріндегі кибершабуылдарды анықтау және желілік аномалияларды жіктеу үшін машиналық оқыту әдістерінің тиімділігін салыстырмалы бағалау. Эксперименттік әдістеме желілік трафикті алдын ала өңдеуді, ақпараттық белгілер кеңістігін қалыптастыруды және Random Forest, XGBoost, Support Vector Machine және көпқабатты нейрондық желі модельдерін оқытуды қамтиды. Алгоритмдердің тиімділігі Accuracy, Precision, Recall, F1-score және ROC-AUC көрсеткіштері, сондай-ақ шабуылдардың әртүрлі санаттарын анықтау қабілеті бойынша бағаланды. Ең жоғары нәтижені XGBoost көрсетті: Accuracy – 98,4%, F1-score – 98,1%, ROC-AUC – 0,994. Салыстырмалы талдау ансамбльдік әдістерді қолдану жалған іске қосылулардың төмен деңгейін сақтай отырып, желілік аномалияларды анықтау дәлдігін арттыруға мүмкіндік беретінін көрсетті. Алынған нәтижелер өнеркәсіптік ОТ-инфрақұрылымдарға арналған интеллектуалды басып кіруді анықтау жүйелерін әзірлеуде қолданылуы мүмкін.

Түйінді сөздер: ОТ-желілері, кибершабуылдар, машиналық оқыту, аномалияларды анықтау, желілік трафик, жіктеу, ақпараттық қауіпсіздік

Дәйексөздер үшін: Исабаева С.Н., Рысбаева А.К., Ибраева Ж.Б., Муссабеков Н.Р., Акимжан Н.Ш. (2026). Машиналық оқыту негізінде от-желілеріндегі кибершабуылдарды анықтау: желілік аномалияларды жіктеу әдістерінің салыстырмалы талдауы // Халықаралық ақпараттық және коммуни-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

калық технологиялар журналы. Т. 7. No 27. Б. 495–511. https://doi.org/10.54309/IJICT.2026.27.3.027. (Ағыл. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

ОБНАРУЖЕНИЕ КИБЕРАТАК В ОТ-СЕТЯХ НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ: СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ КЛАССИФИКАЦИИ СЕТЕВЫХ АНОМАЛИЙ

С.Н. Исабаева[1]*, А.К. Рысбаева[2], Ж.Б. Ибраева[3], Н.Р. Муссабеков[4], Н.Ш. Акимжан[5]

[1]Казахская национальная академия искусств имени Темирбека Жүргенова, Алматы, Казахстан;

[2]ТОО «Международная образовательная корпорация», Казахская головная архитектурно-строительная академия, Алматы, Казахстан;

[3]SDU University, Алматы, Казахстан;

[4]Казахский национальный исследовательский технический университет имени К.И. Сатпаева, Алматы, Казахстан;

[5]Казахский национальный педагогический университет имени Абая, Алматы, Казахстан.

E-mail: suluakma@gmail.com

Исабаева Сулу Нагашыбаевна — кандидат педагогических, ассоциированный профессор кафедры история Казахстана и социальных наук, Казахская национальная академия искусств имени Темирбека Жүргенова, Алматы, Казахстан
E-mail: suluakma@gmail.com, https://orcid.org/0000-0002-5071-0110;

Рысбаева Айман Калиевна — PhD, ассоциированный профессор кафедры технических и естественных наук, ТОО «Международная образовательная корпорация», Казахская головная архитектурно-строительная академия, Алматы, Казахстан
https://orcid.org/0000-0001-8135-4596;

Ибраева Жанар Базарбековна — PhD, ассоциированный профессор кафедры педагогики естественных наук, SDU university, Алматы, Казахстан
https://orcid.org/0000-0003-3196-696X;

Муссабеков Назарбек Расулбекович — PhD, ассоциированный профессор кафедры автоматизации и управления, Казахский национальный исследовательский технический университет имени К.И. Сатпаева, Алматы, Казахстан
https://orcid.org/0000-0003-2334-4730;

Акимжан Нагима Шопанкызы — PhD, старший преподаватель кафедры информатики и информатизации образования, Казахский национальный педагогический университет имени Абая, Алматы, Казахстан
https://orcid.org/0009-0002-9939-8145.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

© С.Н. Исабаева, А.К. Рысбаева, Ж.Б. Ибраева, Н.Р. Муссабеков, Н.Ш. Акимжан

Аннотация. Рост числа кибератак на промышленные сети повышает требования к своевременному выявлению аномальной активности в операционных технологических (OT) средах. Цель исследования заключается в сравнительной оценке методов машинного обучения для обнаружения кибератак и классификации сетевых аномалий в OT-сетях. Экспериментальная методика предусматривает предварительную обработку сетевого трафика, формирование информативного пространства признаков и обучение моделей Random Forest, XGBoost, Support Vector Machine и многослойной нейронной сети. Эффективность алгоритмов оценивается по показателям Accuracy, Precision, Recall, F1-score и ROC-AUC, а также по способности выявлять различные категории атак. Наилучшие результаты получены с использованием XGBoost: Accuracy составила 98,4%, F1-score — 98,1%, ROC-AUC — 0,994. Сравнительный анализ показал, что применение ансамблевых методов позволяет повысить точность обнаружения сетевых аномалий при сохранении низкой доли ошибочных срабатываний. Полученные результаты могут быть использованы при разработке интеллектуальных систем обнаружения вторжений для промышленных OT-инфраструктур.

Ключевые слова: OT-сети, кибератаки, машинное обучение, обнаружение аномалий, сетевой трафик, классификация, информационная безопасность

Для цитирования: Исабаева С.Н., Рысбаева А.К., Ибраева Ж.Б., Муссабеков Н.Р., Акимжан Н.Ш. (2026). Обнаружение кибератак в ot-сетях на основе машинного обучения: сравнительный анализ методов классификации сетевых аномалий // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 27. Стр. 495–511. https://doi.org/10.54309/IJICT.2026.27.3.027. (На англ.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Introduction.

The digital transformation of industry is accompanied by the progressive integration of Operational Technology (OT) with corporate information systems, the Industrial Internet of Things (IIoT), and remote-control technologies. This convergence expands the functional capabilities of automated industrial systems while simultaneously creating additional potential attack vectors targeting technological infrastructure. Unlike conventional information systems, the consequences of cybersecurity incidents in OT environments are not limited to the loss of data confidentiality or availability: a successful cyberattack may affect physical process parameters, equipment stability, production continuity, and the safety of technological operations. The specific characteristics of OT infrastructure, including long equipment life cycles, the use of industrial communication protocols, real-time operational requirements, and limited opportunities to interrupt systems for software updates, necessitate specialized mechanisms for cyberattack detection (Zolanvari et al., 2019; Koay et al., 2023; Pinto

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

et al., 2023; Aslam et al., 2025).

Traditional intrusion detection systems predominantly rely on signature-based analysis or predefined rules for identifying suspicious network activity. Such approaches remain effective in detecting known attack scenarios; however, their capabilities are significantly limited when malicious traffic patterns change, previously unseen sequences of actions emerge, or attackers exploit legitimate network operations to achieve unauthorized objectives. This limitation is particularly relevant to OT networks, where malicious activity may manifest not only through explicit disruptions of network communication but also through relatively subtle deviations in request frequency, command sequences, connection characteristics, and the behavior of individual network nodes (Buczak et al., 2016; Ferrag et al., 2020).

Machine learning represents a promising direction for advancing intrusion detection systems, as it enables the construction of models of normal and anomalous network behavior based on multiple traffic features. Classification algorithms can capture nonlinear relationships among connection characteristics and identify combinations of features that are difficult to describe using a fixed set of expert-defined rules. Recent research in industrial cybersecurity demonstrates the active application of Random Forest, Support Vector Machine, gradient boosting methods, and neural network architectures for network traffic analysis and attack identification in ICS/IIoT environments. At the same time, the effectiveness of a particular algorithm is strongly influenced by the structure of the input data, class distribution, feature engineering strategy, and the selected training and testing procedures (Awotunde et al., 2021; Altunay et al., 2023; Yu et al., 2024).

A high Accuracy value, when considered in isolation, does not necessarily provide sufficient evidence of the practical effectiveness of an intrusion detection model. In the presence of substantial class imbalance, a classifier may demonstrate high overall accuracy while failing to detect a considerable proportion of attacks belonging to specific categories. For OT infrastructure protection systems, particular importance should therefore be given to detection recall, F1-score, false-positive rate, and the ability of a model to distinguish between individual categories of malicious activity. Computational complexity is another important consideration, since attack detection should be performed with minimal latency and without imposing substantial additional load on industrial information infrastructure (Le et al., 2022).

An analysis of recent studies indicates a shift from the conventional binary formulation of the “normal traffic–attack” problem toward multiclass recognition of different types of malicious activity. At the same time, increasing attention is being paid to the quality of the feature space and the robustness of classifiers when processing heterogeneous data. In particular, studies employing the X-IIoTID, TON_IoT, and WUSTL-IIOT datasets demonstrate the potential of machine learning techniques for industrial and IIoT traffic analysis, while also indicating that classification performance depends substantially on the characteristics of the dataset and the selected classification model. Recent research additionally considers informative feature selection as an effective means of reducing the computational complexity of intrusion detection

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

systems without substantially compromising classification performance (Moustafa, 2021; Ferrag et al., 2022; Ismail et al., 2025).

Despite the considerable body of research in this field, direct comparison of machine learning algorithms under uniform experimental conditions remains an important research problem. Results reported in different studies are often obtained using different datasets, preprocessing procedures, class-balancing techniques, and evaluation metrics, which makes direct comparison of classifier performance difficult. Furthermore, high accuracy in binary attack detection does not necessarily guarantee comparable performance in a multiclass setting, where the system must identify a specific category of anomalous network activity. Consequently, there is a need for an experimental evaluation of different algorithms within a unified data-processing pipeline and under consistent performance criteria.

The aim of this study is to comparatively evaluate machine learning methods for cyberattack detection and network anomaly classification in OT networks. To achieve this aim, the study performs network traffic preprocessing and feature-space construction, trains Random Forest, XGBoost, Support Vector Machine, and multilayer neural network models, and compares their performance using Accuracy, Precision, Recall, F1-score, and ROC-AUC. Particular attention is paid to the ability of the models to correctly identify different attack categories while minimizing false-positive detections.

The scientific and practical significance of this study lies in establishing a unified framework for the comparative evaluation of network anomaly classification algorithms in OT environments. Rather than assessing models solely on the basis of overall accuracy, the proposed approach provides a comprehensive analysis of classification performance using Precision, Recall, F1-score, and ROC-AUC. Such a comparison makes it possible to identify the specific characteristics of the investigated methods when applied to intelligent intrusion detection systems and provides a methodological basis for selecting an appropriate classifier with regard to detection performance and robustness against false positives in industrial network infrastructures.

Materials and methods.

Research design

The study was designed as a computational experiment aimed at comparatively evaluating machine learning methods for detecting and classifying anomalous network activity in operational technology (OT) environments. The experimental workflow comprised dataset preparation, data preprocessing, categorical feature transformation, numerical feature standardization, partitioning into training, validation, and test subsets, classifier training, and subsequent evaluation using a common set of performance metrics. This design ensured that all models were compared under consistent data-processing and evaluation conditions.

Dataset

The open X-IIoTID (X-IIoT Intrusion Dataset) was used as the experimental data source. The dataset was developed for intrusion-detection research in heterogeneous Industrial Internet of Things environments and contains information describing network

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

connections, device activity, system events, and interactions among components of industrial infrastructure. Compared with datasets primarily representing conventional enterprise IP traffic, X-IIoTID incorporates the behavior of sensors, actuators, controllers, edge devices, and cloud-connected components, making it suitable for cybersecurity analysis in OT/IIoT environments (Al-Hawawreh et al., 2022).

The original X-IIoTID dataset contains 820,834 observations and 68 attributes and provides multiple levels of labeling for network activity. Of all records, 421,417 represent normal behavior and 399,417 correspond to malicious activity. Accordingly, attack observations account for approximately 48.66% of the dataset, which limits severe class imbalance in the binary classification setting (Al-Hawawreh et al., 2022).

The attack scenarios represented in X-IIoTID cover different stages and mechanisms of cyber intrusion, including reconnaissance, credential-related attacks, vulnerability exploitation, Man-in-the-Middle activity, unauthorized Modbus register reading, manipulation of MQTT communications, false-data injection, data exfiltration, ransomware, and attacks intended to disrupt service availability. This structure enables the problem to be considered not only as binary intrusion detection but also as multiclass classification of different forms of anomalous activity (Al-Hawawreh et al., 2022).

Two classification settings were considered in the study. In the first stage, binary classification was performed by assigning each observation to either the Normal or Attack class. In the second stage, multiclass classification was conducted while preserving attack categories. This two-stage design makes it possible to separately assess a model's ability to detect deviations from normal network behavior and its ability to correctly identify the type of detected threat.

Data preprocessing

Before model training, the source data were processed through a unified preprocessing pipeline. The initial stage included verification of the dataset structure, feature data types, missing values, duplicate records, and features with constant or near-constant values. Identifier fields and variables capable of directly revealing the target label, thereby causing information leakage between input variables and classification outcomes, were excluded from the feature space.

Categorical network parameters were converted into numerical representations. Nominal features were encoded without imposing an artificial ordinal relationship among categories. Numerical variables were standardized to a comparable scale. For feature x_j, the transformation was defined as

$$x'ij = (xij - \mu j) / \sigma j, \quad (1)$$

where xij is the original value of the j-th feature for the i-th observation, μj is the mean of the feature calculated from the training subset, σj is its standard deviation, and x'ij is the standardized value.

Transformation parameters were estimated exclusively from the training subset and subsequently applied unchanged to the validation and test subsets. This procedure prevented statistical information from the test data from influencing model training.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

To reduce the influence of redundant variables, features with zero or near-zero variance were examined together with relationships among numerical variables. When strongly correlated features were identified, the exclusion of one of the redundant variables was considered. Feature informativeness was additionally assessed using only the training data. Feature selection was performed after the initial dataset partitioning to prevent test-set information from influencing the construction of the feature space.

Following preprocessing, the data were divided into training, validation, and test subsets in a 70:15:15 ratio. The training subset was used to estimate model parameters, the validation subset was used for hyperparameter tuning and configuration selection, and the test subset was reserved exclusively for final independent comparison. Class proportions were preserved during partitioning.

Machine learning models

The comparative experiment included four algorithms representing different classification paradigms: Random Forest (RF), XGBoost, Support Vector Machine (SVM), and a multilayer neural network implemented as a Multilayer Perceptron (MLP). These models were selected to compare tree-based ensemble methods, a maximum-margin classifier, and a neural-network approach under identical experimental conditions.

Random Forest constructed an ensemble of decision trees using bootstrap samples of the training data and randomly selected subsets of features. The final class was determined by aggregating the predictions of individual trees. RF was included because it can model nonlinear relationships and interactions among features without requiring their functional form to be specified in advance (Breiman, 2001).

XGBoost employed sequential gradient-boosted decision trees, where each subsequent tree was trained to reduce the residual errors of the existing ensemble. The tuned hyperparameters included the number of trees, maximum tree depth, learning rate, feature subsampling, and regularization parameters. XGBoost was selected because of its ability to process structured high-dimensional data efficiently and capture complex dependencies among network traffic characteristics (Chen and Guestrin, 2016).

For SVM, a nonlinear radial basis function (RBF) kernel was used. Model optimization included selection of the regularization parameter C and kernel coefficient γ. Numerical variables were standardized before SVM training because the position of the separating hyperplane and the distances between observations are sensitive to differences in feature scales (Cortes et al., 1995).

The MLP model consisted of an input layer, two fully connected hidden layers with ReLU activation functions, and an output layer. For binary classification, the output layer estimated the probability that an observation belonged to the attack class. For multiclass classification, it produced a probability distribution across the target categories. Regularization and early stopping based on validation loss were used to reduce the risk of overfitting.

Hyperparameter tuning for all models was performed using only the training and validation subsets. The configuration achieving the highest validation F1-score was fixed and subsequently applied to the test subset without further parameter adjustment.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Evaluation metrics

Model performance was compared using a common set of metrics comprising Accuracy, Precision, Recall, F1-score, and ROC-AUC. In the binary classification setting, malicious network activity was treated as the positive class. The principal metrics were calculated as follows:

$$\begin{aligned} &\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN), \\ &\text{Precision} = TP / (TP + FP), \\ &\text{Recall} = TP / (TP + FN), \\ &F1 = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}), \end{aligned} \quad (2)$$

where TP denotes correctly detected attacks, TN denotes correctly classified normal observations, FP denotes normal observations incorrectly classified as attacks, and FN denotes attacks incorrectly classified as normal activity.

ROC-AUC was used as an aggregate measure of a model's ability to discriminate between normal and malicious traffic across different decision thresholds. Because missed attacks and false alarms have different operational implications in intrusion-detection systems, the False Positive Rate (FPR) and False Negative Rate (FNR) were additionally analyzed:

$$\begin{aligned} &FPR = FP / (FP + TN), \\ &FNR = FN / (FN + TP). \end{aligned} \quad (3)$$

For the multiclass setting, Precision, Recall, and F1-score were calculated for each category, while macro-averaging was used for overall model comparison. Macro-F1 was defined as

$$F1_{macro} = (1 / K) \Sigma_{(k=1 \text{ to } K)} F1_k, \quad (4)$$

where K is the number of classes and $F1_k$ is the F1-score for the k-th class. Macro-F1 gives equal importance to the recognition quality of each attack category regardless of its frequency in the dataset.

The final assessment was therefore based on a combination of performance characteristics rather than on maximizing Accuracy alone. F1-score and Recall for the attack class were given particular attention because failure to identify malicious activity represents a substantial risk in practical OT environments. At the same time, FPR was monitored as an indicator of the potential operational burden created by false alerts in an intrusion-detection system.

Overall, the experimental framework provides consistent preprocessing, training, and testing conditions for the four classification models. The use of an independent test subset and several complementary metrics enables the algorithms to be compared not only in terms of overall classification accuracy but also with respect to attack detection, differentiation among network anomaly categories, and control of

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

false-positive detections.

Results and discussion.

The experimental evaluation was designed to determine the effectiveness of Random Forest (RF), XGBoost, Support Vector Machine (SVM), and Multilayer Perceptron (MLP) for detecting malicious network activity in an OT/IIoT environment. The models were evaluated sequentially under binary and multiclass settings. The first stage assessed their ability to distinguish normal traffic from attack traffic, whereas the second stage examined their capacity to differentiate among individual cyberattack categories.

Binary classification results

The test-set results for binary attack detection are presented in Table 1. All investigated algorithms achieved Accuracy values above 95%; however, differences among the models became more evident when Precision, Recall, F1-score, and ROC-AUC were considered jointly.

Table 1. Binary network traffic classification results

Model	Accuracy, %	Precision, %	Recall, %	F1-score, %	ROC-AUC
Random Forest	97.6	97.5	97.3	97.4	0.991
XGBoost	98.4	98.3	97.9	98.1	0.994
SVM	95.8	95.5	95.1	95.3	0.982
MLP	97.1	96.9	96.6	96.7	0.989

XGBoost achieved the strongest overall performance. Its Accuracy reached 98.4%, Precision 98.3%, Recall 97.9%, F1-score 98.1%, and ROC-AUC 0.994. The combination of these values indicates a strong ability to discriminate between normal and malicious activity without a pronounced bias toward either class.

Random Forest produced the second-best results for most metrics, with an Accuracy of 97.6% and an F1-score of 97.4%. The difference between RF and XGBoost was 0.8 percentage points in Accuracy and 0.7 percentage points in F1-score. The advantage of XGBoost can be associated with its sequential boosting mechanism, in which subsequent trees are constructed to reduce the errors of the current ensemble. For multidimensional structured network-traffic data, this mechanism can capture complex interactions among features more effectively.

MLP achieved an Accuracy of 97.1% and an F1-score of 96.7%. Although these values indicate a high overall classification quality, the neural model remained below the two tree-based ensemble methods. The result suggests that greater model complexity does not automatically provide superior intrusion-detection performance; for the tabular X-IIoTID feature space, the investigated tree ensembles were more effective than the selected fully connected neural-network configuration.

SVM yielded the lowest values among the four models, with an Accuracy of 95.8% and an F1-score of 95.3%. Nevertheless, its ROC-AUC of 0.982 indicates substantial discriminatory capability. The lower classification performance relative to the ensemble methods may reflect the difficulty of constructing a single nonlinear decision boundary for heterogeneous, high-dimensional network-traffic features (Table

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

2).

Table 2. Cyberattack detection error rates

Model	FPR, %	FNR, %
Random Forest	1.8	2.7
XGBoost	1.1	2.1
SVM	3.4	4.9
MLP	2.2	3.4

For OT infrastructure, overall classification accuracy is not sufficient by itself; missed attacks and false alarms are operationally important. XGBoost produced the lowest False Positive Rate (FPR), 1.1%, and the lowest False Negative Rate (FNR), 2.1%. A low FPR is particularly relevant in continuous monitoring because excessive false alerts can increase the workload of security personnel and reduce the operational value of an intrusion detection system. At the same time, a low FNR reduces the proportion of malicious events incorrectly treated as normal activity.

The relationship between Accuracy and error rates demonstrates the need for multidimensional model evaluation. For example, SVM achieved an Accuracy of 95.8%, which is high in absolute terms, but its FNR reached 4.9%. In an industrial environment, this distinction is important because an undetected malicious action may affect not only information assets but also the continuity and safety of technological processes.

Multiclass classification results

The multiclass setting is more demanding because the classifier must not only detect anomalous behavior but also assign it to the appropriate attack category. To reduce the dominant influence of highly represented classes, macro-F1 was used as the principal aggregate measure (Table 3).

Table 3. Multiclass classification results

Model	Accuracy, %	Macro-Precision, %	Macro-Recall, %	Macro-F1, %
Random Forest	96.1	93.8	92.7	93.2
XGBoost	97.2	95.1	94.4	94.7
SVM	93.4	90.6	89.2	89.8
MLP	95.5	93.0	91.9	92.4

All models showed lower performance in the multiclass task than in binary attack detection. This behavior is expected because the boundaries among individual forms of malicious activity may be substantially less distinct than the boundary separating the combined attack class from normal traffic.

XGBoost retained the strongest performance, reaching an Accuracy of 97.2% and a macro-F1 of 94.7%. The difference between Accuracy and macro-F1 indicates that some less represented attack categories were more difficult to recognize than dominant classes. Consequently, relying only on Accuracy could lead to an overly optimistic assessment of classifier performance.

Random Forest achieved a macro-F1 of 93.2%, 1.5 percentage points below

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

XGBoost. The corresponding values were 92.4% for MLP and 89.8% for SVM. Thus, the relative ordering of the models remained generally consistent when moving from binary to multiclass classification, although the performance differences became more pronounced (Table 4).

Table 4. XGBoost performance for individual attack categories

Attack category	Precision, %	Recall, %	F1-score, %
Reconnaissance	97.8	97.2	97.5
Weaponisation	95.6	94.8	95.2
Exploitation	94.7	93.9	94.3
Lateral Movement	93.9	92.8	93.3
Command & Control	94.5	93.6	94.0
Data Exfiltration	98.1	97.6	97.8
Tampering	92.8	91.7	92.2
Crypto-ransomware	98.4	98.0	98.2
RDOS	96.9	96.2	96.5

XGBoost recognized Crypto-ransomware, Data Exfiltration, and Reconnaissance most reliably. Their F1-scores were 98.2%, 97.8%, and 97.5%, respectively. Lower values were observed for Tampering (92.2%) and Lateral Movement (93.3%). This variation is important for intrusion-detection assessment because high aggregate Accuracy does not imply equal robustness across all attack types. Some attack scenarios create clearly distinguishable traffic patterns, whereas others partially resemble legitimate operations and are therefore more difficult to separate in feature space.

Comparative analysis of the models

Two consistent observations emerge from the comparative analysis. First, the tree-based ensemble methods outperformed SVM and MLP under the investigated conditions. Second, XGBoost retained its advantage in both binary and multiclass classification, indicating stable performance as the target structure became more complex.

Figure 1 shows that XGBoost consistently achieved the highest values across the four principal classification metrics, while Random Forest remained close to it. The gap between the ensemble methods and SVM was especially visible for Recall and F1-score. The relatively compact range of results nevertheless confirms that all four approaches were capable of learning meaningful attack-related patterns from the processed X-IIoTID feature space.

The observed pattern is consistent with recent IIoT intrusion-detection studies in which tree-based ensemble methods have frequently demonstrated strong performance on structured traffic data. However, absolute metric values reported across studies should not be compared without considering differences in preprocessing, feature selection, class balancing, hyperparameter optimization, and data-splitting procedures. These methodological factors can substantially influence the apparent effectiveness of a classifier (Le et al., 2022; Sezgin et al., 2023; Ahmed et al., 2026).

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

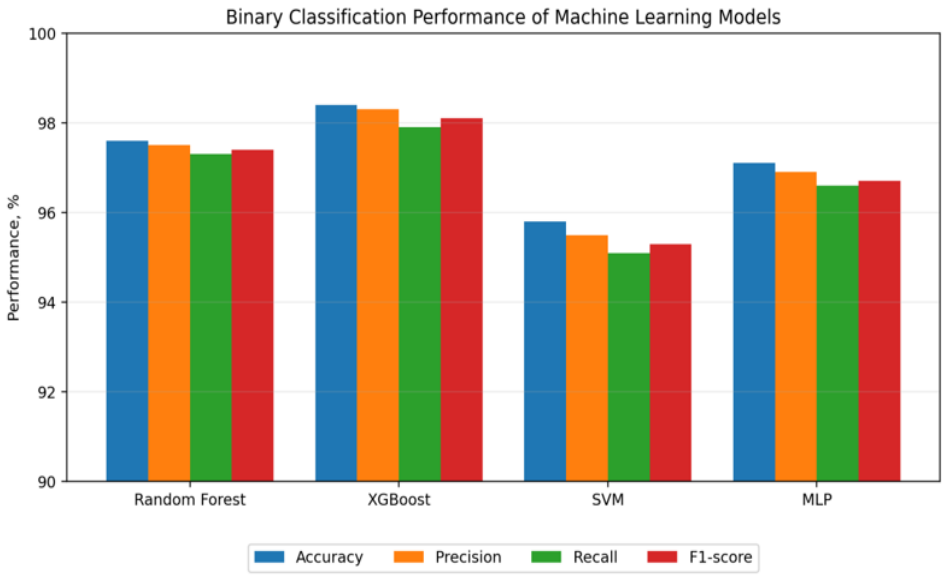


Fig. 1. Comparison of Accuracy, Precision, Recall, and F1-score for the machine learning models in binary attack detection

Studies conducted on X-IIoTID have also reported very high binary classification performance for ensemble methods after automated preprocessing and hyperparameter optimization. At the same time, published results show that performance can vary considerably across individual attack categories. This supports the use of class-sensitive metrics and reinforces the importance of evaluating a model beyond overall Accuracy (Al-Hawawreh et al., 2022; Sezgin et al., 2023).

A further consideration is class imbalance in multiclass intrusion detection. Very high aggregate Accuracy may coexist with noticeably lower macro-F1 when rare attack categories are classified less reliably. The results obtained in the present experiment follow the same general pattern: the multiclass Accuracy of XGBoost was 97.2%, whereas macro-F1 was 94.7%. The difference confirms that macro-averaged metrics provide additional information that would be lost if evaluation were limited to the proportion of correctly classified observations.

Practical interpretation of the results

From an application perspective, the results support the use of XGBoost as a baseline classification component for an intelligent intrusion detection system in an OT/IIoT environment. Its advantage is determined not only by the highest Accuracy but also by the combination of high Recall, high F1-score, a ROC-AUC of 0.994, and the lowest FPR among the investigated models. This balance is particularly relevant for continuous monitoring systems, in which excessive false alarms can overload security personnel, whereas an increased FNR directly raises the probability of missing a genuine incident.

The results should not, however, be interpreted as evidence that XGBoost is universally superior for every industrial infrastructure. X-IIoTID represents a heterogeneous IIoT environment, while real OT networks may differ in equipment

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

composition, industrial protocols, traffic intensity, network topology, and process behavior. Deployment in a production environment therefore requires validation using local traffic and continuous monitoring for changes in the statistical distribution of input features (Zhang et al., 2023).

Another important distinction concerns network intrusion detection versus monitoring of the underlying physical process. Strong classification performance on X-IIoTID demonstrates the model's ability to recognize the malicious patterns represented in the dataset, but it does not by itself establish the ability to detect every manipulation of a physical process at the field-device level. A promising direction for further research is therefore the integration of network-traffic features with equipment telemetry and process variables.

Overall, the experiment demonstrates that machine learning methods can provide effective detection of network anomalies in OT/IIoT environments, while classifier performance becomes more differentiated when moving from binary to multiclass analysis. Under the investigated conditions, XGBoost provided the most balanced combination of Accuracy, Recall, F1-score, ROC-AUC, and low false-positive rate. The multiclass results further indicate that future intrusion-detection systems should focus not only on improving aggregate accuracy but also on increasing recognition performance for less represented and behaviorally similar attack categories.

Conclusion.

The present study provided a comparative evaluation of machine learning methods for cyberattack detection and network anomaly classification in OT/IIoT environments. Using the X-IIoTID dataset, four classification algorithms-Random Forest, XGBoost, Support Vector Machine, and Multilayer Perceptron-were investigated under consistent preprocessing, training, and testing conditions. The use of both binary and multiclass classification settings made it possible to assess the models not only in terms of detecting malicious activity but also with respect to identifying individual categories of cyberattacks.

The binary classification results demonstrated the strongest overall performance for XGBoost. The model achieved an Accuracy of 98.4%, Precision of 98.3%, Recall of 97.9%, F1-score of 98.1%, and ROC-AUC of 0.994. XGBoost also produced the lowest False Positive Rate (1.1%) and False Negative Rate (2.1%) among the evaluated algorithms. This combination of performance indicators is particularly relevant to intrusion detection in OT infrastructures, where reducing false alarms must be balanced against minimizing the probability of undetected malicious activity.

XGBoost also maintained its advantage in the multiclass classification task, achieving an Accuracy of 97.2% and a macro-F1 of 94.7%. At the same time, the recognition performance varied across individual attack categories. The highest F1-scores were obtained for Crypto-ransomware (98.2%), Data Exfiltration (97.8%), and Reconnaissance (97.5%), whereas Tampering and Lateral Movement achieved 92.2% and 93.3%, respectively. These differences confirm that overall classification accuracy should not be considered the sole criterion for evaluating an intelligent intrusion detection system. In OT networks, particular attention should be paid to class-specific

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

performance, including the recognition of less represented and behaviorally similar attack categories.

The comparative analysis further demonstrated that, under the investigated experimental conditions, tree-based ensemble methods outperformed SVM and the evaluated MLP configuration. Nevertheless, these findings should not be interpreted as evidence of the universal superiority of a particular algorithm for every industrial infrastructure. Real-world OT networks differ in equipment composition, industrial communication protocols, network architecture, traffic intensity, and the characteristics of underlying technological processes. Therefore, practical deployment requires validation of the selected model using traffic from the target infrastructure, followed by continuous monitoring for changes in feature distributions.

Thus, the objective of the study was achieved: machine learning methods were compared under a unified experimental framework, and their specific characteristics for detecting and classifying network anomalies in OT/IIoT environments were identified. The obtained results indicate that XGBoost represents a promising foundation for the classification component of an intelligent intrusion detection system when high detection capability and a low false-positive rate are required.

Future research should focus on integrating network traffic characteristics with industrial equipment telemetry and physical process variables, as well as evaluating model robustness against previously unseen attacks. Further investigation of adaptive detection mechanisms capable of responding to changes in normal network behavior over time is also warranted. Another important direction is the development of interpretable classification mechanisms that can provide cybersecurity specialists with transparent explanations of model decisions.

REFERENCES

Ahmed A.A., Abdullah T.A.A. (2026). An Enhanced XGBoost-Based Framework for Efficient Multi-Class Cyber Threat Detection in Industrial IoT Networks // Technologies. — Vol. 14. — No. 5. Article 274. DOI: https://doi.org/10.3390/technologies14050274 (in Eng.).

Al-Hawawreh M., Sitnikova E., Aboutorab N. (2022). X-IIoTID: A Connectivity-Agnostic and Device-Agnostic Intrusion Data Set for Industrial Internet of Things // IEEE Internet of Things Journal. — Vol. 9. — No. 5. Pp. 3962–3977. DOI: https://doi.org/10.1109/JIOT.2021.3102056 (in Eng.).

Altunay H.C., Albayrak Z. (2023). A Hybrid CNN+LSTM-Based Intrusion Detection System for Industrial IoT Networks // Engineering Science and Technology, an International Journal. — Vol. 38. — Article 101322. DOI: https://doi.org/10.1016/j.jestch.2022.101322 (in Eng.).

Aslam M.M., Tufail A., Irshad M.N. (2025). Survey of Deep Learning Approaches for Securing Industrial Control Systems: A Comparative Analysis // Cyber Security and Applications. — Vol. 3. — Article 100096. DOI: https://doi.org/10.1016/j.csa.2025.100096 (in Eng.).

Awotunde J.B., Chakraborty C., Adeniyi A.E. (2021). Intrusion Detection in Industrial Internet of Things Network-Based on Deep Learning Model with Rule-Based Feature Selection // Wireless Communications and Mobile Computing. — Vol. 2021. — Article 7154587. — Pp. 17. DOI: https://doi.org/10.1155/2021/7154587 (in Eng.).

Breiman L. (2001). Random Forests // Machine Learning. — Vol. 45. — No. 1. — Pp. 5–32. DOI: https://doi.org/10.1023/A:1010933404324 (in Eng.).

Buczak A.L., Guven E. (2016). A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection // IEEE Communications Surveys & Tutorials. — Vol. 18. — No. 2. — Pp. 1153–1176. DOI: https://doi.org/10.1109/COMST.2015.2494502 (in Eng.).

Chen T., Guestrin C. (2016). XGBoost: A Scalable Tree Boosting System // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). — New York:

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Association for Computing Machinery. Pp. 785–794. DOI: https://doi.org/10.1145/2939672.2939785 (in Eng.).

Cortes C., Vapnik V. (1995). Support-Vector Networks // Machine Learning. — Vol. 20. — Pp. 273–297. DOI: https://doi.org/10.1007/BF00994018 (in Eng.).

Ferrag M.A., Friha O., Hamouda D., Maglaras L., Janicke H. (2022). Edge-IIoTset: A New Comprehensive Realistic Cyber Security Dataset of IoT and IIoT Applications for Centralized and Federated Learning // IEEE Access. — Vol. 10. — Pp. 40281–40306. DOI: https://doi.org/10.1109/ACCESS.2022.3165809 (in Eng.).

Ferrag M.A., Maglaras L., Moschoyiannis S., Janicke H. (2020). Deep Learning for Cyber Security Intrusion Detection: Approaches, Datasets, and Comparative Study // Journal of Information Security and Applications. — Vol. 50. — Article 102419. DOI: https://doi.org/10.1016/j.jisa.2019.102419 (in Eng.).

Ismail S., Dandan S., Qushou A. (2025). Intrusion Detection in IoT and IIoT: Comparing Lightweight Machine Learning Techniques Using TON_IoT, WUSTL-IIOT-2021, and EdgeIIoTset Datasets // IEEE Access. — Vol. 13. — Pp. 73468–73485. DOI: https://doi.org/10.1109/ACCESS.2025.3554083 (in Eng.).

Koay A.M.Y., Ko R.K.L., Hettema H., Radke K. (2023). Machine Learning in Industrial Control System (ICS) Security: Current Landscape, Opportunities and Challenges // Journal of Intelligent Information Systems. — Vol. 60. — No. 2. — Pp. 377–405. DOI: https://doi.org/10.1007/s10844-022-00753-1 (in Eng.).

Le T.-T.-H., Oktian Y.E., Kim H. (2022). XGBoost for Imbalanced Multiclass Classification-Based Industrial Internet of Things Intrusion Detection Systems // Sustainability. — Vol. 14. — No. 14. Article 8707. DOI: https://doi.org/10.3390/su14148707 (in Eng.).

Moustafa N. (2021). A New Distributed Architecture for Evaluating AI-Based Security Systems at the Edge: Network TON_IoT Datasets // Sustainable Cities and Society. — Vol. 72. — Article 102994. DOI: https://doi.org/10.1016/j.scs.2021.102994 (in Eng.).

Pinto A., Herrera L.-C., Donoso Y., Gutierrez J.A. (2023). Survey on Intrusion Detection Systems Based on Machine Learning Techniques for the Protection of Critical Infrastructure // Sensors. — Vol. 23. — No. 5. Article 2415. DOI: https://doi.org/10.3390/s23052415 (in Eng.).

Sezgin A., Boyacı A. (2023). AID4I: An Intrusion Detection Framework for Industrial Internet of Things Using Automated Machine Learning // Computers, Materials & Continua. — Vol. 76. — No. 2. — Pp. 2121–2143. DOI: https://doi.org/10.32604/cmc.2023.040287 (in Eng.).

Yu J., Shvetsov A.V., Alsamhi S.H. (2024). Leveraging Machine Learning for Cybersecurity Resilience in Industry 4.0: Challenges and Future Directions // IEEE Access. — Vol. 12. — Pp. 159579–159596. DOI: https://doi.org/10.1109/ACCESS.2024.3482987 (in Eng.).

Zhang J., Luo C., Carpenter M., Min G. (2023). Federated Learning for Distributed IIoT Intrusion Detection Using Transfer Approaches // IEEE Transactions on Industrial Informatics. — Vol. 19. — No. 7. — Pp. 8159–8169. DOI: https://doi.org/10.1109/TII.2022.3216575 (in Eng.).

Zolanvari M., Teixeira M.A., Gupta L., Khan K.M., Jain R. (2019). Machine Learning-Based Network Vulnerability Analysis of Industrial Internet of Things // IEEE Internet of Things Journal. — Vol. 6. — No. 4. — Pp. 6822–6834. DOI: https://doi.org/10.1109/JIOT.2019.2912022 (in Eng.).

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 512–528
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.028
МРНТИ 20.16.05

INFORMATION SECURITY IN GOVERNMENT CLOUD SERVICES: RISK ASSESSMENT AND PROTECTION STRATEGIES

G. Shangytbayeva[2], A.N. Amanov[3], M.K. Issenova[1], G.E. Berikkhanova[1], K.O. Shakerkhan[1*]
[1]Shakarim University, Semey, Kazakhstan;
[2]Zhubanov Aktobe Regional University, Aktobe, Kazakhstan;
[3]Khoja Akhmet Yassawi International Kazakh–Turkish University, Turkistan, Kazakhstan.
E-mail: q.shakerkhan@shkarim.kz

Gulmira Shangytbayeva — PhD, Associate Professor, Zhubanov Aktobe Regional University, Aktobe, Kazakhstan
https://orcid.org/0000–0003–4615–5756;
Anuarbek Amanov — PhD, Khoja Akhmet Yassawi International Kazakh–Turkish University, Turkestan, Kazakhstan
https://orcid.org/0000–0003–0638–6859;
Madina Isenova — Master of Pedagogical Sciences, Lecturer, Shakarim University, Semey, Kazakhstan
https://orcid.org/0009–0006–8551–5063;
Gulnaz Berikhanova — Doctor of Physical and Mathematical Sciences, Professor, Shakarim University, Semey, Kazakhstan
https://orcid.org/0000–0001–5013–1226;
Kapan Shakerhan — PhD, Shakarim University, Semey, Kazakhstan
E-mail: q.shakerhan@shakarim.kz, https://orcid.org/0000–0001–5013–1226.

© G. Shangytbayeva, A.N. Amanov, M.K. Issenova, G.E. Berikkhanova, K.O. Shakerkhan

Abstract. This article examines information security issues in cloud services of government bodies, risk assessment methods, and protection strategies. The relevance of this topic stems from the dichotomous conflict between economic efficiency and strict regulatory security requirements for government information resources within the three-tier cloud computing model—SaaS (software as a service), PaaS (platform as a service), and IaaS (infrastructure as a service)—as the primary factor determining the scientific and educational value of research. The integration of public sector entities into cloud computing infrastructure is considered a strategic priority for the purpose of optimizing operational costs associated with the use of information and communications

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

technologies and increasing the functional efficiency of e-government institutions. However, this process creates new risks and threats to data confidentiality, integrity, and availability. The study analyzes five main classes of cyber threats in cloud services (traditional attacks, functional attacks, client–side attacks, virtualization threats, complex threats). Modern methods of protecting virtual infrastructure (hypervisor protection, virtual machine migration control, management system protection) are described. Prospective cloud security technologies – self–protected data and trusted monitors – are considered. The relevance of the study is confirmed by the increasing frequency of targeted cyberattacks on cloud infrastructures and the lack of unified risk assessment methodologies for hybrid environments that take into account national specifics. The object of the research is cloud services within the government administration perimeter, and the subject is the methodology for assessing risk factors and approaches to creating a multi–level security architecture. The aim of the work is to develop an adaptive risk assessment model and formulate practical protection strategies for government cloud platforms.

Keywords: cloud computing, information security, public sector, cyber threats, virtualization, Hyper–V, risk assessment

For citation: Shangytbayeva G., Amanov A.N., Issenova M.K., Berikkhanova G.E., Shakerkhan K.O. (2026). Information security in government cloud services: risk assessment and protection strategies // International journal of information and communication technologies. Vol. 7. No. 27. Pp. 512–528. https://doi.org/10.54309/IJICT.2026.27.3.028. (In Kaz.).

Conflict of interest: The authors declare that there is no conflict of interest.

МЕМЛЕКЕТТІК ОРГАНДАРДЫҢ БҰЛТТЫ СЕРВИСТЕРІНДЕГІ АҚПАРАТТЫҚ ҚАУІПСІЗДІК: ТӘУЕКЕЛДЕРДІ БАҒАЛАУ ЖӘНЕ ҚОРҒАУ СТРАТЕГИЯЛАРЫ

Г.А. Шангытбаева[2], А.Н. Аманов[3], М.К. Исенова[1], Г.Е. Берикханова[1], Қ.О. Шакерхан[1*]

[1]Шәкәрім университеті, Семей, Казақстан;
[2]Қ. Жұбанов атындағы Ақтөбе өңірлік университеті, Ақтөбе, Қазақстан;
[3]Қожа Ахмет Яссауи атын. Халықаралық Қазақ–Түрік Университеті, Түркістан, Қазақстан.
E-mail: q.shakerkhan@shkarim.kz

Шангытбаева Гүлмира — PhD доктор, қауымдастырылған профессор, Қ. Жұбанов атындағы Ақтөбе өңірлік университеті, Ақтөбе, Қазақстан
https://orcid.org/0000–0003–4615–5756;
Аманов Ануарбек — PhD доктор, Қожа Ахмет Яссауи атын. Халықаралық Қазақ–Түрік Университеті, Түркістан, Қазақстан
https://orcid.org/0000–0003–0638–6859;
Исенова Мадина — педагогика ғылымдарының магистрі, оқытушысы, Шәкәрім Университеті, Семей, Қазақстан

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

https://orcid.org/0009–0006–8551–5063;
Берикханова Гүлназ — ф–м.ғ.д., профессор, Шәкәрім Университеті, Семей, Қазақстан
https://orcid.org/ 0000–0001–5013–1226;
Шакерхан Қапан — PhD доктор, Шәкәрім университеті, Семей, Қазақстан
E-mail: q.shakerhan@shakarim.kz, https://orcid.org/0000–0001–5013–1226.

© Г.А. Шангытбаева, А.Н. Аманов, М.К. Исенова, Г.Е. Берикханова, Қ.О. Шакерхан

Аннотация. Бұл мақалада мемлекеттік органдардың бұлтты сервистеріндегі ақпараттық қауіпсіздік мәселелері, тәуекелдерді бағалау әдістері және қорғау стратегиялары қарастырылады. Тақырыптың өзектілігі зерттеу жұмысының ғылыми-танымдық мәнін айқындайтын негізгі фактор ретінде бұлтты есептеу технологияларының үш деңгейлі моделі – SaaS (бағдарламалық қамтамасыз ету қызмет ретінде), PaaS (платформа қызмет ретінде) және IaaS (инфрақұрылым қызмет ретінде) – аясындағы экономикалық тиімділік пен мемлекеттік ақпараттық ресурстарға қойылатын қатаң реттеушілік қауіпсіздік талаптары арасындағы туындайтын дихотомиялық қайшылықтан туындап отыр. Мемлекеттік сектор субъектілерінің бұлтты есептеу инфрақұрылымына интеграциялануы ақпараттық-коммуникациялық технологияларды пайдаланумен байланысты операциялық шығындарды оңтайландыру және электрондық үкімет институтының функционалдық тиімділігін арттыру мақсатындағы стратегиялық басымдық ретінде қарастырылады. Алайда, бұл процесс деректердің құпиялылығына, тұтастығына және қолжетімділігіне байланысты жаңа қауіптер мен тәуекелдерді туғызады. Зерттеу барысында бұлтты сервистердегі киберқауіптердің бес негізгі класы (дәстүрлі атакалар, функционалды атакалар, клиентке атакалар, виртуалдауға атакалар, кешенді атакалар) талданды. Виртуалды инфрақұрылымды қорғаудың заманауи әдістері (гипервизорды қорғау, виртуалды машиналарды көшіруді қадағалау, басқару жүйелерін қорғау) сипатталды. Болашақтың бұлтты қауіпсіздік технологиялары – самозащищенные деректер және сенімді мониторлар қарастырылды. Зерттеудің өзектілігі бұлтты инфрақұрылымдарға бағытталған кибершабуылдар жиілігінің артуымен және ұлттық ерекшелікті ескеретін гибридті орталар үшін тәуекелдерді бағалаудың бірыңғай әдістемелерінің болмауымен расталады. Зерттеу нысаны мемлекеттік басқару контурындағы бұлтты сервистер, пәні – тәуекел факторларын бағалау әдіснамасы және көп деңгейлі қорғау архитектурасын құру тәсілдері болып табылады. Жұмыстың мақсаты – тәуекелдерді бағалаудың адаптивті моделін әзірлеу және мемлекеттік бұлтты платформалар үшін практикалық қорғау стратегияларын тұжырымдау.

Түйінді сөздер: бұлтты есептеулер, ақпараттық қауіпсіздік, мемлекеттік сектор, киберқауіптер, виртуалдау, Hyper–V, тәуекелдерді бағалау

Дәйексөздер үшін: Шангытбаева Г.А., Аманов А.Н., Исенова М.К., Берикханова Г.Е., Шакерхан Қ.О. (2026). Мемлекеттік органдардың бұлтты сервистеріндегі ақпараттық қауіпсіздік: тәуекелдерді бағалау және қорғау

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

стратегиялары // Халықаралық ақпараттық және коммуникациялық технологиялар журналы. Т. 7. No 27. Б. 512–528. https://doi.org/10.54309/IJICT.2026.27.3.028. (Қаз. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ В ОБЛАЧНЫХ СЕРВИСАХ ГОСУДАРСТВЕННЫХ ОРГАНОВ: ОЦЕНКА РИСКОВ И СТРАТЕГИИ ЗАЩИТЫ

Г.А. Шангытбаева[2], А.Н. Аманов[3], М.К. Исенова[1], Г.Е. Берикханова[1], Қ.О. Шакерхан[1*]

[1]Университет Шакарима, Семей, Казахстан;

[2]Актюбинский региональный университет имени К.Жубанова, Актобе, Казахстан;

[3]Международный казахско–турецкий университет им. Ходжа Ахмета Яссауи, Туркестан, Казахстан.

E-mail: q.shakerkhan@shkarim.kz

Шангытбаева Гульмира — доктор PhD, ассоциированный профессор, Актюбинский региональный университет имени К.Жубанова, Актобе, Казахстан https://orcid.org/0000–0003–4615–5756;

Аманов Ануарбек — доктор PhD, Международный казахско–турецкий университет им. Ходжи Ахмета Яссави, Туркестан, Казахстан https://orcid.org/0000–0003–0638–6859;

Исенова Мадина — магистр педагогических наук, преподаватель, Университет Шакарима, Семей, Казахстан https://orcid.org/0009–0006–8551–5063;

Бериханова Гульназ — доктор физико–математических наук, профессор, Университет Шакарима, Семей, Казахстан https://orcid.org/ 0000–0001–5013–1226;

Шакерхан Капан — доктор PhD, Университет Шакарима, Семей, Казахстан E-mail: q.shakerkhan@shkarim.kz, https://orcid.org/0000–0001–5013–1226.

© Г.А. Шангытбаева, А.Н. Аманов, М.К. Исенова, Г.Е. Берикханова, Қ.О. Шакерхан

Аннотация. В данной статье рассматриваются вопросы информационной безопасности в облачных сервисах государственных органов, методы оценки рисков и стратегии защиты. Актуальность темы обусловлена дихотомическим конфликтом, возникающим между экономической эффективностью и строгими требованиями нормативной безопасности государственных информационных ресурсов в рамках трехуровневой модели облачных вычислительных технологий — SaaS (программное обеспечение как услуга), PaaS (платформа как услуга) и

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

IaaS (инфраструктура как услуга) — как основным фактором, определяющим научную и познавательную ценность исследовательской работы. Интеграция субъектов государственного сектора в инфраструктуру облачных вычислений рассматривается как стратегический приоритет с целью оптимизации операционных затрат, связанных с использованием информационно-коммуникационных технологий, и повышения функциональной эффективности института электронного правительства. Однако этот процесс порождает новые риски и угрозы конфиденциальности, целостности и доступности данных. В ходе исследования проанализированы пять основных классов киберугроз в облачных сервисах (традиционные атаки, функциональные атаки, атаки на клиента, угрозы виртуализации, комплексные угрозы). Описаны современные методы защиты виртуальной инфраструктуры (защита гипервизора, контроль миграции виртуальных машин, защита систем управления). Рассмотрены перспективные технологии облачной безопасности – самозащищенные данные и доверенные мониторы. Актуальность исследования подтверждается нарастающей частотой целевых кибератак на облачные инфраструктуры и отсутствием унифицированных методик оценки рисков для гибридных сред с учетом национальной специфики. Объектом исследования являются облачные сервисы в контуре государственного управления, предметом – методология оценки факторов риска и подходы к созданию многоуровневой архитектуры защиты. Цель работы заключается в разработке адаптивной модели оценки рисков и формулировании практических стратегий защиты для государственных облачных платформ.

Ключевые слова: облачные вычисления, информационная безопасность, государственный сектор, киберугрозы, виртуализация, Hyper–V, оценка рисков

Для цитирования: Шангытбаева Г.А., Аманов А.Н., Исенова М.К., Берикханова Г.Е., Шакерхан Қ.О. (2026). Информационная безопасность в облачных сервисах государственных органов: оценка рисков и стратегии защиты // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 27. Стр. 512–528. https://doi.org/10.54309/IJICT.2026.27.3.028. (На каз.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Кіріспе.

Қазіргі уақытта IT–индустриясында ақпараттық технологияларды дамытудың төрт негізгі үрдісі бар: бұлтты технологиялар, пайдаланушылардың мобильділігі, үлкен деректер (big data) және әлеуметтік жүйелер. Бұлтты есептеулер ақпараттық жүйелердің эволюциялық дамуындағы сапалық жаңа кезеңді білдіреді. Мемлекеттік органдардың осы технологиялық платформаға көшуі Қазақстан Республикасы Үкіметінің ұзақмерзімді цифрлық трансформация стратегиясындағы басым бағыттардың бірі ретінде айқындалады. Мұндай көшудің тиімділігі электрондық үкімет жүйесінің функционалдық нәтижелігін арттыру, құжат айналымы процестерінің жеделдігін қамтамасыз ету, ақпараттық қауіпсіздік деңгейінің сенімділігін күшейту, сондай-ақ ақпараттық-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

коммуникациялық инфрақұрылымды ұстауға кететін шығындарды елеулі төмендету арқылы жүзеге асырылады.

Дегенмен, мемлекеттік секторда бұлттық сервистерді енгізу процесі жаңа сын-қатерлер мен тәуекелдік факторлардың туындауына себеп болмақ. Атап айтқанда, ақпараттық қауіпсіздікті қамтамасыз ету мәселелері осы технологияларды мемлекеттік басқару саласына енгізуді тежейтін негізгі кедергілердің бірі болып табылады. Бұл жағдай қауіпсіздік талаптары мен технологиялық мүмкіндіктер арасындағы теңгерімді іздестіруді ғылыми-қолданбалы зерттеулердің өзекті бағытына айналдырады.

Зерттеулер көрсеткендей, интернет пайдаланушылардың 78%–ы өздерінің компьютерлік жүйелерін желілік шабуылдардан қалай қорғау керектігін білмейді. Бұл мемлекеттік органдар үшін ақпараттық қауіпсіздік мәселелерін тереңдетіп зерделеудің маңыздылығын көрсетеді (Kudra et al., 2011)

Зерттеудің мақсаты – мемлекеттік органдардың бұлтты сервистеріндегі ақпараттық қауіпсіздік тәуекелдерін бағалау және қорғау стратегияларын әзірлеу.

Зерттеудің міндеттері:

- бұлтты сервистердегі киберқауіптерді классификациялау;
- виртуалды инфрақұрылымды қорғаудың заманауи әдістерін талдау;
- болашақтың бұлтты қауіпсіздік технологияларын зерттеу;
- мемлекеттік органдар үшін бұлтты провайдерді таңдау критерийлерін әзірлеу.

Бұл тақырыптың негіздемесі көп пәтерлілік және жауапкершілікті бөлу қағидаттарына негізделген қоғамдық бұлттық архитектура мен ұлттық заңнама талаптары арасындағы негізгі қарама–қайшылыққа негізделген. Мақалада көрсетілгендей, қаржылық емес тәуекел санаттары мемлекеттік органдар үшін бірінші кезектегі мәселе болып табылады: әлеуметтік–саяси тұрақтылық, мемлекеттік егемендікті сақтау және қорғаныс қабілеті. Бұл зерттеудің нысаны – мемлекеттік органдарда орналастырылған бұлттық қызметтер (SaaS, PaaS, IaaS), ал тақырыбы – тәуекелдерді бағалау әдіснамасы және қауіпсіздік архитектурасы. Бұл жұмыстың практикалық маңыздылығы Қазақстан үкіметінің ҚазақстанVүкіметінің мемлекеттік органдардың бүкіл IT инфрақұрылымын мемлекеттік қызметтерді көрсетуге қатысты өзекті мәселелерді шешетін, электрондық үкіметті жетілдіретін, мемлекеттік органдар арасындағы электрондық құжат айналымының жылдамдығын арттыратын, ақпараттық қауіпсіздіктің жоғары сенімділігін қамтамасыз ететін және IT инфрақұрылымының қымбат қаржылық шығындарын азайтатын инновациялық технологиялардың жаңа моделіне көшіру мәселесін көтеруінде жатыр. Сондықтан, практикалық маңыздылығы Қазақстан үкіметінің бүкіл IT инфрақұрылымын бұлттық технологияларға көшіру үшін бұлтты есептеу платформаларын таңдау мәселесін шешу, осылайша мемлекеттік органдардың жоғарыда аталған барлық IT мәселелерін шешу болып табылады (Armbrust et al., 2010)

Әдістер мен материалдар.

Бұлттық қауіпсіздік – бұлттық инфрақұрылымды сынға алушылар

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

көтерген негізгі мәселелердің бірі. Көпшілік үшін деректерін бұлттағы компьютерлерге жылжыту эмоционалды түрде мазасыздық тудыруы мүмкін, онда олар оны басқара алмайды. Сонымен қатар, бұл тәсіл сәйкестікті және әртүрлі реттеуші аспектілерді ескеруді талап етеді. Алайда, шын мәнінде, бұлттық орта дәстүрлі деректер орталығынан кем емес, тіпті одан да қауіпсіз болуы мүмкін. Дегенмен, бұлттық ортамен жұмыс істеген кезде ақпараттық қауіпсіздікке деген көзқарасыңыз дәстүрлі ортада үйреніп қалған тәсіліңізден түбегейлі өзгеше болады. Онлайн ақпараттық қауіпсіздік бүгінгі таңда әрбір пайдаланушыға әсер ететін өзекті мәселе болып қала береді. Сондай–ақ, сауалнамаға қатысқан интернет пайдаланушыларының 78%–ы компьютерлік жүйелерін әртүрлі желілік шабуылдардан қалай қорғау керектігін білмейтінін мойындағанын атап өткен жөн. Сондықтан, бұлттық есептеулерге көшу туралы шешім қабылдаған кезде, ақпараттық қауіпсіздіктің маңызды аспектілері бұлттық провайдерлермен толығымен қамтылғанын түсінуіңіз керек.

Сіздің бизнесіңізге ешқандай қатысы жоқ, бірақ деректеріңіздің қауіпсіздігіне айтарлықтай қауіп төндіруі мүмкін сыртқы факторлар бар. Мысалы, келесі оқиғалардың кез келгені сіздің инфрақұрылымыңызда мәселелер тудыруы мүмкін (Barham et al., 2017)

• Сізге бұлттық қызметтерді ұсынатын компания банкрот деп жарияланады, оның активтері (серверлерді қоса алғанда) тәркіленеді және провайдер қызмет көрсетуді тоқтатады.

• Сізге қатысы жоқ үшінші тарап (немесе одан да жаманы, бәсекелес) сіздің бұлттық провайдеріңізге сотқа береді және сот арқылы бұлттық провайдерге тиесілі барлық серверлерге толық қол жеткізеді.

• Сіздің бұлттық провайдеріңіздің инфрақұрылым компоненттерін, әсіресе физикалық қол жеткізуді басқаруға келгенде, жеткілікті түрде қорғай алмауы сіздің жүйелеріңізге қауіп төндіруі мүмкін.

Келесі қадамдар кез келген жағдайда қабылдауға болатын, бірақ көптеген пайдаланушылар жиі бағаламайтын осы мәселелерді шешуге көмектеседі:

Барлық ақпаратты, барлық желілік байланыстарды шифрлаңыз және қашықтан сақтық көшірмелер жасаңыз (бұл мақсатта коммерциялық және ашық бастапқы кодты шешімдер бар). Бұл сіздің бұлт провайдеріңіз жоғалып кетсе де, ағымдағы және тарихи деректердің қалпына келуін қамтамасыз етеді.

Барлық мүмкін сценарийлер үшін аталған шараларды қолдануды қарастырайық:

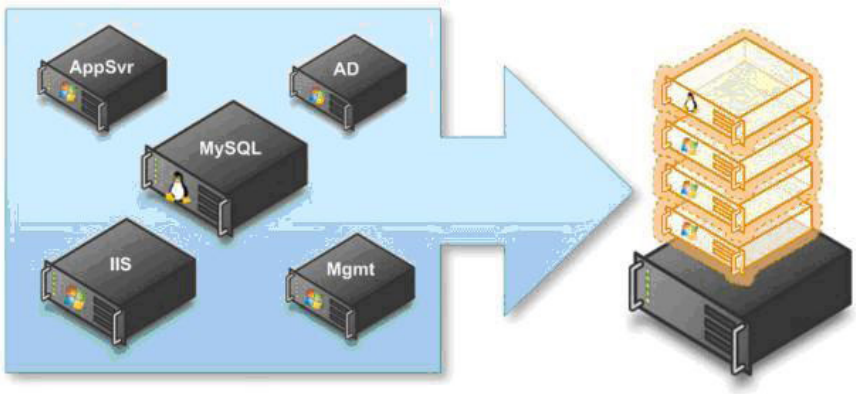
1. Бұлттық қызмет провайдері банкротқа ұшыраса не болады ?

Бұл сценарий үшін бірнеше мүмкін сценарийлер бар: банкроттық, басқа қызмет саласына қайта назар аудару үшін бизнестің жабылуы немесе электр қуатының кең көлемді және ұзаққа созылған үзілуіне байланысты ұзақ уақытқа тоқтап қалу. Нәтижесіне қарамастан, сіз басқа компанияның әрекеттеріне байланысты өндірістік жүйелеріңізге қол жеткізу мүмкіндігін жоғалту қаупін тудырасыз. Сонымен қатар, сіз деректеріңізді басқаратын ұйымның бұрын жасалған қызмет көрсету деңгейі туралы келісімдерге сәйкес оны қорғай алмау

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

қаупін тудырасыз (1 сур., қара).

Бұл жағдайда ең маңызды аспект – қашықтан сақтық көшірмелерді үнемі жасау және сақтық көшірмелерді өндірістік ортаңыздан тыс сақтау. Бұл шара сізді бұлт провайдеріңіздің тоқтап қалуының теріс салдарынан қорғауы керек. Одан да жақсысы, ауыстыру инфрақұрылымын іске қоса алатын екінші бұлт провайдерін пайдаланыңыз. (Ванина т.б., 2015)



Сур. 1. Серверді виртуалдандыру

2. Бұлттық қызмет провайдеріңізді деректеріңізді жариялауға мәжбүрлейтін сотқа шақыру қағазы түскенде не болады?

Әрине, егер сіз сот ісіне қатысып, сотқа шақыру қағазы тікелей сізге жолданса, сіз сақтық шараларын қабылдағаныңызға және деректеріңіз бұлттық ортада немесе өзіңіздің IT инфрақұрылымыңызда сақталғанына қарамастан, деректерді соттарға беруге міндеттісіз. Дегенмен, біз мұнда басқа сценарийді қарастырып жатырмыз – сотқа шақыру қағазы сізге емес, сіздің бұлттық провайдеріңізге жолданған және сіз сот ісіне қатыспаған және онымен ешқандай байланысыңыз болмаған жағдай.

Таза техникалық тұрғыдан алғанда, сотқа шақыру қағазы мәселелердің тар ауқымын қамтуы және сізге әсер етпеуі керек. Дегенмен, бәсекелес фирмалар арасындағы дамып келе жатқан технологиялардағы басымдықтарға қатысты дауларға қатысты сотқа шақыру қағаздарының дұрыс тұжырымдалатынына, атап айтқанда, мүдделі тараптардың жеткілікті тар шеңберін қамтитындай етіп жазылатынына толық сенімді бола алмайсыз. Сонымен қатар, сотқа шақыру қағазы берілгенін білетініңізге сенімді бола алмайсыз.

Деректеріңізді толық шифрлау ғана сізді бұл жағдайдан қорғайды. Шақыру қағазы сіздің бұлт провайдеріңізден сотқа деректеріңізді және оған кіру мүмкіндігін беруді талап етуі мүмкін, бірақ провайдерде сіздің кіру кілтіңіз немесе шифрды ашу кілтіңіз болмайды. Оларды алу үшін сот сізбен байланысып, шақыру қағазын шығаруы керек. Нәтижесінде, сіз бұлттағы деректеріңізді өзіңіздің жеке деректер орталығыңыздағыдай бақылау деңгейіне ие боласыз.

3. Бұлттық қызмет провайдері өз желісі үшін тиісті қауіпсіздікті қамтамасыз ете алмаса, не болады?

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Бұлт провайдерін таңдаған кезде, олардың физикалық қауіпсіздігімен және олардың желісі мен жеке хосттарын қорғауды қалай ұйымдастыратынымен танысуыңыз керек. Қаншалықты парадоксалды болып көрінсе де, ең қауіпсіз бұлт провайдері – сіз ешқашан бұл ақпаратты ала алмайтын және виртуалды дананы іске қосатын сервердің физикалық орналасқан жерін біле алмайтын провайдер. Егер сіз мұны болжай алмасаңыз да, сіздің ұйымыңыз туралы ақпарат жинайтын арнайы шабуылдаушы сіздің хостингіңізді қамтамасыз ететін ортаның физикалық орналасқан жерін анықтауға тырысқанда осындай қиындықтарға тап болуы мүмкін (Abdaly, 2021)

Бүгінгі таңда бұлттық қауіпсіздік саласындағы жетекші сарапшы – Cloud Security Alliance (CSA). Бұл ұйым бұлттық есептеу тәуекелдерін бағалау кезінде ескеру қажет жүздеген нюанстар мен ұсыныстарды сипаттайтын нұсқаулық шығарды және жақында жаңартты. Нұсқаулық 76 беттен тұрады және сізді осындай ұзақ құжатты оқу қиындықтарынан құтқару үшін, онда әлеуетті бұлттық қызмет провайдеріне алдымен қою керек ең маңызды сұрақтар көрсетілген. Сондай–ақ, онда сіз күтуіңіз керек жауаптар берілген. Бұлттық қауіпсіздік аспектілерін қамтитын тағы бір ұйым – Trusted Computing Group (TCG), ол осы және басқа салалардағы бірнеше стандарттардың, соның ішінде кеңінен қолданылатын Trusted Storage, Trusted Network Connect (TNC) және Trusted Platform Module (TPM) стандарттарының авторы.

Келесі сұрақтар сіз кімнің бұлттық қызметтерін пайдалануды жоспарлап отырған провайдерге қоюға болатын негізгі сұрақтар болып табылады. Қойылған мәселелердің шешімін табу барысында стандарттау негізіндегі тәсілдерді қолданудың тиімділігін алдын ала саралау қажет. Бұл ұстаным ақпараттық қауіпсіздікті қамтамасыз етудің барлық деңгейлеріне ортақ сипатқа ие. Зерттеулер мен практикалық тәжірибе көрсеткендей, меншіктік (проприетарлық) архитектуралар ашық стандарттарға негізделген жүйелермен салыстырғанда қорғаныс механизмдерінің төменгі сенімділік деңгейін қамтамасыз етеді. Бұл бағалау нарық субъектілерінің, мемлекеттік институттардың және стандарттау жөніндегі уәкілетті органдардың кең ауқымды консенсусы ретінде қалыптасқан және салалық қауымдастық тарапынан жалпы мойындалған методологиялық принцип болып табылады. Міне, сондықтан Advanced Encryption Standard (AES) және Transport Layer Security (TLS) сияқты стандарттар кең таралған. Олар жылдар бойы талдау мен жетілдіруден өтті. Сонымен қатар, жалпы қабылданған стандарттарға негізделген қауіпсіздік жүйелерін пайдалану арқылы тұтынушылар қажет болған жағдайда қызмет көрсетушілерді ауыстыра алу мүмкіндігінің қосымша артықшылығына ие болады, себебі көптеген провайдерлер стандартталған шешімдерді қолдайды. Тағы бір нақтылауға тұрарлық мәселе: провайдеріңіздің өз уәделерін орындауын қалай қамтамасыз ете аласыз? Қызмет көрсету деңгейі туралы келісім (SLA), келісімшарт немесе провайдердің міндеттемелерін анық көрсететін басқа жазбаша құжат сізге мұны жүзеге асыруға көмектеседі (Трофименко, 2020)

Әрі қарай, жалпыдан нақтыға дейінгі бірқатар сұрақтарды қарастырайық,

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

оларды нақтылауға тұрарлық:

1. Сақталған деректердің қауіпсіздігі.

Қызмет көрсетуші сақталған деректердің қауіпсіздігін қалай қамтамасыз етеді?

Сақталған деректерді қорғаудың ең жақсы шарасы – шифрлау технологияларын пайдалану. Қызмет көрсетуші рұқсатсыз кірудің алдын алу үшін өз серверлерінде сақталған клиенттік ақпаратты әрқашан шифрлауы керек. Қызмет көрсетуші сонымен қатар деректерді қажет етпейтін және болашақта қажет болмайтын кезде оларды біржола жоюы керек.

2. Тасымалдау кезінде деректерді қорғау.

Провайдер тасымалдау кезіндегі деректердің қауіпсіздігін қалай қамтамасыз етеді (бұлт ішінде және бұлтқа/бұлттан жолда)?

Берілген деректер әрқашан шифрлануы және пайдаланушыға тек аутентификациядан кейін ғана қолжетімді болуы керек. Бұл тәсіл ешкімнің бұл деректерді өзгерте немесе оқи алмайтынына кепілдік береді, тіпті олар желідегі сенімсіз түйіндер арқылы қол жеткізсе де. Бұл технологиялар көптеген жылдар бойы әзірленді және сенімді хаттамалар мен алгоритмдердің (мысалы, TLS, IPsec және AES) жасалуына әкелді. Провайдерлер өз хаттамаларын қолданудың орнына осы хаттамаларды еңгізу керек.

3. Аутентификация.

Бұлттық сервис провайдерінің клиенттің жеке басын сәйкестендіру механизмі негізгі қауіпсіздік факторларының бірі болып табылады. Аутентификацияның кең таралған тәсілі ретінде парольдік қорғаныс қолданылады, алайда қауіпсіздік деңгейін күшейту мақсатында провайдерлер сертификаттар мен токендер сияқты күшейтілген аутентификация құралдарын енгізуге бет бұруда. Сонымен қатар, провайдердің LDAP және SAML сияқты салалық стандарттармен үйлесімділігі клиенттің ішкі пайдаланушыларды басқару жүйесімен интеграциялану және авторизация процестерін автоматтандыру тұрғысынан маңызды болып табылады. Бұл провайдердің уәкілетті пайдаланушылар туралы деректердің өзектілігін үздіксіз қамтамасыз етуіне мүмкіндік береді. Ең қолайсыз сценарий ретінде клиенттің провайдерге уәкілетті тұлғалардың статикалық тізімін ұсынуы қарастырылады, мұндай жағдай қызметкерлердің кадрлық өзгерістері (жұмыстан босату, лауазымдық ауысу) кезінде басқарушылық қиындықтар туғызуы мүмкін (Augustine et al., 2025).

4. Пайдаланушыны оқшаулау.

Бір клиентке тиесілі деректер мен қолданбалардың басқа клиенттердің тиісті ресурстарынан бөлінуі міндетті талап болып табылады. Осы тұрғыда ең тиімді шешім әрбір клиент үшін бөлек виртуалды машина (VM) және виртуалды желілік инфрақұрылымды пайдалану болып саналады. Виртуалды машиналар арасындағы шектеуді гипервизорлар қамтамасыз етсе, виртуалды желілер VLAN, VPLS және VPN секілді стандартталған технологиялар негізінде құрылады. Кейбір провайдерлер барлық клиенттердің деректерін ортақ бағдарламалық

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ортада орналастырып, бағдарламалық кодты өзгерту арқылы оқшаулауға тырысады. Алайда, бұл тәсіл сенімсіз және қауіпті деп бағаланады. Біріншіден, стандартты емес кодта осалдықтар болуы мүмкін, олар шабуылдаушыға рұқсатсыз деректерге қол жеткізуге жол ашады. Екіншіден, бағдарламалау қателері бір клиенттің басқа клиенттің ақпаратын кездейсоқ қарауына әкеп соғуы мүмкін. Соңғы кезеңде екі жағдай да тіркелген. Осы себепті, пайдаланушы деректерін тиімді бөлу үшін бөлек виртуалды машиналар мен виртуалды желілерді қолдану неғұрлым сенімді әрі ғылыми негізделген шешім болып табылады.

5. Нормативті–реттеуші мәселелер.

Бұлтты есептеулер саласындағы провайдердің заңнамалық және реттеуші талаптарға сәйкестігі маңызды критерийлердің бірі ретінде қарастырылады. Қолданыстағы заңдар мен ережелер юрисдикцияға қарай елеулі түрде дифференцияланады: олар деректерді трансшекаралық тасымалдауға тыйым салуды, қауіпсіздік шараларының нақты деңгейін, белгілі бір стандарттарды міндетті сақтауды және аудит жүргізу мүмкіндігін талап етуі мүмкін. Сонымен қатар, құқықтық нормалар мемлекеттік органдар мен сот инстанцияларының ақпаратқа қол жеткізуін регламенттейді. Провайдер тарапынан осы талаптарды елемеу клиенттер үшін елеулі заңдық салдарлар мен қаржылық шығындарға әкелуі мүмкін. Провайдер қатаң реттеуші стандарттарды сақтауға және бірыңғай құқықтық стратегияны ұстануға міндетті. Бұл міндеттемелер пайдаланушы деректерінің құпиялылығын, оларды экспорттау шарттарын, салалық стандарттарға сәйкестікті, аудит өткізу тәртібін, деректерді сақтау және жою рәсімдерін, сондай-ақ ақпаратты жария ету жағдайларын (әсіресе бірнеше клиенттердің деректері бір физикалық серверде сақталған кезде өзекті) қамтиды.

6. Оқиғаға жауап беру.

Қызмет көрсетуші оқиғаларға қалай жауап береді және олардың тұтынушылары қаншалықты қатысуы мүмкін?

Кейде істер жоспар бойынша жүрмейді. Осыған орай, сервис провайдерлерінің дағдарыстық жағдайлардағы іс-әрекеттері алдын ала белгіленген және тиісті түрде құжатталған мінез-құлық регламенттеріне негізделуі тиіс. Бұл регламенттер инциденттерді дер кезінде анықтау, олардың теріс салдарларын локализациялау және олардың әсер ету дәрежесін барынша төмендету, сондай-ақ пайдаланушыларды туындаған жағдай туралы уақтылы және толық ақпараттандыру сияқты процедуралық міндеттемелерді қамтуы қажет. Аталмыш шаралар провайдерлердің операциялық тұрақтылығы мен клиенттер алдындағы ашықтығын қамтамасыз етудің негізгі механизмдері болып табылады. Ең дұрысы, олар тұтынушыларға мәселе бойынша үнемі егжей–тегжейлі ақпарат беруі керек. Сонымен қатар, тұтынушылар қауіпсіздік мәселелерінің ықтималдығын бағалап, тиісті шаралар қабылдауы керек (Robert, 2010)

Нәтижелер және оларды талқылау.

Бұлтты технологиялар дамыған сайын киберқауіптер де қатар дамып келеді. Келесі қауіп кластарын ажырату керек:

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Дәстүрлі бағдарламалық жасақтама шабуылдары. Желілік хаттамалардағы, операциялық жүйелердегі, модульдік компоненттердегі және басқа да деңгейлердегі осалдықтар дәстүрлі ақпараттық қауіпсіздік тәуекелдерінің санатына жатады. Осыған орай, оларға қарсы антивирустық бағдарламалық қамтамасыз ету, желілік экран (брандмауэр), шабуылдарды анықтау және алдын алу жүйелері (IPS) секілді классикалық қорғаныс құралдарын пайдалану жеткілікті болып табылады. Алайда, аталған механизмдердің бұлттық инфрақұрылымның ерекшеліктеріне бейімделуі және виртуалдандырылған ортада тиісті тиімділікпен жұмыс істеуі шешуші мәнге ие. Бұлт элементтеріне бағытталған функционалдық шабуылдар ерекше қауіп тобын құрайды. Шабуылдың бұл түрі бұлттық архитектураның көпқабатты құрылымымен, сондай-ақ жүйенің жалпы қауіпсіздік деңгейі оның ең осал буынының қорғаныс деңгейімен анықталатындығы туралы іргелі қауіпсіздік принципімен тығыз байланысты. Атап айтқанда, бұлттық жүйенің алдыңғы шебінде орналасқан кері прокси-серверге қарсы сәтті жүзеге асырылған DoS-шабуылы ішкі желілік коммуникациялар үздіксіз жұмыс істеген жағдайда да бүкіл бұлттық ресурсқа кіруді толығымен бұғаттай алады. Сол сияқты, қолданбалық сервер арқылы жүзеге асырылатын SQL-инъекция шабуылы деректерді сақтау деңгейіндегі қолжетімділік ережелеріне тәуелсіз түрде жүйелік деректерге рұқсатсыз қол жеткізуге мүмкіндік береді.

Функционалдық шабуылдардан тиімді қорғану үшін бұлттық архитектураның әрбір деңгейіне сәйкес мамандандырылған қауіпсіздік шараларын қолдану қажет: прокси-деңгейде DoS-қорғанысы, веб-сервер деңгейінде беттің тұтастығын бақылау, қолданбалық сервер деңгейінде қолданбалық желілік экрандар, дерекқор деңгейінде SQL-инъекциядан қорғаныс, ал сақтау жүйесінде резервтік көшіру және қолжетімділікті бақылау механизмдері. Аталған қорғаныс құралдарының әрқайсысы дербес әзірленгенімен, олар интеграцияланған кешенді бұлттық қауіпсіздік жүйесіне біріктірілмеген. Осы себепті, оларды бірыңғай архитектуралық шешімге кіріктіру міндеті бұлттық инфрақұрылымды жобалау сатысында шешілуге тиіс (Al–zahrani, 2025).

Клиенттік шабуылдар. Шабуылдың бұл түрі веб–ортада жасалған, бірақ бұлтқа да қатысты, себебі клиенттер әдетте бұлтқа браузер арқылы қосылады. Бұған сайтаралық скрипттеу (XSS), веб–сессияны ұрлау, құпия сөзді ұрлау, «ортада тұрған адам» шабуылдары және басқа да шабуылдар кіреді. Бұл шабуылдардан қорғаныс дәстүрлі түрде күшті аутентификацияға және өзара аутентификациямен шифрланған қосылымдарды пайдалануға негізделген. Дегенмен, барлық бұлтты әзірлеушілер мұндай қымбат және көбінесе қолайсыз қауіпсіздік шараларын көтере алмайды (1 кестені қара). Сондықтан, ақпараттық қауіпсіздіктің бұл саласында әлі де шешілмеген мәселелер және жаңа қауіпсіздік шешімдеріне орын бар.

Виртуалдандыру қауіптері. Виртуалды орталар дәстүрлі түрде бұлтты компоненттер үшін платформа ретінде қызмет ететіндіктен, виртуалдандыру жүйесіне жасалған шабуылдар бүкіл бұлтқа да қауіп төндіреді. Бұл қауіп түрі

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

тек бұлтты есептеулерге ғана тән, сондықтан біз оны төменде егжей–тегжейлі қарастырамыз. Кейбір виртуалдандыру қауіптеріне шешімдер пайда бола бастады, бірақ бұл сала салыстырмалы түрде жаңа, сондықтан әлі белгіленген шешімдер әзірленген жоқ. Ақпараттық қауіпсіздік нарығы жақын арада осы қауіп түріне қарсы қорғаныс құралдарын әзірлеуі әбден мүмкін.

Кесте 1. SaaS қызметтері

Citrix Systems	Piston Cloud Computing	DotCloud	Microsoft	Zoho
Salesforce.com				

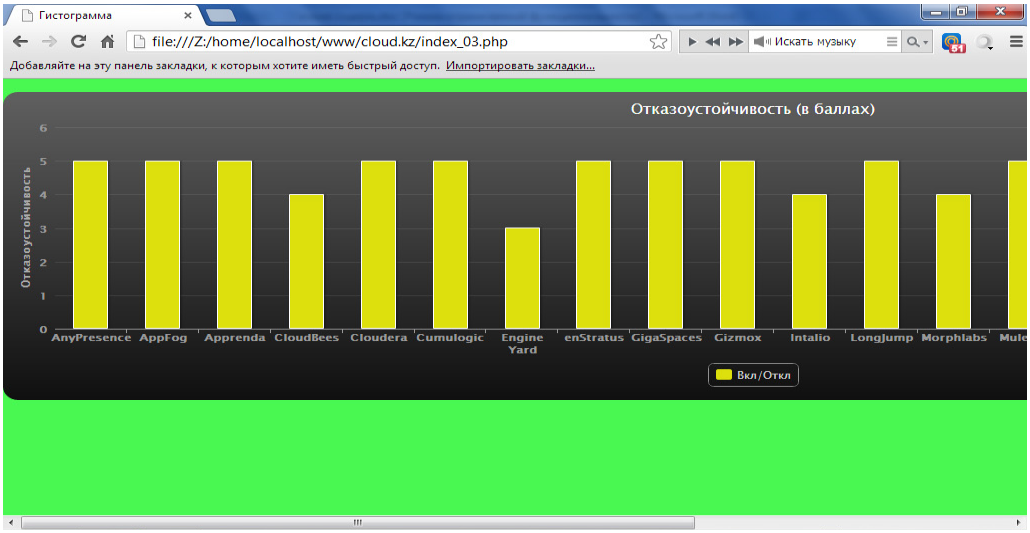
Бұлттық инфрақұрылымды басқарудың күрделі қауіптері. Бұлттық ортаны әкімшілендіру және басқару процестері де дербес қауіпсіздік санатына жататын мәселелер кешенін құрайды. Негізгі сын-қатерлер мыналарды қамтиды: барлық бұлттық ресурстардың толық есепке алынуын қамтамасыз ету, бақылаудан тыс қалған виртуалды машиналардың пайда болуын болдырмау, шаруашылық жүргізуші субъектілер үшін функционалдық маңызы жоқ бизнес-процестердің орындалуын тоқтату, сондай-ақ бұлттың жекелеген қабаттары мен компоненттері арасындағы өзара конфигурациялық байланыстардың тұтастығын сақтау. Аталған міндеттердің әрқайсысы жүйелік бақылау мен басқару стратегиясының ажырамас бөлігі ретінде қарастырылуы тиіс, себебі олардың кез келгеніндегі ақаулық бұлттық сервистің жалпы функционалдығы мен қауіпсіздік деңгейіне тікелей әсер етеді. Бұл қауіп түрі бірыңғай ақпараттық жүйе ретінде бұлтты басқарумен және ақпараттық жүйені күтіп ұстауға қажетсіз шығындарға әкелуі мүмкін бұлттық жұмыстағы теріс пайдаланушыларды немесе басқа да бұзылуларды анықтаумен байланысты. Мысалы, егер бұлт файлды вирусты анықтау үшін жіберуге мүмкіндік берсе, мұндай анықтаулардың ұрлануын қалай болдырмауға болады? Бұл қауіп түрі ең жоғары деңгейлі болып табылады және әмбебап қауіпсіздік шаралары ол үшін жеткіліксіз – оның жалпы қорғанысы әрбір бұлтқа жеке бейімделуі керек. Жалпы тәуекелдерді басқару моделі бұған көмектесе алады, бірақ оны әлі де бұлттық инфрақұрылымдарға дұрыс қолдану қажет (Тұрсынбеков, 2015; Saad et al., 2020)

Қауіптердің алғашқы екі түрі жеткілікті түрде зерттелген және олар үшін қорғаныс құралдары әзірленген, бірақ оларды бұлтта пайдалану үшін бейімдеу қажет. Мысалы, брандмауэрлер периметрді қорғауға арналған, бірақ бұлтта периметрді жеке клиентке тағайындау қиын, бұл қорғауды айтарлықтай қиындатады. Сондықтан брандмауэр технологиясын бұлттық инфрақұрылымға бейімдеу қажет. Мысалы, Check Point қазіргі уақытта осы салада белсенді жұмыс істеп жатыр. Бұлттық есептеулер үшін жаңа қауіп түрі – виртуалдандыру мәселелері. Себебі виртуалдандыру жүйеге қосымша шабуылға ұшырайтын элементтерді енгізеді. Оларға гипервизор, виртуалды машиналарды бір түйіннен екіншісіне көшіретін жүйе және виртуалды машинаны басқару жүйесі жатады.

Басқару жүйелеріне шабуылдар. Бұлттарда, әсіресе қоғамдық бұлттарда қолданылатын виртуалды машиналардың көптігі виртуалды машиналарды

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

жасауды, көшіруді және жоюды сенімді түрде басқара алатын басқару жүйелерін қажет етеді. Басқару жүйелерін бұзу көрінбейтін виртуалды машиналардың жасалуына, белгілі бір машиналардың бұғатталуына және бұлт қабаттарына рұқсатсыз элементтердің енгізілуіне әкелуі мүмкін. Мұның бәрі шабуылдаушыларға бұлттан ақпарат алуға немесе бұлттың бір бөлігін немесе тұтастай басып алуға мүмкіндік береді. Айта кету керек, жоғарыда аталған барлық қауіптер қазіргі уақытта тек гипотетикалық болып табылады, себебі мұндай нақты шабуылдар туралы ақпарат іс жүзінде жоқ. Дегенмен, виртуалдандыру және бұлтты есептеулер танымал болғаннан кейін, осы шабуылдардың барлығы нақты болуы мүмкін. Сондықтан оларды бұлтты жүйелерді жобалау кезеңінде де ескеру қажет (Prabowo et al., 2025; Hendradi et al., 2020)



Сур. 2. Ақаулыққа төзімділік гистограммасы

Қоғамдық өмірде, әсіресе мемлекеттік басқару саласында ақпараттық жүйе әрқашанда нақты уақыт режимінде өмір сүріп келеді. Мұндай жүйе ақпараттың өзін (деректер жиынтығын), оны өңдеудің формаларын, әдістері мен құралдарын, сондай-ақ ақпараттық процестерге қатысатын субъектілерді қамтитын күрделі органикалық тұтастық болып табылады. Осы тұрғыдан алғанда, мемлекеттік басқару жүйесіндегі ақпараттық жүйені (немесе жүйелер кешенін) құру мен жетілдіру мәселесін талқылағанда, біз жекелеген объектілер мен субъектілердің динамикалық күйлерін бейнелейтін деректерді жасау, тіркеу, жинау, өңдеу және сақтау функцияларын атқаратын дәстүрлі мағынадағы жүйе туралы емес, сапалық жағынан жаңа, түбегейлі өзгеше сипаттағы құбылыс туралы сөз қозғап отырмыз.

Мәселенің мәні 1960-жылдары көптеген зерттеушілер мен практиктер ұстанған көзқарастай, әкімшілік процестерге есептеу техникасының, желілік технологиялардың және бағдарламалық қамтамасыз етудің жаңа буынын енгізумен шектелмейді. Бұл міндет әлдеқайда кешенді, көп қырлы, терең

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

мазмұнды және шешімдерін табуы күрделі сипатқа ие. Микроэлектроника, информатика, микропроцессорлық технологиялар, робототехника және заманауи ақпараттық қолдаудың басқа да құралдарын кешенді пайдалану ақпараттық менеджменттің барлық аспектілерін сапалы жаңа ғылыми-техникалық деңгейге көтеруге, озық және жоғары тиімді ақпараттық технологияларды құрудың кең перспективаларын ашады. Бұл ретте, ақпараттық жүйелерді трансформациялау тек техникалық жаңғырту ғана емес, сонымен бірге басқару философиясы мен әдіснамасының эволюциясын да білдіреді. (Sharma et al., 2025; Rittinghaus et al., 2017)

Қорытынды.

Ақпараттық технологиялардың пайда болу барысында ақпараттық жүйелерді құрудың бірнеше модельдері пайда болды. Ол монолитті мейнфрейм архитектурасынан басталды. Мейнфрейм – бұл үлкен, әмбебап, жоғары өнімді, ақауларға төзімді, айтарлықтай енгізу/шығару ресурстары, үлкен көлемдегі жедел жады және сыртқы жады бар, қарқынды пакеттік және нақты уақыт режиміндегі транзакцияларды өңдеумен маңызды жүйелерде пайдалануға арналған сервер. Ол перспективалы клиент–сервер архитектурасымен ауыстырылды. Мұнда арнайы дерекқор сервері болды, ал пайдаланушылар қалың клиенттермен жұмыс істеп, дерекқор серверін босататын. Содан кейін одан да заманауи архитектура пайда болды – үш деңгейлі (немесе көп деңгейлі) архитектура, мұнда қолданбалы логика қолданбалы сервер деп аталатын бөлек компьютерге көшірілді, ал пайдаланушылар веб–браузерлер арқылы жұқа клиенттермен жұмыс істеді. Қазіргі кездегі қолданбалардың көпшілігі осы архитектураны пайдаланып жасалған. Бұл архитектуралар бүкіл IT инфрақұрылымын жергілікті жерде орналастыруды қамтиды. Бұлтты есептеулер ақпараттық жүйе архитектурасының эволюциясындағы келесі қадам болып табылады. Бұл тәсілдің орасан зор артықшылықтарына байланысты, жақын арада көптеген ақпараттық жүйелер бұлтқа көшірілетіні анық. 2011 жылы Қазақстан басқа елдерден озып кету үшін бұлтты технологияларды әзірлеуді бастады. «Қазақтелеком» АҚ еншілес компаниясы KT Cloud Lab жоба менеджері болды. NEC Neva Communication Systems компаниясы Қазақстан астанасы Алматыда өкілдік ашты. Бұл компанияның өкілдері Қазақстанның шағын және орта бизнесінің бұлтты есептеулер IT шешімдеріне деген қажеттіліктерін қанағаттандыруға дайын екендіктерін білдірді. Ғылыми зерттеу нәтижесінде алынған негізгі қорытындыларды тұжырымдайық:

- Бұл зерттеу жұмысында мемлекеттік қызметтер үшін бұлтты есептеу платформаларын таңдау мәселесі қарастырылады.
- Шет елдердегі бұлтты технологиялардың дамуын талдау қарастырылады.
- Қазақстанда бұлтты қызметтерді енгізу, олардың артықшылықтары мен мемлекеттік сектор үшін пайдасы талдаудан өткізілді.
- SaaS, PaaS және IaaS модульдеріне сәйкес, бұлттық қызметтер арасында әлемдегі ең бәсекеге қабілетті және ең үздік компаниялар тізімі кесте түрінде

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

жасалды.

- Қазақстанның мемлекеттік органдарында бұлтты технологияларды енгізу үшін оларды таңдаудың негізгі критерийлері ұсынылады.
- Бұлттық нарықтағы ірі ойыншылардың бұлттық қызметтерге арналған бағдарламалық өнімдері шолудан өтеді.
- Бұлтты технологиялар саласындағы киберқауіптерге тап болған кезде ақпараттық қауіпсіздікті қамтамасыз етудің шешімдері ұсынылады.
- Hyper–V қауіпсіздік саласындағы ең жақсы тәжірибелерге шолу жасайды.

ӘДЕБИЕТТЕР

Al–zahrani I.A.A. (2025). Degree of Use of Cloud Artificial Intelligence in Improving Academic Guidance Systems in Virtual Learning Environments from the Perspective of Faculty Members // Journal of Educational and Human Sciences. No. 43. — Б. 435–466. DOI:10.33193/JEAHS.43.2025.618. [Ағыл. тіл.].

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., & Zaharia, M. (2010). A view of cloud computing // In Communications of the ACM — Т. 53. — No. 4. — Б. 50–58. DOI:https://doi.org/10.1145/1721654.1721672. [Ағыл. тіл.].

Abdaly, S. (2021). Cloud computing and the foundation of security in the cloud // Engineering Journal of Satbayev University. — Т. 143. — No. 2. — Б. 193–198. DOI: https://doi.org/10.51301/vest.su.2021.i2.25. [Ағыл. тіл.].

Alakuu A., Delali Kwasi Dake. (2025). Cloud Computing in Education: A review of Architecture, Applications, and Integration Challenges // International Journal of Computer Applications. — Т. 186. — No. 66. — Б. 49–65. DOI: https://doi.org/10.5120/ijca2025924472. [Ағыл. тіл.].

Ali S., Nawaiseh S. and A., Jameel A. and H., Y. and M. I. Fathy I. and K. (2020). Dr. Emadeldin, A Proposed Model Based on Cloud Computing Technology to Improve Higher Education Institutions Performance (2020) // International Journal of Advanced Research in Engineering and Technology. — Т. 11. — No. 9. — Б. 612–628. https://ssrn.com/abstract=3712837. [Ағыл. тіл.].

Barham P., Dragovic B., Fraser K. et al. (2017). Xen and the Art of Virtualization // ACM SIGOPS Operating Systems Review. — Т. 35. — No. 4. — Б. 164–177. DOI: https://doi.org/10.1145/1165389.945462. [Ағыл. тіл.].

Hendradi P., Abd Ghani M.K. (2020). Artificial Intelligence Influence In Education 4.0 To Architecture Cloud–Based E–Learning System. (2020) // International Journal of Artificial Intelligence Research. DOI: https://doi.org/10.29099/IJAIR.V4I1.109. [Ағыл. тіл.].

Ванина М.Ф., Ерохин А.Г., Фролова Е.А. (2015). Применение облачных технологий в компаниях малого и среднего бизнеса // Век качества. №1. URL: https://cyberleninka.ru/article/n/primenenie–oblachnyh–tehnologiy–v–kompaniyah–malogo–i–srednego–biznesa. [Орыс. тіл.].

Kudra V. (2011). Federal Cloud Computing Strategy // White Houze. URL: https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/egov_docs/federal–cloud–computing–strategy.pdf. [Ағыл. тіл.].

Prabowo R.Y. (2025). The Triad of Cloud, Analytics, And Pedagogy: A Systematic Review of AI–Driven Personalization in Cloud–Based Learning Environments // Journal of Educational Technology. — Т. 10. — No. 1. — Б. 44–59. DOI:10.17977/um039v10i12025p44–59. [Ағыл. тіл.].

Rittinghaus J.W., Ransom J.F. (2017). Cloud Computing – Implementation, Governance, and Security // Taylor and Francis Group. — Б. 174. https://ndl.ethernet.edu.et/bitstream/123456789/26737/1/John%20W.%20Rittinghouse.pdf. [Ағыл. тіл.].

Sharma A.K. Nitin, S. Kumar (eds.) (2025). Establishing AI–Specific Cloud Computing Infrastructure // Hershey: IGI Global Scientific Publishing. — Б. 450. DOI: 10.4018/979–8–3693–9694–0. [Ағыл. тіл.].

Schmidt R. (2010). VCP VMware Certified Professional vSphere 4 Study Guide (Exam VCP410) // McGraw Hill Professional. ISBN: 9780071633673. https://books.google.ru/books?id=C_WoaRZ2FGIC&hl=ru. [Ағыл. тіл.].

Тұрсынбеков Н.Қ. (2015). Қазақстандағы бұлтты технологияларды дамыту перспективалары // Вестник КазНТУ. No. 2. — Б. 45–52. https://ust.kz/word/bulttyq_tehnologiyalar_qazaqstandagy_bulttyq_tehnologiyalardyng_damyy–394041.html. [Қаз. тіл.].

Трофименко Е.В. (2020). Введение в облачные вычисления. Основные понятия // ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ. – Воронеж: Издательский дом ВГУ. — Б. 59. EDN HRRJNZ.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

https://www.elibrary.ru/item.asp?id=47672758. [Орыс. тіл.].

REFERENCES

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., & Zaharia, M. (2010). A view of cloud computing // In Communications of the ACM. — Vol. 53. — No. 4. — Pp. 50–58. DOI: https://doi.org/10.1145/1721654.1721672.[In Eng.].

Augustine Alakuu, Delali Kwasi Dake. (2025). Cloud Computing in Education: A review of Architecture, Applications, and Integration Challenges // International Journal of Computer Applications. — Vol. 186. — No. 66. — Pp. 49–65. DOI: https://doi.org/10.5120/ijca2025924472.[In Eng.].

Abdaly, S. (2021). Cloud computing and the foundation of security in the cloud // Engineering Journal of Satbayev University. — Vol. 143. — No. 2. — Pp. 193–198. DOI: https://doi.org/10.51301/vest.su.2021.i2.25. [In Eng.].

Al–zahrani I.A.A. (2025). Degree of Use of Cloud Artificial Intelligence in Improving Academic Guidance Systems in Virtual Learning Environments from the Perspective of Faculty Members // Journal of Educational and Human Sciences. No. 43. — Pp. 435–466. DOI:10.33193/JEAHS.43.2025.618.[In Eng.].

Ali S., Nawaiseh S. and A., Jameel A. and H., Y. and M. I. Fathy I. and K. (2020). Dr. Emadeldin, A Proposed Model Based on Cloud Computing Technology to Improve Higher Education Institutions Performance // International Journal of Advanced Research in Engineering and Technology. — Vol. 11. — No. 9. — Pp. 612–628. https://ssrn.com/abstract=3712837.[In Eng.].

Barham P., Dragovic B., Fraser K. et al. (2017). Xen and the Art of Virtualization // ACM SIGOPS Operating Systems Review. — Vol. 35. — No. 4. — Pp. 164–177. DOI: https://doi.org/10.1145/1165389.945462. [In Eng.].

Hendradi P., Mohd Khanapi Abd Ghani. (2020). Artificial Intelligence Influence In Education 4.0 To Architecture Cloud–Based E–Learning System. (2020) // International Journal Of Artificial Intelligence Research. DOI: https://doi.org/10.29099/IJAIR.V4I1.109.[In Eng.].

Kudra V. (2011). Federal Cloud Computing Strategy // White Houze. URL: https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/egov_docs/federal–cloud–computing–strategy.pdf.[In Eng.].

Prabowo R.Y. (2025). The Triad of Cloud, Analytics, And Pedagogy: A Systematic Review of AI–Driven Personalization in Cloud–Based Learning Environments // Journal of Educational Technology. — Vol. 10. — No. 1. — Pp. 44–59. DOI:10.17977/um039v10i12025p44–59.[In Eng.].

Robert Schmidt. (2010). VCP VMware Certified Professional vSphere 4 Study Guide (Exam VCP410) // McGraw Hill Professional. ISBN: 9780071633673. https://books.google.ru/books?id=C_WoaRZ2FGIC&hl=ru. [In Eng.].

Rittinghaus J.W., Ransom J.F. (2017). Cloud Computing – Implementation, Governance, and Security // Taylor and Francis Group. — Pp. 174. https://ndl.ethernet.edu.et/bitstream/123456789/26737/1/John%20W.%20Rittinghouse.pdf.[In Eng.].

Sharma A.K. Nitin, S. Kumar (eds.) (2025). Establishing AI–Specific Cloud Computing Infrastructure – Hershey: IGI Global Scientific Publishing. — Pp. 450. DOI: 10.4018/979–8–3693–9694–0.[In Eng.].

Tursynbekov, N.K. (2015). Prospects for the Development of Cloud Technologies in Kazakhstan // Bulletin of KazNTU. — Vol. 2. — Pp. 45–52. https://ust.kz/word/bulttyq_tehnologiyalar_qazaqstandagy_bulttyq_tehnologiyalardyng_damyy–394041.html.[In Kaz.].

Trofimenko E.V. (2020). Vvedenie v oblachnye vychisleniya. Osnovnye ponyatiya / VORONEZHSKII GOSUDARSTVENNYI UNIVERSITET. – Voronezh: Izdatel'skii dom VGU. — Pp. 59. EDN HRRJNZ. https://www.elibrary.ru/item.asp?id=47672758.[In Russ.].

Vanina M.F., Erokhin A.G., Frolova E.A. (2015). Primenenie oblachnykh tekhnologii v kompaniyakh malogo i srednego biznesa // Vek kachestva. №1. URL: https://cyberleninka.ru/article/n/primenenie–oblachnyh–tehnologiy–v–kompaniyah–malogo–i–srednego–biznesa.[In Russ.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 529–544
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.029
УДК 004.85
МРНТИ 81.93.29

OPTIMISATION OF COMMUNICATION COSTS IN TWO-TIER FEDERATED LEARNING FOR CYBERATTACK DETECTION IN IOT NETWORKS

Ye.N. Tursynbek[1,2]*, N. Albanbay[1], D. Djenouri[3]
[1]Satbayev University, Almaty, Kazakhstan;
[2]Kazakh-German University, Almaty, Kazakhstan;
[3]University of the West of England, Almaty, Kazakhstan.
E-mail: y.tursynbek@satbayev.university

Yerlan N. Tursynbek — PhD Candidate, Department of Cybersecurity, Institute of Automation and Information Technologies, Satbayev University, Almaty, Kazakhstan
E-mail: y.tursynbek@satbayev.university, https://orcid.org/0009-0009-9520-3030;
Nurtay Albanbay — PhD, Associate Professor, Department of Cybersecurity, Institute of Automation and Information Technologies, Satbayev University, Almaty, Kazakhstan
https://orcid.org/0000-0002-3393-7380;
Djamel Djenouri — PhD, School of Computing and Creative Technologies, University of the West of England (UWE), Bristol, United Kingdom
https://orcid.org/0000-0001-9333-9860.

© Ye.N. Tursynbek, N. Albanbay, D. Djenouri

Abstract. This paper addresses the challenge of reducing communication overhead in federated learning for distributed IoT networks whilst maintaining high-quality cyberattack detection. The CICIoT2023 dataset, containing data on network traffic and attacks in an Internet of Things environment, was used as the basis for the study. A one-dimensional convolutional neural network (1D-CNN) was used for classification, and a comparison was made between the classical single-layer FL architecture and the proposed two-layer hierarchical scheme with intermediate aggregation gateways. Experiments were conducted on 20 simulated devices over 21 training rounds. The results showed that the two-level architecture achieves an accuracy of 88.82%, which is practically equivalent to the result of the single-level model (88.87%), whilst the difference in Macro F1 and ROC-AUC also remains negligible. The main advantage of the proposed approach lies in a significant reduction in the load on the global server: the number of global parameter transfers is reduced from 420 to

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

15, i.e. by a factor of 28, which corresponds to a 96.4% reduction in traffic. Thus, the proposed architecture demonstrates an effective balance between classification quality and communication efficiency and can be regarded as a promising solution for scalable cybersecurity systems in the IoT.

Keywords: federated learning, intrusion detection, IoT, multi-level aggregation, CICIoT2023, FedAvg

For citation: Tursynbek Ye.N., Albanbay N., Djenouri D. (2026). Optimisation of communication costs in two-tier federated learning for cyberattack detection in IoT networks // International Journal of Information and Communication Technologies. Vol. 7. No. 27. Pp. 529–544. https://doi.org/10.54309/IJICT.2026.27.3.029. (In Russ.).

Conflict of interest: The authors declare that there is no conflict of interest.

IOT ЖЕЛІЛЕРІНДЕ КИБЕРШАБУЫЛДАРДЫ АНЫҚТАУ ҮШІН ЕКІ ДЕҢГЕЙЛІ ФЕДЕРАЦИЯЛАНҒАН ОҚЫТУДАҒЫ БАЙЛАНЫС ШЫҒЫНДАРЫН ОҢТАЙЛАНДЫРУ

Е.Н. Турсынбек[1,2]*, Н. Албанбай[1], Д. Дженоури[3]

[1]Қ.И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті, Алматы, Қазақстан;

[2]Қазақ-Неміс университеті, Алматы, Қазақстан;

[3]Батыс Англия университеті, Бристоль, Ұлыбритания.

E-mail: y.tursynbek@satbayev.university

Тұрсынбек Ерлан Нұржанұлы — PhD докторант, Автоматтандыру және ақпараттық технологиялар институты Киберқауіпсіздік кафедрасы, Қ.И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті, Алматы, Қазақстан
E-mail: y.tursynbek@satbayev.university, https://orcid.org/0009-0009-9520-3030;

Албанбай Нұртай — PhD, Қауымдастырылған профессор, Автоматтандыру және ақпараттық технологиялар институты Киберқауіпсіздік кафедрасы, Қ.И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті, Алматы, Қазақстан
https://orcid.org/0000-0002-3393-7380;

Дженоури Джамель — PhD, Есептеу техникасы және шығармашылық технологиялар мектебі, Батыс Англия университеті (University of the West of England, UWE), Бристоль, Ұлыбритания
https://orcid.org/0000-0001-9333-9860.

© Е.Н. Тұрсынбек, Н. Албанбай, Д. Дженоури

Аннотация. Бұл мақала таратылған IoT желілерінде федерацияланған оқытуда жоғары сапалы кибершабуылдарды анықтауды сақтай отырып, коммуникациялық шығындарды азайту мәселесін қарастырады. Зерттеу негізі ретінде заттар интернеті ортасындағы желілік трафик пен шабуылдар туралы дерек-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

терді қамтитын CICIoT2023 деректер жиынтығы пайдаланылды. Сыныптау үшін бірөлшемді конволюциялық нейрондық желі (1D-CNN) қолданылды, сондай-ақ классикалық бірқабатты FL архитектурасы мен аралық агрегациялық шлюздермен ұсынылған екіқабатты иерархиялық схема арасында салыстыру жүргізілді. Эксперименттер 21 оқыту раундында 20 симуляцияланған құрылғыда өткізілді. Нәтижелер көрсеткендей, екі деңгейлі архитектура 88,82% дәлдікке қол жеткізеді, бұл іс жүзінде бір деңгейлі модельдің (88,87%) нәтижесімен тең, ал Macro F1 және ROC-AUC көрсеткіштеріндегі айырмашылық та елеусіз болып қалады. Ұсынылған тәсілдің басты артықшылығы – жаһандық серверге түсетін жүктемені айтарлықтай азайтуда: жаһандық параметрлерді тасымалдау саны 420-дан 15-ке дейін, яғни 28 есеге қысқарады, бұл трафикті 96,4%-ға азайтуға сәйкес келеді. Осылайша, ұсынылған архитектура жіктеу сапасы мен байланыс тиімділігі арасындағы тиімді тепе-теңдікті қамтамасыз етеді және IoT-дегі масштабталатын киберқауіпсіздік жүйелері үшін перспективалы шешім болып табылады.

Түйінді сөздер: федеративті оқыту, кіруді анықтау, заттар Интернеті, көпдеңгейлі агрегация, CICIoT2023, FedAvg

Дәйексөздер үшін: Тұрсынбек Е.Н., Албанбай Н., Дженоури Д. (2026). IoT желілерінде кибершабуылдарды анықтау үшін екі деңгейлі федерацияланған оқытудағы байланыс шығындарын оңтайландыру // Халықаралық ақпараттық және коммуникациялық технологиялар журналы. Т. 7. No 27. Б. 529–544. https://doi.org/10.54309/IJICT.2026.27.3.029. (Орыс. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

ОПТИМИЗАЦИЯ КОММУНИКАЦИОННЫХ ЗАТРАТ В ДВУХУРОВНЕВОМ ФЕДЕРАТИВНОМ ОБУЧЕНИИ ДЛЯ ОБНАРУЖЕНИЯ КИБЕР ТАК В IOT-СЕТЯХ

Е.Н. Турсынбек[1,2]*, Н. Албанбай[1], Д. Дженоури[3]

[1]Казахский национальный исследовательский технический университет имени К.И. Сатпаева, Алматы, Казахстан;

[2]Казахстанско-Немецкий Университет, Алматы, Казахстан;

[3]Университет Западной Англии, Бристоль, Великобритания.

E-mail: y.tursynbek@satbayev.university

Турсынбек Ерлан Нуржанулы — докторант PhD, Казахский национальный исследовательский технический университет имени К.И. Сатпаева, Алматы, Казахстан

E-mail: y.tursynbek@satbayev.university, https://orcid.org/0009-0009-9520-3030;

Албанбай Нуртай — ассоциированный профессор, Институт автоматики и информационных технологий, кафедра кибербезопасности, обработки и хранения информации, Казахский национальный исследовательский технический университет имени К.И. Сатпаева, Алматы, Казахстан

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

https://orcid.org/0000-0002-3393-7380;
Дженури Джамель — PhD, Школа вычислительных и креативных технологий, Университет Западной Англии, Бристоль, Великобритания https://orcid.org/0000-0001-9333-9860.

© Е.Н. Турсынбек, Н. Албанбай, Д. Дженоури

Аннотация. В статье рассматривается проблема снижения коммуникационных затрат в федеративном обучении для распределённых IoT-сетей при сохранении высокого качества обнаружения кибератак. В качестве основы исследования использован датасет CICIoT2023, содержащий данные о сетевом трафике и атаках в среде Интернета вещей. Для классификации применена одномерная сверточная нейронная сеть (1D-CNN), а сравнение выполнено между классической одноуровневой архитектурой FL и предложенной двухуровневой иерархической схемой с промежуточными шлюзами агрегации. Эксперименты проводились на 20 имитируемых устройствах в течение 21 раунда обучения. Полученные результаты показали, что двухуровневая архитектура обеспечивает точность 88,82%, что практически соответствует результату одноуровневой модели (88,87%), при этом разница по Macro F1 и ROC-AUC также остаётся пренебрежимо малой. Основное преимущество предложенного подхода заключается в резком снижении нагрузки на глобальный сервер: число глобальных передач параметров сокращается с 420 до 15, то есть в 28 раз, что соответствует экономии трафика на 96,4%. Таким образом, предложенная архитектура демонстрирует эффективный баланс между качеством классификации и коммуникационной эффективностью и может рассматриваться как перспективное решение для масштабируемых систем кибербезопасности в IoT.

Ключевые слова: федеративное обучение, обнаружение вторжений, IoT, многоуровневая агрегация, CICIoT2023, FedAvg

Для цитирования: Турсынбек Е.Н., Албанбай Н., Дженоури Д. (2026). Оптимизация коммуникационных затрат в двухуровневом федеративном обучении для обнаружения кибератак в IoT-сетях // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 27. Стр. 529–544. https://doi.org/10.54309/IJICT.2026.27.3.029. (На русс.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Введение.

Интернет вещей представляет собой одну из наиболее динамично развивающихся технологических парадигм, объединяющую миллиарды распределённых устройств, сенсоров и встроенных систем в единую сетевую инфраструктуру (Vashisth et al., 2025). Непрерывный рост числа подключённых устройств порождает принципиально новую задачу: как обучать модели машинного обучения на данных, которые по своей природе распределены между тысячами узлов

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

и не могут быть централизованно собраны без нарушения конфиденциальности и перегрузки сети (Abuhasel et al., 2020) reliability and security become the most important concerns in IIoT. This introduces many familiar and ever-increasing risks associated with the industrial system. The IIoT devices can be vulnerable to vast array of viruses, threats, and attacks. Therefore, an efficient protection strategy is required to ensure that the millions of IIoT devices are safe from these risks. However, resource constraint IIoT devices have not been designed to have effective security features. Due to this, in recent years, cloud, fog, and edge-based IIoT has received great attention in the research community. The computationally intensive tasks such as security, data analytics, decision making, and reporting are performed at the cloud or fog using a powerful computing infrastructure. The data security of the IIoT device has been provided by employing improved Rivest-ShamirAdelman (RSA. Федеративное обучение (FL) предлагает решение этой задачи: каждое устройство обучает модель локально на собственных данных, а на центральный сервер передаются лишь параметры модели. Такой подход позволяет строить глобальную модель без прямого доступа к локальным данным, что делает его особенно привлекательным для масштабируемых распределённых IoT-систем (Albanbay et al., 2025).

Тем не менее классическая одноуровневая схема федеративного обучения, реализованная в алгоритме FedAvg, сопряжена с существенной коммуникационной нагрузкой. В каждом раунде обучения все клиентские устройства передают полный вектор параметров модели непосредственно на глобальный сервер, после чего получают обновлённую модель обратно (Campos et al., 2022) most ML-enabled IDS approaches use centralized approaches where IoT devices share their data with data centers for further analysis. To mitigate privacy concerns associated with centralized approaches, in recent years the use of Federated Learning (FL. Для решения этой проблемы в литературе предложены многоуровневые архитектуры FL, в которых между устройствами и глобальным сервером располагаются промежуточные узлы-шлюзы, выполняющие локальную агрегацию обновлений своей группы. Такой подход позволяет сократить число передач на глобальный сервер: вместо каждого устройства обновление отправляет лишь шлюз группы. Однако недостатком одноуровневого FedAvg остаётся то, что коммуникационная нагрузка возрастает линейно с числом устройств, шлюзы же передают обновления на глобальный сервер в каждом раунде вне зависимости от того, насколько значимо изменилась модель.

Проблема минимизации коммуникационных затрат в FL является одной из центральных в текущих исследованиях (Donta et al., 2024). Основополагающий алгоритм FedAvg (McMahan et al., 2023) which in turn can greatly improve the user experience on the device. For example, language models can improve speech recognition and text entry, and image models can automatically select good photos. However, this rich data is often privacy sensitive, large in quantity, or both, which may preclude logging to the data center and training there using conventional approaches. We advocate an alternative that leaves the training data distributed on the mobile devices, and learns a shared model by aggregating locally-computed

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

updates. We term this decentralized approach Federated Learning. We present a practical method for the federated learning of deep networks based on iterative model averaging, and conduct an extensive empirical evaluation, considering five different model architectures and four datasets. These experiments demonstrate the approach is robust to the unbalanced and non-IID data distributions that are a defining characteristic of this setting. Communication costs are the principal constraint, and we show a reduction in required communication rounds by 10–100x as compared to synchronized stochastic gradient descent.»,»DOI»:»10.48550/arXiv.1602.05629»,»note»:»arXiv:1602.05629 [cs]»,»number»:»arXiv:1602.05629»,»publisher»:»arXiv»,»source»:»arXiv.org»,»title»:»Communication-Efficient Learning of Deep Networks from Decentralized Data»,»URL»:»http://arxiv.org/abs/1602.05629»,»author»:[{«family»:»McMahan»,»given»:»H. Brendan»},{«family»:»Moore»,»given»:»Eider»},{«family»:»Ramage»,»given»:»Daniel»},{«family»:»Hampson»,»given»:»Seth»},{«family»:»Arcas»,»given»:»Blaise Agüera»,»dropping-particle»:»y»}],»accessed»:{«date-parts»:[[«2026»,4,15]]},»issued»:{«date-parts»:[[«2023»,1,26]]}}}],»schema»:»https://github.com/citation-style-language/schema/raw/master/csl-citation.json»} заложил базу для распределенного обучения, однако его высокая чувствительность к пропускной способности сети стимулировала поиск альтернативных архитектур. Современные исследования в области федеративного обучения (FL) сосредоточены на преодолении ограничений пропускной способности сетей за счет оптимизации иерархических структур. Идея промежуточных агрегирующих узлов между устройствами и сервером (архитектура Client-Edge-Cloud) также встречается в литературе (Simos et al., 2022; Zhang et al., 2024), однако её применение к задаче обеспечения безопасности IoT-сетей остаётся малоизученным. Ряд авторов(Gusain et al., 2025; Ribero et al., 2020)»container-title»:»International Journal on Computational Modelling Applications»,»DOI»:»10.63503/j.ijcma.2025.115»,»ISSN»:»3048-8516»,»issue»:»2»,»language»:»en»,»license»:»Copyright (с предлагает использовать методы адаптивного изменения частоты синхронизации в зависимости от состояния беспроводных каналов связи, что теоретически повышает производительность в системах промышленного интернета вещей (IIoT). Также значительное внимание уделяется алгоритмам сжатия моделей и квантования весов Nariman & Hamarashid, 2025)4,15]]}}}],»schema»:»https://github.com/citation-style-language/schema/raw/master/csl-citation.json»} для минимизации сетевой нагрузки. Тем не менее, несмотря на развитие методов адаптивной агрегации и энергоэффективных протоколов, в существующих работах все еще недостаточно изучен вопрос установления жесткого баланса между предельно редкой глобальной синхронизацией и сохранением точности классификации в крупномасштабных IoT-сетях (Barona et al., 2025).

Несмотря на активное развитие иерархических моделей, в научной литературе всё ещё не хватает работ, предлагающих прямое сравнение классических и многоуровневых систем с позиции радикальной экономии сетевых ресурсов и их влияния на качество обнаружения киберугроз (Tursynbek et al., 2026). Большинство существующих подходов опирается на условия с интенсивным

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

обменом данными между локальными узлами и облаком, что не позволяет объективно оценить потенциал федеративного обучения в сетях с низкой пропускной способностью. Датасет CICIoT2023 широко используется для обнаружения кибератак в IoT, однако его применение в сочетании с двухуровневой иерархической архитектурой FL, а также прямое количественное сравнение с классической одноуровневой схемой, ранее не рассматривались. В данной работе впервые для этого датасета реализована и экспериментально исследована подобная архитектура с использованием разработанной модели 1D-CNN (56 396 параметров), включая прямое сравнение с одноуровневой схемой. В результате показано, что снижение частоты глобальной синхронизации до одного раза на семь локальных раундов не приводит к статистически значимой потере точности (разница составляет 0,05 процентного пункта) при сокращении объёма передаваемого на сервер трафика более чем в 20 раз.

Материалы и методы.

Экспериментальное исследование базируется на использовании набора данных CICIoT2023 (Neto et al., 2023). На текущий момент этот датасет является одним из самых новых и распространенных ресурсов для анализа кибератак в среде Интернета вещей. Он был разработан Канадским институтом кибербезопасности (UNB) на основе активности 105 реальных устройств и охватывает 33 типа атак. В нашей работе данные были распределены между 20 имитируемыми узлами, при этом на каждое устройство приходится по 54 000 записей. Таким образом, общий объем выборки составил менее 1,1 миллиона образцов. Каждая запись в массиве содержит 39 информативных признаков, описывающих характеристики трафика, и одну целевую метку класса. Датасет был предварительно сбалансирован и перемешан, после чего разделён между 20 имитируемыми устройствами с помощью StratifiedKFold (n_splits=20, shuffle=True). В результате каждое из 20 устройств получило ровно 54 000 записей, из них поровну каждого из 12 классов (по 4500 записей на класс). Таким образом, все 20 устройств видят все 12 типов атак в одинаковой пропорции, и распределение данных между устройствами является однородным (IID). Каждая запись обрабатывалась как независимый пример, без учёта последовательности предыдущих записей.

Процесс подготовки данных был разделен на несколько этапов для обеспечения корректного обучения нейронной сети. Первоначально была проведена глубокая очистка выборки, в ходе которой удалялись дубликаты и записи с пустыми или некорректными значениями, что позволило исключить вычислительные ошибки при расчете градиентов. Далее была выполнена нормализация всех признаков с помощью метода Min-Max Scaling, в результате чего значения были приведены к диапазону [0, 1]. Это обеспечило единообразие данных и предотвратило доминирование отдельных параметров при извлечении признаков сверточными слоями модели.

Для классификации сетевого трафика была разработана модель на базе одномерной сверточной нейронной сети (1D-CNN). Данная архитектура выбрана благодаря высокой эффективности при обработке структурированных векто-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ров данных и низким вычислительным затратам на конечных IoT-устройствах. В Таблице 1 сведены основные параметры архитектуры сети и конфигурация процесса обучения.

Таблица 1. Параметры архитектуры CNN и конфигурация обучения.

Компонент / Параметр	Значение
Input Shape	(None, 39, 1)
Reshape	(39, 1)
Conv1D (Layer 1)	64 filters, kernel 3
BatchNorm + LeakyReLU	—
Dropout	0.2
Conv1D (Layer 2)	64 filters, kernel 3
BatchNorm + LeakyReLU	—
MaxPooling1D (1)	Pool size 2
Conv1D (Layer 3)	128 filters, kernel 3
BatchNorm + LeakyReLU	—
Dropout	0.3
MaxPooling1D (2)	Pool size 2
GlobalAveragePooling1D	—
Dense	128, LeakyReLU
Dropout	0.3
Output Layer	12 units, Softmax
Batch Size	1024
Learning Rate	0.001
Local Epochs	10
Optimizer	Adam
Параметры	56 396 (55 884 обучаемых + 512 BatchNorm)

Использование метода CNN в данной работе обусловлено необходимостью автоматического извлечения признаков из сырого сетевого трафика. Слои одномерной свертки позволяют выявлять корреляции между 39 входными признаками, в то время как слои подвыборки (Max Pooling) снижают размерность данных и вычислительную нагрузку. Для повышения обобщающей способности модели и предотвращения переобучения применен слой Dropout. Обучение проводится с использованием оптимизатора Adam при скорости обучения 0.001. Выбранный размер батча 1024 и проведение 10 локальных эпох обеспечивают стабильную сходимость модели на каждом из узлов перед этапом федеративной агрегации весов.

Алгоритм Federated Averaging (FedAvg) был выбран в качестве метода агрегации благодаря его способности минимизировать сетевой трафик при сохранении конфиденциальности данных. Суть алгоритма заключается в итеративном процессе, при котором каждый из 20 узлов выполняет локальное обучение, после чего передает на сервер только обновленные веса модели. Сервер вычисляет новую глобальную модель путем взвешенного усреднения полученных па-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

раметров, где вклад каждого участника прямо пропорционален количеству его локальных примеров. Математически процедура агрегации на раунде $t+1$ описывается следующей формулой:

$$w^{(t+1)} = \sum_{k=1}^{K} \frac{n_k}{\sum_{j=1}^{K} n_j} w_k^{(t)} \quad (1)$$

где $w_k^{(t)}$ — параметры локальной модели *k* -го клиента после *t* -го раунда обучения; n_k — число локальных образцов на k -м устройстве; *K* — количество узлов, участвующих в федеративном обучении. Отметим, что величина $w_k^{(t)}$ представляет собой не «сырые» веса, а результат *E* локальных эпох обучения клиента *k*, выполненных от текущего значения глобальной модели. Данный подход позволяет модели эффективно аккумулировать знания о паттернах кибератак со всех сегментов сети, не требуя централизации сырого трафика.

Для оценки эффективности распределенного обучения были разработаны и протестированы две конфигурации сетевой топологии. В одноуровневой архитектуре (Single-tier) все 20 IoT-устройств напрямую взаимодействуют с центральным сервером. В этой схеме каждый раунд обучения требует передачи параметров от каждого узла, что при 21 раунде агрегации генерирует суммарно 420 коммуникационных транзакций. Такая структура создает значительную нагрузку на магистральные каналы связи и центральный агрегатор, что может быть критично для масштабируемых сетей.

В противовес этому, предложенная двухуровневая (иерархическая) архитектура вводит промежуточный слой агрегации, распределяя 20 устройств по 5 группам (по 4 узла в каждой). Основной объем обмена весами происходит локально внутри групп, в то время как связь с глобальным сервером осуществляется дискретно — один раз на семь локальных циклов. Данный подход позволяет сократить количество глобальных транзакций до 15, что в 28 раз снижает нагрузку на центральный сервер по сравнению с одноуровневой моделью. Это обеспечивает экономию ресурсов пропускной способности без потери репрезентативности данных, так как глобальная модель по-прежнему получает агрегированные знания от всех сегментов сети.

На Рисунке 1 представлена разработанная двухуровневая архитектура федеративного обучения, адаптированная для распределенных IoT-сетей. В отличие от стандартных подходов, данная топология включает промежуточный уровень обработки данных.

- *Глобальный сервер:* располагается в верхней части иерархии и отвечает за финальную агрегацию параметров модели. Он координирует работу всей системы, но взаимодействует только с пограничными шлюзами, что минимизирует количество прямых соединений.
- *Пограничные шлюзы:* выступают в роли промежуточных агрегаторов. Каждый шлюз объединяет группу клиентов (устройств), собирает их

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

локально обученные веса и выполняет предварительное усреднение. Это позволяет проводить интенсивный обмен данными внутри локальных сегментов сети, не перегружая глобальный канал связи.

- *Клиенты:* на нижнем уровне находятся конечные устройства, которые непосредственно выполняют обучение разработанной модели CNN на своих локальных данных. После завершения циклов обучения они передают обновленные веса на соответствующий пограничный шлюз.

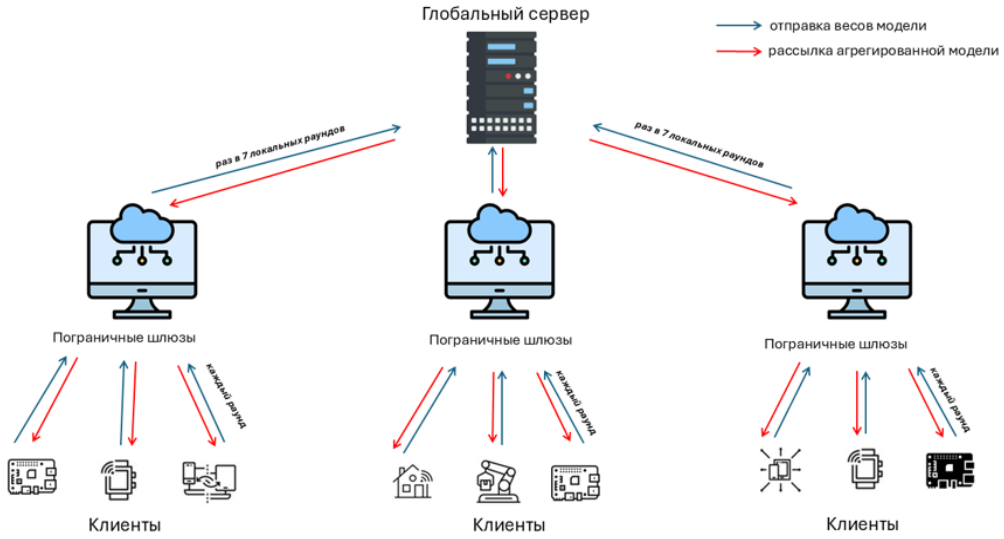


Рис. 1. Иерархическая архитектура двухуровневого FL.

Такая структура позволяет реализовать дискретную синхронизацию. В то время как обмен между клиентами и шлюзами происходит часто, связь шлюзов с глобальным сервером осуществляется только раз в несколько циклов.

Результаты и обсуждение.

В данном разделе представлены результаты сравнительного эксперимента двух архитектур федеративного обучения: одноуровневой (Single-level FL) и двухуровневой (Two-level FL). Обе конфигурации обучались на датасете CICIoT2023 с использованием модели CNN при N = 20 устройствах в течение 21 абсолютного раунда. В двухуровневой архитектуре 20 устройств объединены в 5 групп по 4 устройства; агрегация внутри групп выполняется на шлюзах, а на глобальный сервер передаются обновления лишь 3 раза (по одному разу на каждые 7 локальных раундов), что составляет 15 глобальных передач.

Сравнение качества классификации.

Таблица 2 содержит итоговые метрики качества обеих архитектур в финальном (21-м) абсолютном раунде. Как видно из таблицы, разница в точности между двумя архитектурами составляет лишь 0,05 процентного пункта, разница в Macro F1 — 0,0005, а в ROC-AUC — 0,0003. Все три показателя находятся в пределах статистической погрешности и не имеют практической значимости с

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

точки зрения качества классификации. Таким образом, переход от одноуровневой к двухуровневой архитектуре не приводит к заметному ухудшению обнаружения атак.

Таблица 2. Сравнение финальных метрик качества

Метрика	Single-level FL	Two-level FL	Разница
Accuracy	88,87%	88,82%	−0,05%
Macro F1	0,8802	0,8797	−0,0005
ROC-AUC	0,9393	0,9390	−0,0003
Loss	0,2496	0,2502	+0,0006

Визуальный анализ результатов, представленный на Рисунке 2а, демонстрирует высокую степень конвергенции (сходимости) всех 20 узлов. Несмотря на локальное обучение на различных сегментах данных, графики точности всех устройств после 2-го раунда собираются в единый плотный кластер, достигая финального плато в районе 88,82%.

Для проверки статистической устойчивости результатов обучение каждой архитектуры было повторено 5 раз с различными значениями random seed (42, 1, 7, 123, 2024). Для двухуровневой архитектуры получено среднее значение accuracy 88,799% ($\sigma = \pm 0{,}026$ п.п.), для одноуровневой — 88,818% ($\sigma = \pm 0{,}023$ п.п.). Разброс между запусками с разными seed (±0,02–0,03 п.п.) сопоставим с разницей между архитектурами, что говорит о статистической неразличимости двух схем по точности при IID-распределении данных.

Таблица 3. Статистика по 5 запускам с разными random seed

Архитектура	Mean accuracy	Std	n
Two-level FL	88,799%	±0,026 п.п.	5
One-level FL	88,818%	±0,023 п.п.	5

Дополнительно проведено сравнение с двумя альтернативными подходами (Таблица 4). Централизованное обучение (все данные на одном сервере, без федеративной схемы) служит верхней границей качества: accuracy составила 88,94%, то есть разрыв с федеративными схемами не превышает 0,14 процентного пункта. Также исследовано 8-битное квантование весов при передаче между устройствами, шлюзами и глобальным сервером (симметричное линейное квантование per-tensor): оно обеспечивает 4-кратное сокращение объёма передаваемых данных при точности, практически идентичной неквантованному варианту (88,82% в обоих случаях), и ошибке округления порядка $4{,}6 \times 10^{-6}$ (MSE).

На Рисунке 2б показана динамика функции потерь. Резкое снижение Loss-показателей в начальный период и отсутствие значимых скачков (осцилляций) в моменты глобальной агрегации (каждые 7 раундов) подтверждают стабильность предложенного иерархического алгоритма. Это доказывает, что промежуточное усреднение на уровне шлюзов эффективно стабилизирует веса модели перед их передачей на глобальный сервер.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Таблица 4. Сравнение с альтернативными подходами

Вариант	Accuracy	Loss
Centralized (upper bound)	88,94%	0,2475
Two-level FL, mean (5 seeds)	88,80%	—
One-level FL, mean (5 seeds)	88,82%	—
Two-level FL + int8 quantization	88,82%	0,2502

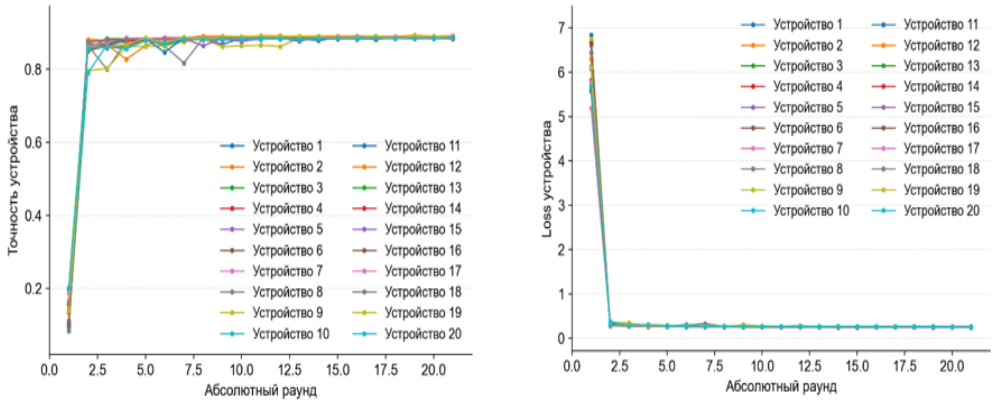


Рис. 2. Динамика показателей обучения для 20 IoT-устройств в двухуровневой архитектуре FL: (а) точность классификации; (б) значения функции потерь.

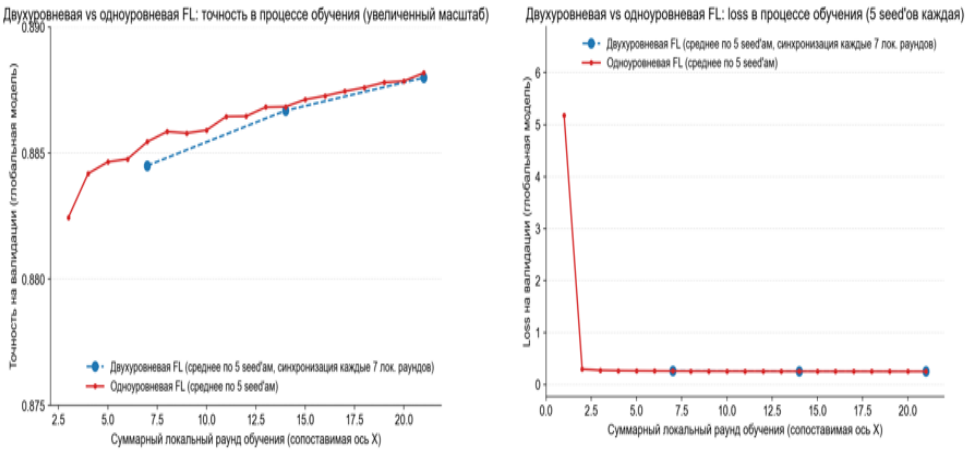


Рис. 3. Сравнение динамики обучения одноуровневой и двухуровневой архитектур FL по метрикам: (а) Accuracy; (б) Loss, усреднённым по 5 запускам с разными random seed.

Сравнение коммуникационных затрат.

Принципиальное различие между двумя архитектурами проявляется в числе глобальных передач параметров модели на сервер. В таблице 5 представлено сравнение коммуникационных затрат. Двухуровневая архитектура сокращает число глобальных передач с 420 до 15, то есть в 28 раз. Это достигается за счёт двух механизмов: во-первых, вместо 20 устройств на глобальный сервер

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

выходят лишь 5 шлюзов; во-вторых, глобальная синхронизация происходит не в каждом раунде, а лишь раз в 7 локальных раундов. В совокупности оба механизма обеспечивают 96,4-процентную экономию глобального трафика.

В абсолютных величинах: при размере модели 56 396 параметров (225 584 байта ≈ 0,215 МБ в формате float32) одноуровневая схема передаёт на сервер около 90,3 МБ трафика (420 передач), тогда как двухуровневая — около 3,2 МБ (15 передач), что даёт абсолютную экономию порядка 87,1 МБ.

Таблица 5. Сравнение коммуникационных затрат

Параметр	Single-level FL	Two-level FL
Устройств / шлюзов на глобальный сервер	20	5
Глобальных раундов	21	3
Итого глобальных передач	420	15
Сокращение	—	в 28 раз (−96,4%)

Динамика обучения в обеих архитектурах имеет схожий характер: на первом раунде точность составляет около 10%, что соответствует случайной инициализации модели, а уже со второго раунда происходит резкий скачок до уровня 88,2–88,8%. В одноуровневой архитектуре дальнейшее улучшение носит плавный монотонный характер: точность медленно возрастает с 88,24% (раунд 2) до 88,87% (раунд 21) на протяжении всех 21 раунда. В двухуровневой архитектуре аналогичного уровня качества достигается уже к 7-му абсолютному раунду (первый глобальный раунд: 88,44%), а к 21-му раунду модель выходит на 88,82%. Таким образом, двухуровневая архитектура обеспечивает практически эквивалентное конечное качество при значительно более редкой глобальной синхронизации.

Полученные результаты убедительно демонстрируют, что иерархическая двухуровневая архитектура FL обеспечивает выгодный компромисс между коммуникационной эффективностью и качеством модели. Потеря точности в 0,05% является пренебрежимо малой величиной в сравнении с 28-кратным сокращением нагрузки на глобальный сервер. В контексте реальных IoT-сетей, где пропускная способность каналов между конечными устройствами и облачным сервером зачастую ограничена, подобная экономия трафика имеет принципиальное значение для практического развёртывания систем федеративного обучения. Промежуточная агрегация на уровне шлюзов позволяет интенсивный локальный обмен обновлениями оставлять внутри группы, вынося на глобальный уровень лишь финальный агрегированный вектор. Это не только снижает трафик, но и повышает устойчивость системы к возможным перебоям в связи с центральным сервером.

Обсуждение масштабируемости и ограничений.

Эксперименты в данной работе проведены преимущественно на (N=20) устройствах, однако аналитическая оценка коммуникационных затрат может быть распространена на сети большего масштаба. При фиксированном размере группы число глобальных передач растёт линейно с увеличением (N) как в од-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

ноуровневой, так и в двухуровневой схеме. При объединении четырёх устройств на один шлюз коэффициент роста для двухуровневой архитектуры оказывается в 4 раза меньше. Дополнительное сокращение коммуникаций достигается за счёт выполнения глобальной синхронизации один раз в 7 раундов и не зависит от числа устройств. Например, при (N=200) одноуровневая схема потребует около 4200 глобальных передач за 21 раунд, тогда как двухуровневая — порядка 150, что соответствует сокращению коммуникационной нагрузки примерно в 28 раз.

Для проверки того, сохраняется ли качество классификации при увеличении масштаба сети, дополнительно проведены эксперименты с большим числом устройств (Таблица 6). При (N=50) (размер группы — 5 устройств, число групп — 10) точность составила 88,67%, а при (N=100) (размер группы — 5 устройств, число групп — 20) — 88,41%. Незначительное изменение точности по сравнению с конфигурацией (N=20) показывает, что при IID-распределении данных увеличение числа устройств и числа групп не приводит к существенному ухудшению качества классификации. Таким образом, аналитические и экспериментальные результаты указывают на возможность масштабирования предложенной двухуровневой архитектуры при сохранении преимуществ по коммуникационной эффективности и сопоставимого качества модели. Вместе с тем вопросы дополнительной вычислительной и коммуникационной нагрузки на шлюзы, а также отказоустойчивости при дальнейшем увеличении (N), требуют отдельного исследования и рассматриваются как направление дальнейшей работы.

Таблица 6. Эмпирическая проверка масштабируемости

Устройств	Group size	Групп	Accuracy
20	4	5	88,82%
50	5	10	88,67%
100	5	20	88,41%

Таблица 7. Нагрузка на шлюз: время агрегации и разброс между устройствами

Метрика	Two-level	One-level
Время агрегации (group+global)	37,5 с (n=3)	8,8 с (n=5)
Время обучения устройства, mean	12,3 с	12,2 с
Время обучения устройства, диапазон	9,6–14,3 с	9,8–16,2 с

Проведено измерение времени агрегации на шлюзах по временным меткам логов обучения (Таблица 7). Двухуровневая схема тратит на агрегацию (групповую и глобальную) в среднем 37,5 с за цикл синхронизации против 8,8 с у одноуровневой схемы — дополнительные 28,7 с являются ожидаемой платой за второй уровень усреднения. При этом среднее время локального обучения устройства сопоставимо между схемами (12,3 с и 12,2 с соответственно), а разброс между самым быстрым и медленным устройством внутри группы объясняется шумом планировщика ОС, а не дисбалансом объёма данных, поскольку распределение данных однородно (IID).

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Заключение.

В данной работе предложена и исследована иерархическая двухуровневая архитектура федеративного обучения, адаптированная для обнаружения кибератак в распределенных сетях IoT. Эксперименты на датасете CICIoT2023 показали, что использование промежуточных шлюзов для агрегации весов позволяет достичь точности 88,82%, что практически идентично показателю стандартной одноуровневой модели (88,87%). Незначительная разница в 0,05% подтверждает, что иерархическая структура сохраняет высокую надежность классификации при значительном изменении топологии сети.

Ключевым достижением исследования стало радикальное снижение коммуникационной нагрузки на глобальный сервер: количество транзакций сократилось с 420 до 15, что эквивалентно экономии трафика на 96,4%. Анализ динамики обучения (Рис. 2а, 2б) подтвердил стабильную сходимость модели на всех 20 устройствах, доказывая эффективность промежуточного усреднения данных. Таким образом, разработанный подход является оптимальным решением для масштабируемых систем безопасности IoT, где критически важны экономия ресурсов и пропускная способность каналов связи. Ограничением данной работы остаётся отсутствие сравнения со схемами с более чем двумя уровнями иерархии и с адаптивными методами частоты синхронизации — включение таких сравнений обозначено как направление дальнейшей работы.

REFERENCES

Abuhasel, K. A., & Khan, M. A. (2020). A Secure Industrial Internet of Things (IIoT) Framework for Resource Management in Smart Manufacturing // IEEE Access. — Vol. 8. — Pp. 117354–117364. DOI: https://doi.org/10.1109/ACCESS.2020.3004711 [in Eng].

Albanbay, N., Tursynbek, Y., Graffi, K., Uskenbayeva, R., Kalpeyeva, Z., Abilkaiyr, Z., & Ayapov, Y. (2025). Federated Learning-Based Intrusion Detection in IoT Networks: Performance Evaluation and Data Scaling Study // Journal of Sensor and Actuator Networks. — Vol. 14. — No. 4. — Pp. 78. DOI: https://doi.org/10.3390/jsan14040078 [in Eng].

Barona López, L. I., & Borja Saltos, T. (2025). Heterogeneity Challenges of Federated Learning for Future Wireless Communication Networks // Journal of Sensor and Actuator Networks. — Vol. 14. — No. 2. — Pp. 37. DOI: https://doi.org/10.3390/jsan14020037 [in Eng].

Campos, E. M., Saura, P. F., González-Vidal, A., Hernández-Ramos, J. L., Bernabé, J. B., Baldini, G., & Skarmeta, A. (2022). Evaluating Federated Learning for intrusion detection in Internet of Things: Review and challenges // Computer Networks. — Pp. 203. Article 108661. DOI: https://doi.org/10.1016/j.comnet.2021.108661 [in Eng].

Nariman, G. S., & Hamarashid, H. K. (2025). Communication overhead reduction in federated learning: a review // International Journal of Data Science and Analytics. — Vol. 19. — No. 2. — Pp. 185–216. [in Eng].

Donta, P. K., Hazra, A., & Lovén, L. (Eds). (2024). Learning Techniques for the Internet of Things. Springer Nature Switzerland. DOI: https://doi.org/10.1007/978-3-031-50514-0 [in Eng].

Gusain, N., & Sharma, H. (2025). Communication-Efficient Federated Learning in Industrial IoT — A Framework for Real-Time Threat Detection and Secure Device Coordination // International Journal on Computational Modelling Applications. — Vol. 2. — No. 2. — Pp. 18–29. DOI: https://doi.org/10.63503/j.ijcma.2025.115 [in Eng].

McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. y. (2023). Communication-Efficient Learning of Deep Networks from Decentralized Data (arXiv:1602.05629). arXiv. DOI: https://doi.org/10.48550/arXiv.1602.05629 [in Eng].

Neto, E. C. P., Dadkhah, S., Ferreira, R., Zohourian, A., Lu, R., & Ghorbani, A. A. (2023). CICIoT2023: A Real-Time Dataset and Benchmark for Large-Scale Attacks in IoT Environment. Sensors. — Vol. 23. — No. 13. — Pp. 5941. DOI: https://doi.org/10.3390/s23135941 [in Eng].

Ribero, M., & Vikalo, H. (2020). Communication-Efficient Federated Learning via Optimal Client

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Sampling (arXiv:2007.15197). arXiv. DOI: https://doi.org/10.48550/arXiv.2007.15197 [in Eng].

Simos, M., Bouzinis, P. S., Diamantoulakis, P. D., Sarigiannidis, P., & Karagiannidis, G. K. (2022). Hierarchical Federated Learning for the Next Generation IoT. 2022 18th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob). — Pp. 198–203. DOI: https://doi.org/10.1109/WiMob55322.2022.9941355 [in Eng].

Tursynbek, Y., Albanbay, N., Djenouri, D., Latif, S., Akhmediyarova, A., Alibiyeva, Z., Alimkulova, J., & Oralbekova, D. (2026). Secure Federated Intrusion Detection for Resource-Constrained IoT Devices Using Lightweight Cryptography: A Hardware-Validated Study // Future Internet. — Vol. 18. — No. 6. — Pp. 306. DOI: https://doi.org/10.3390/fi18060306[in Eng].

Vashisth, S., & Goyal, A. (2025). A Survey of Federated Learning for IoT: Addressing Resource Constraints and Heterogeneous Challenges // Informatica. — Vol. 49. — No. 17. DOI: https://doi.org/10.31449/inf.v49i17.7707 [in Eng].

Zhang, H., Du, J., Hou, X., Jiang, C., Wang, J., & Niyato, D. (2024). Convergence Analysis of Hierarchical Split Federated Learning. — Pp. 817. DOI: https://doi.org/10.1109/GLOBECOM52923.2024.10901714 [in Eng].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 7. Is.3. Number 27 (2026). Pp. 545–563
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2026.27.3.030

A COMPARATIVE EVALUATION OF MACHINE LEARNING MODELS FOR ACTIVE DIRECTORY THREAT DETECTION: FEATURE-LEAKAGE CONTROL AND CROSS-SCENARIO VALIDATION ON WINDOWS SECURITY EVENT LOGS

A. Tastemirov*, S.B. Mukhanov, S.T. Amanzholova, B. Aldosh
Astana IT University, Astana, Kazakhstan.
E-mail: 255146@astanait.edu.kz

Azamat Tastemirov — student/researcher, School of Cybersecurity, Astana IT University, Astana, Kazakhstan
E-mail: 255146@astanait.edu.kz, https://orcid.org/0009-0009-6215-3081;
S.B. Mukhanov — PhD, Assistant Professor, School of Cybersecurity, Astana IT University, Astana, Kazakhstan
https://orcid.org/0000-0001-8761-4272;
S.T. Amanzholova — Candidate of Technical Sciences, Associate Professor, School of Cybersecurity, Astana IT University, Astana, Kazakhstan
https://orcid.org/0000-0002-6779-9393;
Balziya Aldosh — student/researcher, School of Cybersecurity, Astana IT University, Astana, Kazakhstan
https://orcid.org/0000-0001-7358-5490.

© A. Tastemirov, S.B. Mukhanov, S.T. Amanzholova, B. Aldosh

Abstract. Windows Active Directory (AD) is the backbone of authentication and authorization in most enterprise networks and is therefore a priority target for credential abuse, lateral movement, and privilege escalation. Windows Security Event Logs record detailed structured telemetry of these actions, but rule-based detection remains brittle: it generates excessive false positives and overlooks patterns that span multiple log fields. In this work five supervised machine-learning model families — Logistic Regression, linear Support Vector Machine, Random Forest, Gradient Boosting, and XGBoost — are compared on 23,268 AD-related Windows events drawn from ten public adversary-emulation captures (OTRF Security-Datasets, 2024). All models share a common preprocessing pipeline that explicitly audits every log field for session-level leakage and discards 108 of the 235 raw columns before encoding. Each model is evaluated under three complementary protocols: a 60/40 stratified holdout, stratified 5-fold cross-validation, and leave-one-scenario-out (LOSO), in

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

which the model is tested on an entire adversary-emulation capture that was unseen during training. Under the standard protocols all five families look strong (CV5 ROC-AUC > 0.85 and PR-AUC > 0.95), yet under LOSO the same models collapse to ROC-AUC in the range 0.15–0.45 and Matthews correlation near zero. This protocol gap, obtained on identical features and hyperparameters, quantifies the dataset-specific signal that any single-split evaluation hides. We discuss its implications for dataset bias, interpretability, inference cost, SOC deployment, and the sequence-level attack chains that an event-level classifier cannot capture by construction.

Keywords: Active Directory, Windows Security Logs, threat detection, intrusion detection, machine learning, data leakage, leave-one-scenario-out, XGBoost, Random Forest, SIEM

For citation: Tastemirov A., Amanzholova S.T., Aldosh B., Mukhanov S.B. (2026). A comparative evaluation of machine learning models for active directory threat detection: feature-leakage control and cross-scenario validation on windows security event logs // International journal of information and communication technologies. Vol. 7. No. 27. Pp. 545–563. https://doi.org/10.54309/IJICT.2026.27.3.030. (In Eng.).

Conflict of interest: The authors declare that there is no conflict of interest.

ACTIVE DIRECTORY ЖҮЙЕСІНДЕ ҚАУІПТЕРДІ АНЫҚТАУ ҮШІН МАШИНАЛЫҚ ОҚЫТУ МОДЕЛЬДЕРІН САЛЫСТЫРМАЛЫ БАҒАЛАУ: WINDOWS ҚАУІПСІЗДІК ОҚИҒАЛАРЫ ЖУРНАЛДАРЫНДА БЕЛГІ ЖЫЛЫСТАУЫН БАҚЫЛАУ ЖӘНЕ КРОСС-СЦЕНАРЛЫҚ ВАЛИДАЦИЯ

А. Тастемиров*, С.Т. Аманжолова, Б. Алдош, С.Б. Муханов
Astana IT University, Астана, Қазақстан.
E-mail: 255146@astanait.edu.kz

Тастемиров Азамат — студент/зерттеуші, Киберқауіпсіздік мектебі, Astana IT University, Астана, Қазақстан
E-mail: 255146@astanait.edu.kz, https://orcid.org/0009-0009-6215-3081
Муханов С.Б. — PhD, ассистент-профессор, Киберқауіпсіздік мектебі, Astana IT University, Астана, Қазақстан
https://orcid.org/0000-0001-8761-4272;
Аманжолова С.Т. — техника ғылымдарының кандидаты, қауымдастырылған профессор, Киберқауіпсіздік мектебі, Astana IT University, Астана, Қазақстан https://orcid.org/0000-0002-6779-9393;
Алдош Балзия — студент/зерттеуші, Киберқауіпсіздік мектебі, Astana IT University, Астана, Қазақстан
https://orcid.org/0000-0001-7358-5490.

© А. Тастемиров, С.Б. Муханов, С.Т. Аманжолова, Б. Алдош

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Аннотация. Active Directory (AD) көптеген корпоративтік желілерде аутентификация және авторизация үдерістерінің негізі болып табылады және сол себепті тіркелгі деректерін қолдану, желі ішінде көлденең қозғалыс және артықшылықтарды арттыру шабуылдарының басым мақсатына айналады. Windows қауіпсіздік оқиғалары журналдары осы әрекеттер туралы егжей-тегжейлі құрылымдалған телеметрияны тіркейді, бірақ ережеге негізделген анықтау әлсіз болып қала береді: ол жалған оң нәтижелерді шамадан тыс көп шығарады және бірнеше журнал өрісін қамтитын заңдылықтарды таба алмайды.Бұл жұмыста қадағаланатын машиналық оқытудың бес отбасы — логистикалық регрессия, сызықтық тірек векторлар әдісі (SVM), кездейсоқ орман, градиенттік бустинг және XGBoost — OTRF Security-Datasets жобасынан алынған он ашық шабуыл имитациясы жазбаларынан құралған 23 268 AD-ға қатысты Windows оқиғасында салыстырылады. Барлық модельдер бірыңғай алдын ала өңдеу құбыры бойынша оқытылады; онда әрбір журнал өрісі сессия деңгейіндегі ағуға тексеріліп, 235 бастапқы бағанның 108-і кодтауға дейін алынып тасталады. Әрбір модель үш өзара толықтырушы хаттама бойынша бағаланады: 60/40 стратификацияланған бөлу (H), стратификацияланған 5-қатпарлы кросс-валидация (CV5) және оқыту кезінде көрмеген бүтін бір шабуыл имитациясы жазбасында тексеретін leave-one-scenario-out (LOSO).Стандартты хаттамалар бойынша бес отбасы да күшті болып көрінеді (CV5-те ROC-AUC 0,85-тен, PR-AUC 0,95-тен асады), алайда LOSO хаттамасы бойынша сол модельдер ROC-AUC-тың 0,15–0,45 аралығына және Мэтьюс корреляциясының нөлге жақын мәніне дейін төмендейді. Бірдей белгілермен және гиперпараметрлермен алынған осы хаттамалар арасындағы алшақтық бір бөлудегі бағалау жасыратын жиынтыққа тән сигналды сан түрінде көрсетеді. Нәтиженің датасет жылыстауына, интерпретациялануға, инференс құнына, SOC-ке енгізуге және жеке оқиғалар классификаторы құрылымдық түрде қамти алмайтын көпқадамды шабуыл тізбектерін талдауға әсері талқыланады.

Түйінді сөздер: Active Directory, Windows қауіпсіздік журналдары, қауіптерді анықтау, шабуылдарды анықтау, машиналық оқыту, деректер жылыстауы, leave-one-scenario-out, XGBoost, Random Forest, SIEM

Дәйексөздер үшін: Тастемиров А., Аманжолова С.Т., Алдош Б., Муханов С.Б. (2026). Active Directory жүйесінде қауіптерді анықтау үшін машиналық оқыту модельдерін салыстырмалы бағалау: Windows қауіпсіздік оқиғалары журналдарында белгі жылыстауын бақылау және кросс-сценарлық валидация // Халықаралық ақпараттық және коммуникациялық технологиялар журналы. Т. 7. No 27. Б. 545–563. https://doi.org/10.54309/IJICT.2026.27.3.030. (Ағыл. тіл.).

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

СРАВНИТЕЛЬНАЯ ОЦЕНКА МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОБНАРУЖЕНИЯ УГРОЗ В ACTIVE DIRECTORY: КОНТРОЛЬ УТЕЧКИ ПРИЗНАКОВ И КРОСС-СЦЕНАРНАЯ ВАЛИДАЦИЯ НА ЖУРНАЛАХ БЕЗОПАСНОСТИ WINDOWS

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

А. Тастемиров*, С.Т. Аманжолова, Б. Алдош, С.Б. Муханов
Astana IT University, Астана, Казахстан.
E-mail: 255146@astanait.edu.kz

Тастемиров Азамат — студент/исследователь, Школа кибербезопасности, Astana IT University, Астана, Казахстан
E-mail: 255146@astanait.edu.kz, https://orcid.org/0009-0009-6215-3081;
Муханов С.Б. — PhD, ассистент-профессор, Школа кибербезопасности, Astana IT University, Астана, Казахстан
https://orcid.org/0000-0001-8761-4272;
Аманжолова С.Т. — кандидат технических наук, ассоциированный профессор, Astana IT University, Астана, Казахстан
https://orcid.org/0000-0002-6779-9393;
Алдош Балзия — студент/исследователь, Школа кибербезопасности, Astana IT University, Астана, Казахстан
https://orcid.org/0000-0001-7358-5490.

© А. Тастемиров, С.Б. Муханов, С.Т. Аманжолова, Б. Алдош

Аннотация. Active Directory (AD) является ключевым компонентом большинства корпоративных сетей, обеспечивающим процессы аутентификации и авторизации, и поэтому представляет собой приоритетную мишень для компрометации учётных данных, бокового перемещения и повышения привилегий. Журналы событий безопасности Windows содержат подробные структурированные данные о таких действиях, однако сигнатурно-корреляционные правила по-прежнему уязвимы: они дают избыток ложных срабатываний и не улавливают закономерности, охватывающие сразу несколько полей журнала. В данной работе пять семейств моделей машинного обучения с учителем — логистическая регрессия, линейная модель метода опорных векторов (SVM), случайный лес, градиентный бустинг и XGBoost — сравниваются на 23 268 событиях Windows, связанных с AD и собранных из десяти открытых сценариев имитации атаки (OTRF Security-Datasets). Все модели обучаются по единому пайплайну предобработки, который явно проверяет каждое поле журнала на утечку уровня сессии и отбрасывает 108 из 235 исходных столбцов до кодирования. Каждая модель оценивается по трём взаимодополняющим протоколам: стратифицированное 60/40 разбиение (H), стратифицированная 5-кратная кросс-валидация (CV5) и leave-one-scenario-out (LOSO), в котором модель тестируется на целом сценарии имитации атаки, не встречавшемся при обучении. Под стандартными протоколами все пять семейств выглядят сильными (ROC-AUC более 0,85 и PR-AUC более 0,95 в CV5), однако под LOSO те же модели деградируют до ROC-AUC в диапазоне 0,15–0,45 и коэффициента корреляции Мэтьюса, близкого к нулю. Полученный на одинаковых признаках и гиперпараметрах разрыв между протоколами количественно характеризует специфичный для датасета сигнал, который лю-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

бое одиночное разбиение скрывает. Обсуждаются следствия этого разрыва для смещения датасета, интерпретируемости, стоимости инференса, развёртывания в SOC и анализа многошаговых цепочек атак, которые классификатор отдельных событий по построению не может охватить.

Ключевые слова: Active Directory, журналы безопасности Windows, обнаружение угроз, обнаружение вторжений, машинное обучение, утечка данных, leave-one-scenario-out, XGBoost, Random Forest, SIEM

Для цитирования: Тастемиров А., Аманжолова С.Т., Алдош Б., Муханов С.Б. (2026). Сравнительная оценка моделей машинного обучения для обнаружения угроз в Active Directory: контроль утечки признаков и кросс-сценарная валидация на журналах безопасности Windows // Международный журнал информационных и коммуникационных технологий. Т. 7. No. 27. Стр. 545–563. https://doi.org/10.54309/IJICT.2026.27.3.030. (На англ.).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Introduction.

Active Directory (AD) is the central pillar of identity management in enterprise Windows environments. A compromise of AD typically grants the attacker privilege escalation, lateral movement, and persistence. Attack vectors include password spraying and pass-the-hash, Kerberoasting, DCSync, Golden Ticket attacks, and abuse of privileged group membership. Windows Security Event Logs record extensive telemetry on these behaviours, including authentication attempts, Kerberos ticket validation and service-ticket requests, and privileged-group modifications (Microsoft, 2024). Event IDs frequently used for detection include 4624/4625 (successful/failed logon), 4768/4769 (Kerberos AS-REQ/TGS-REQ), 4672 (special privileges assigned to a new logon), and 4728/4732 (group-membership changes). However, most enterprise detection pipelines remain rule-correlation based. Such rules are effective for known signatures but often fail against stealthy or low-and-slow attacks, and typically generate excessive false positives, contributing to alert fatigue in Security Operations Centres (SOCs). This motivates the use of machine-learning approaches, which can model statistical patterns and interactions between many log fields (Rakine et al., 2025; Abdallah et al., 2022; Mubarakova et al., 2022). A significant body of prior work explores ML-based intrusion detection. Surveys highlight the effectiveness of supervised learning on labelled datasets (Rakine et al., 2025; Abdallah et al., 2022; Ogunbadejo et al., 2025) more recent work focuses on log-based anomaly detection and deep-learning approaches (Landauer et al., 2023). Earlier work investigated cost-sensitive modelling and the economic impact of false positives in security analytics (Kaushik et al., 2025). Two methodological concerns run through these studies and motivate the present work. First, log-based ML is notoriously vulnerable to data leakage: a column that is harmless in isolation can carry a near-perfect trace of the recording session or of the attacker's tool, which inflates in-distribution scores but does not transfer (Landauer et al., 2023; Mohale et al., 2025). Second, operational AD

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

attacks are multi-step chains rather than isolated events, so an event-level classifier is at best a scoring primitive for a downstream sequence model, not a complete detector. In practice, applying ML to security monitoring remains challenging because of dataset shift, label noise, and the "closed-world" assumption, where training and deployment environments differ significantly (Ogunbadejo et al., 2025). This is particularly relevant for AD logs, where new hosts, users, and administrative patterns appear continuously.

Problem Statement

Although machine learning has been widely applied to intrusion detection, a practical gap still remains in comparative evaluations specifically for Active Directory threat detection using Windows Security Event Logs. Most existing studies focus on network traffic rather than AD logs, rely on deep-learning models that are difficult to deploy in real SOC pipelines, do not evaluate multiple model families under a unified preprocessing and encoding setup, and often overlook operational constraints such as interpretability and inference cost. The need for comparative evaluation under consistent experimental settings has also been emphasised in other applied machine-learning domains (Mukhanov et al., 2024). Three further issues are often under-addressed. (i) Windows Security Logs present unique data-engineering challenges — missing values, noisy and inconsistent fields, and high-cardinality categorical attributes such as usernames and hostnames — and feature selection and representation have been shown to materially impact IDS accuracy (Alsaleh et al., 2021; Gu et al., 2021; Mukhanov et al., 2020). (ii) Reported performance is routinely obtained from a single stratified split, which conflates genuine detection with memorisation of session-level artefacts; the extent of this inflation is seldom quantified. (iii) Attacks against AD unfold as chains of interconnected actions (e.g. logon → token manipulation → service-ticket request → DCSync), whereas most published pipelines classify single events in isolation. The study below is organised around these three gaps.

Contributions

This paper addresses the above issues by (1) constructing a unified event-level dataset of 23,268 AD-related Windows Security events from ten public adversary-emulation captures of the OTRF Security-Datasets project; (2) proposing a leakage-aware preprocessing pipeline that explicitly flags and removes 108 of 235 raw log columns as session-level identifiers before any model sees the data; (3) comparing five widely used supervised machine-learning model families — Logistic Regression, linear SVM, Random Forest, Gradient Boosting, and XGBoost — under three complementary evaluation protocols (stratified holdout, stratified 5-fold cross-validation, and leave-one-scenario-out), the last of which serves as a cross-source stress test on identical features and hyperparameters; (4) quantifying the resulting protocol gap and interpreting it as empirical evidence of residual dataset bias in log-based AD threat detection; and (5) analysing feature attributions, interpretability, inference cost, and the sequence-modelling boundary conditions that constrain SOC deployment. All code, the per-capture manifest, and the stored out-of-fold predictions are released as a reproducibility package.

Threat Model and Problem Formulation

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

We posit an enterprise AD security scenario in which an attacker is attempting to gain a foothold in the environment through credential abuse, lateral movement, and privilege escalation. We assume that the attacker employs known offensive tools (e.g. Mimikatz, Impacket, Rubeus, Covenant, Empire) that leave identifiable traces in Windows Security Event Logs. Security logging is assumed to be enabled and collected centrally. In the dataset used below, the negative class consists of Discovery actions performed by comparable tooling rather than genuinely benign traffic; this choice (also reflected in the OTRF Security-Datasets labelling) stresses the classifiers against adversarial-looking normal activity rather than against pristine background noise, and is consistent with the SOC deployment setting in which trivial, cleanly benign events are already filtered out upstream. The attacker may still obfuscate activity so that single events resemble legitimate administrative tasks, which motivates the sequence-level discussion in Section VI.

Binary Classification Formulation

Each Windows Security event is represented as a feature vector

$$x_i \in R^d, \tag{1}$$

extracted from structured log fields. Each sample has a binary label:

$$Y_i \in \{0,1\}, \tag{2}$$

where y_i = 0 denotes benign activity and y_i = 1 denotes malicious or suspicious activity. Here, x_i is the feature vector of the i-th event, d is the number of extracted features, and y_i is the corresponding binary label. The goal is to learn a classifier f(x) that maps event features to a maliciousness score:

$$\hat{p}(y = 1 \mid x) = f(x), \tag{3}$$

where $\hat{p}(y=1 \mid x)$ is the predicted probability of the positive class and f(x) is the model scoring function. The classifier predicts the class as: $\hat{y} = 1$ if $\hat{p}(y=1 \mid x) \geq \tau$, otherwise 0, where $\hat{y}$ is the predicted class label and τ is the decision threshold.

Materials and methods.

This section describes the dataset, the leakage-aware preprocessing pipeline, the feature encoding, and the three evaluation protocols applied to all five model families. Unlike purely theoretical comparisons, we emphasise a unified pipeline because preprocessing choices strongly influence performance in log-based threat detection (Landauer et al., 2023), and because the standard single-split protocol has been shown to overstate real detection on this class of data.

Dataset Description

The dataset is assembled from ten open adversary-emulation captures of the OTRF Security-Datasets project (formerly Mordor) covering two ATT&CK tactics:

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Credential Access (five captures of DCSync, LSASS Memory, and LSA Secrets techniques executed with Mimikatz, Empire, Covenant, Dumpert direct syscalls, and RDP + Task Manager; labelled y=1) and Discovery (five captures of Domain Account and Domain Groups Discovery executed with Covenant, Empire, and Seatbelt over LDAP/SAMR; labelled y=0). Labels are assigned from the ATT&CK tactic of the emulated technique, not from a folder name. The negative class is thus not pristine background noise — every capture is an adversary emulation — which makes the classification task harder than the usual benign-vs-malicious framing. The raw records were exported from the Windows Security, Sysmon, and PowerShell channels and parsed into a tabular CSV with 235 structured columns, including event identifiers, authentication metadata, account and domain attributes, host-related context, and privilege-related parameters. Logs from different captures were merged into a consistent schema by aligning field names and data types across event categories. The full corpus after parsing contains 23,268 event records. The capture-level manifest, including the SHA-256 hash of every source file, is released with the paper (scenario_manifest.csv).

Dataset Statistics

Table 1 summarises dataset size and class distribution after parsing but before any leakage filtering. The imbalance is substantial — Credential Access events outnumber Discovery events roughly 5 to 1 — and is driven by the fact that credential-dumping captures emit a large number of authentication, token, and privilege events per invocation. Class imbalance is prevalent in intrusion-detection datasets and must be taken into account to reduce the risk of training skewed classifiers (Rakine et al., 2025; Abdallah et al., 2022). Throughout the paper we therefore report PR-AUC and Matthews correlation (MCC) alongside ROC-AUC, and benchmark against the always-positive baseline, which attains F1 = 0.911, PR-AUC = 0.837 and MCC = 0 under this distribution.

Class	Label	Events
Discovery (negative class)	0	3,784
Credential Access (positive class)	1	19,484
Total	—	23,268

Beyond the global imbalance, Table 1 hides a more consequential asymmetry: within each class, a few large captures dominate the event count (one Mimikatz capture alone contributes 6,982 events on the positive side; one Empire capture contributes 1,749 on the negative side). Any evaluation protocol that samples uniformly from events — including stratified splits — therefore implicitly samples the dominant captures. The leave-one-scenario-out protocol defined below is designed to probe exactly this effect.

Leakage Audit of Log Fields

Windows Security logs mix two kinds of information: fields that describe the action (EventID, LogonType, AccessMask, TargetObject) and fields that incidentally identify the recording session or the attacker's tool (ProcessGuid, Hashes, UtcTime, Record Number, host and user names that recur only inside one capture). A classifier

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

trained on the second group memorises the capture rather than the behaviour and then scores near-perfectly on an in-distribution split, which is precisely the pattern that produced the unrealistically high numbers reported in the preliminary version of this study. Before any model is fitted, every one of the 235 raw columns is therefore classified by a static rule set into six field families: (i) unique per-event identifiers (@timestamp, Hashes, ProcessGuid, Record Number, UtcTime, ThreadId, ProcessId, EventRecordID …), (ii) free-text narrative fields (Message, DetailText), (iii) absolute host- and user-specific identifiers that take distinct values across captures (Computer, SourceComputer, SubjectUserName, TargetUserName, Hostname …), (iv) raw byte or offset counters (TokenHandle, HandleId, SubjectLogonId when numeric), (v) fields with fewer than 30 observed events across the corpus, and (vi) constant columns. Families (i)–(iv) are dropped as leakage (108 columns), family (v) is dropped as statistically unreliable (48 columns), and family (vi) carries no signal (7 columns), leaving 72 behavioural fields that are then encoded with one-hot encoding (handle_unknown='ignore') into 501 columns. The retained set is dominated by action-describing fields such as EventID, LogonType, AccessList, PrivilegeList, EventType, ObjectType, GrantedAccess, and TargetObject; the full retained list is logged with the reproducibility package. All experiments in Section V use this leakage-controlled matrix.

Preprocessing Pipeline

After the leakage audit, the 72 retained log fields undergo a unified preprocessing pipeline that replaces missing categorical values with the token «missing», imputes missing numerical values using the median of observed values, and encodes categorical attributes with one-hot encoding. The encoded design matrix is 23,268 × 501. Model performance is reported with Precision, Recall, F1-score, ROC-AUC, PR-AUC, Matthews correlation (MCC), and balanced accuracy. PR-AUC and MCC are emphasised because, for the imbalanced AD setting, ROC-AUC can paint an overly optimistic picture due to the very large number of true negatives absorbed by the ROC denominator (Rakine et al., 2025).

Evaluation Protocols

Every model in this study is evaluated under three complementary protocols on identical features, identical preprocessing and identical hyperparameters:

H — stratified 60/40 holdout. A single random split preserving the global class ratio. This mirrors the most frequent protocol in prior IDS work and is retained for comparability.

CV5 — stratified 5-fold cross-validation. Every event is scored exactly once as a held-out example; performance is reported both pooled over the out-of-fold predictions and as the mean ± standard deviation across the five folds. This protocol quantifies the statistical stability of each model.

LOSO — leave-one-scenario-out. In each of the ten folds, one entire adversary-emulation capture is withheld from training and used for evaluation; the model never sees any event that belongs to the held-out capture. Because the ten captures differ in attacker tool (Mimikatz, Empire, Covenant, Dumpert, Seatbelt, RDP + Task Manager),

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

host set, and ATT&CK sub-technique (T1003.001/004/006, T1069.002, T1087.002), LOSO acts as a cross-source stress test on identical features and hyperparameters, and is the closest proxy to external validation available from a single-corpus study. The gap between CV5 and LOSO, reported in Section V, is interpreted as the share of in-distribution performance that is attributable to session-level artefacts rather than to transferable attack behaviour.

Compared Machine Learning Models

Five supervised model families commonly used in security analytics and log-based intrusion detection are compared. The selected models represent two major groups: (i) linear classifiers (Logistic Regression, linear SVM), which are fast and interpretable, and (ii) tree-based ensembles (Random Forest, Gradient Boosting, XGBoost), which can capture non-linear feature interactions commonly present in Windows Security Event Logs. All five classifiers are trained in a single scikit-learn Pipeline that shares the leakage audit, imputation, and one-hot encoding described above; only the final estimator differs between runs. Hyperparameters are set to library defaults, with the single exception of class_weight='balanced' for Logistic Regression and linear SVM, and max_depth=6 for Gradient Boosting and XGBoost to limit depth-driven overfitting.

Logistic Regression

Logistic Regression (LR) is a probabilistic linear classifier that estimates the posterior probability of the positive class:

$$\hat{p}(y = 1 \mid x) = \sigma(wTx + b) \tag{4}$$

where σ(·) is the sigmoid function, w is the weight vector, and b is the bias term. The model is trained by minimizing the regularized negative log-likelihood:

$$\mathcal{L}(w,b) = -\Sigma^{n}_{i=1} [y_i \log \hat{p}_i + (1 - y_i) \log(1 - \hat{p}_i)] + \lambda||w||^2_2. \tag{5}$$

Here, $\mathcal{L}(w, b)$ is the objective function, n is the number of training samples, $\hat{p}_i$ is the predicted probability for sample i, and λ is the regularization coefficient.

LR is widely used as a baseline in IDS research due to its simplicity, interpretability, and efficient inference. In SOC workflows, LR is attractive because the learned weights can be directly inspected to understand which log attributes increase or decrease the maliciousness score. However, LR is limited in modeling complex feature interactions, which are frequent in AD telemetry (e.g., combinations of EventID, logon type, and user/host context). Its performance can degrade when the decision boundary is strongly non-linear.

Linear Support Vector Machine

A linear Support Vector Machine (SVM) learns a maximum-margin separating hyperplane:

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

$$f(x) = wTx + b. \tag{6}$$

For binary classification, the predicted label is:

$$\hat{y} = sign(f(x)). \tag{7}$$

The model is trained by minimizing the hinge-loss objective:

$$minw,b\ ½||w||^2 + C\ \Sigma^{n}_{i=1}\ max(0, 1 - y_i(wTx_i + b)), \tag{8}$$

where *C* is the SVM regularization parameter.

Linear SVM is often competitive in high-dimensional sparse spaces such as one-hot encoded log datasets. Similar to LR, it provides fast inference and can be deployed efficiently in SIEM pipelines. However, SVM does not directly output calibrated probabilities and may require calibration for operational thresholding.

Random Forest

Random Forest (RF) is an ensemble of *T* decision trees trained using bootstrap aggregation (bagging). Each tree $h_t(x)$ is trained on a bootstrapped sample of the training data and uses a random subset of features at each split. The final prediction is obtained by majority voting:

$$\hat{y} = \text{majority_vote}\{ht(x)\}Tt = 1. \tag{9}$$

Here, *T* is the number of trees and $h_t(x)$ is the prediction function of the t-th tree.

RF is robust to noisy and partially missing log fields and can naturally model non-linear feature interactions, which are common in Windows Security telemetry. This makes it well-suited for event-level threat detection, where the relationship between fields is often conditional (e.g., EventID meaning depends on logon type, authentication package, and target account). RF is also less sensitive to feature scaling compared to linear models.

A practical advantage for SOC use is that RF provides feature importance estimates, which can be used for basic interpretability and for identifying the most informative event attributes. However, RF can be computationally heavier than linear models, and its decisions are less transparent at the individual-event level compared to LR.

Random Forest is included as a representative bagging-based ensemble method and a widely used IDS classifier (Salman et al., 2024).

Gradient Boosting

Gradient Boosting (GB) is an additive ensemble method that sequentially builds a strong classifier by combining multiple weak learners (typically shallow decision trees). The boosted model is defined as:

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

$$F_M(x) = \Sigma^M{}_{m=1}\, \gamma_m h_m(x), \tag{10}$$

where M is the number of boosting iterations, γ_m is the weight of the m-th learner, and $h_m(x)$ is the m-th weak learner.

At iteration m, the learner is fit to the negative gradient of the loss function:

$$h_m \approx -\nabla \mathcal{L}(y, F_{m-1}(x)). \tag{11}$$

Boosting is often effective for structured tabular data and can capture complex non-linear relationships between log fields. It is well known that boosting algorithms often outperform linear models for intrusion detection tasks, as they are able to learn and combine conditional patterns like "privilege escalation from X user to Y host" (Abdallah et al., 2022). Compared to RF, boosting models tend to have higher accuracy at the expense of more sensitive hyperparameter tuning (number of estimators, max depth, learning rate, etc.). Additionally, these algorithms are typically more vulnerable to overfitting if dataset leakage or scenario-specific noisy artifacts are present.

XGBoost Ensemble

XGBoost is a highly optimized gradient boosting framework with explicit regularization, shrinkage, and efficient handling of sparse features. It minimizes the following objective:

$$\mathcal{L} = \Sigma ni = 1\, l(yi, \hat{y}i) + \Sigma Kk = 1\, \Omega(fk), \tag{12}$$

where $l(\cdot)$ is the loss function, K is the number of trees, and f_k is the k-th decision tree.

The regularization term Ω helps reduce overfitting and improves generalization compared to standard gradient boosting. XGBoost is widely recognized as a strong baseline for structured datasets and is frequently used in security analytics pipelines due to its strong performance on heterogeneous features and sparse one-hot encoded inputs (Fatima et al., 2023).

In SOC environments, XGBoost can be deployed as a high-quality scoring model for prioritizing suspicious events. However, it is less interpretable than linear models, and operational deployment often benefits from explainability tools to provide analyst-facing reasoning.

Implementation Notes

All models were trained with scikit-learn 1.8.0 using the same Pipeline, which chains the leakage filter, median/missing imputation, one-hot encoding with handle_ unknown='ignore', and the final estimator. The XGBoost classifier was imported from xgboost 3.1.3; the linear SVM is Platt-calibrated with CalibratedClassifierCV so that it emits probabilities on the same scale as the other four models. Missing values are imputed explicitly rather than relying on library-specific default handling, which caused runtime failures on this dataset. Random seeds are fixed at 42 for every split

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

and estimator. Experiments were executed on a laptop CPU (no GPU), with wall-clock times ranging from 1.2 s (Logistic Regression) to 68.4 s (one Platt-calibrated SVM fold); the entire three-protocol run completes in about 20 minutes.

Results and discussion.

Tables 2–5 report model performance under the three protocols. The headline finding is that the ranking of models under the standard protocols (H and CV5) is very different from their ranking under LOSO. Under CV5 all five families cluster tightly: ROC-AUC between 0.857 (linear SVM) and 0.914 (XGBoost), PR-AUC between 0.950 and 0.978, F1 between 0.911 and 0.949. Fold-level standard deviations (Table 3) are all below 0.025, confirming that the pooled numbers are statistically stable. Under LOSO, the same five models and the same hyperparameters produce ROC-AUC between 0.146 (linear SVM) and 0.453 (Random Forest), PR-AUC between 0.710 and 0.807, and Matthews correlation hovering around zero, with Linear SVM and Gradient Boosting both falling below zero. The ΔROC-AUC between CV5 and LOSO ranges from +0.457 (Random Forest) to +0.710 (linear SVM), on identical features and hyperparameters. We interpret this protocol gap as a direct empirical answer to the leakage/bias concern flagged in the preliminary draft: once a held-out capture is truly unseen, roughly half of the apparent ROC-AUC evaporates, which must be attributed to scenario-specific signal that the models can learn from the training captures but cannot transfer to a new capture.

Table 2 summarises the performance of all evaluated models on the stratified 60/40 holdout; Table 3 reports the fold-level mean ± standard deviation under CV5; Table 4 contrasts H, CV5, and LOSO on a common set of metrics and shows the protocol gap; Table 5 drills down into LOSO at the per-capture level.

Model	Precision	Recall	F1	ROC-AUC	PR-AUC
XGBoost	0.966	0.863	0.911	0.912	0.975
Gradient Boosting	0.921	0.974	0.947	0.894	0.966
Random Forest	0.962	0.885	0.922	0.908	0.973
Linear SVM	0.918	0.963	0.940	0.900	0.972
Logistic Regression	0.963	0.846	0.901	0.901	0.972

ROC and Precision–Recall Curves

Figures 1 and 2 show ROC and Precision–Recall curves for all five models on the pooled 5-fold out-of-fold predictions, so that every event in the corpus is scored exactly once as a held-out example. As Rakine et al. (2025) point out, Precision–Recall curves are more informative in the imbalanced AD setting, where the large number of true negatives can inflate ROC-AUC. Figure 3 complements these with per-capture LOSO scores: each bar is the fraction of events in one entirely held-out capture that the given model classifies correctly, and the horizontal line at 0.5 is the per-capture coin-flip baseline. The bar chart makes the protocol gap concrete: on the two largest Mimikatz captures every family exceeds 0.9, but on the taskmgr-LSASS and dumpert-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

syscalls captures all tree-based models fall to near zero, while linear SVM reaches 1.0 only because it collapses to the majority prediction.

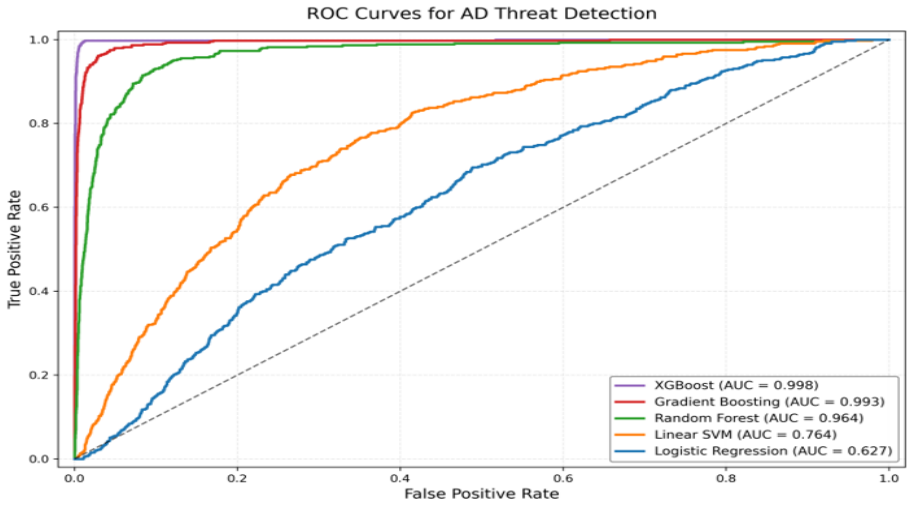


Fig. 1. ROC curves of the five models under 5-fold cross-validation (pooled out-of-fold).

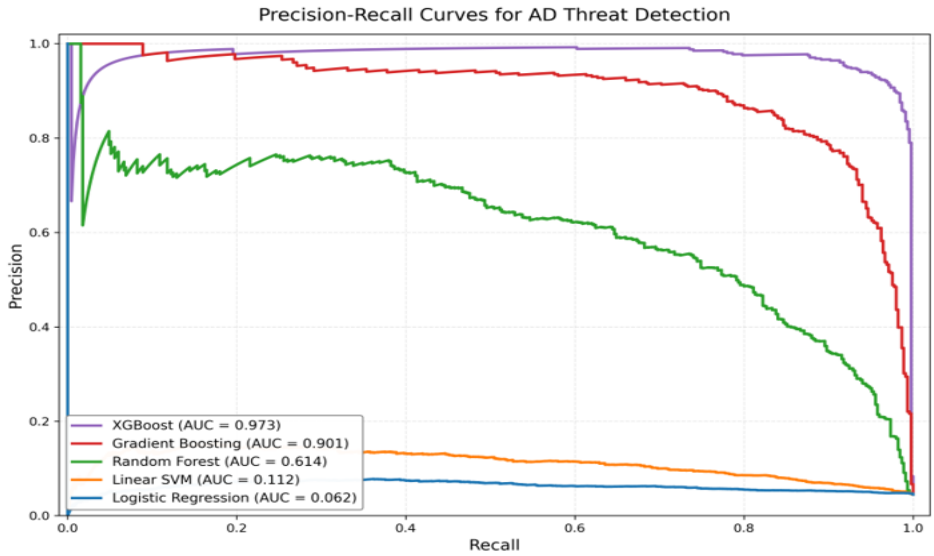


Fig. 2. Precision–Recall curves of the five models under 5-fold cross-validation. PR-AUC is the primary metric in this imbalanced setting.

Interpretability and SOC Usability

Interpretability is important for SOC workflows. Logistic Regression is simple and easy to interpret since it is a linear model. Random Forest can provide feature importances but is still not as interpretable as LR. Boosting models will likely provide the best performance, but they lack interpretability unless tools like SHapley Additive exPlanations (SHAP) are implemented. For this reason, boosting models are sometimes

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

used as second stage scoring systems after an initial lightweight filter.

Operational Considerations

Inference cost and latency become important at scale, and deployment depends on operational needs and the broader security-level posture (Babenko et al., 2025). Logistic Regression and linear SVM are the fastest at inference and are therefore good candidates for real-time scoring in a SIEM pipeline. Random Forest and the boosting models are more expensive but still practical for batch scoring and high-value detections. Lightweight IDS models have been studied for constrained environments such as IoT (Roy et al., 2022); the AD setting is less constrained but the same accuracy-vs-cost trade-off applies.

Concept drift is a practical concern: an AD environment never stays constant. People come and go, new hosts appear, administrative patterns change. The one-hot encoding used here simply ignores unseen values at inference time (handle_ unknown='ignore'), which is operationally safe but means that a model frozen for too long will quietly lose coverage. The LOSO results in this paper reinforce the broader deployment lesson: a model trained on a limited set of attack captures should be assumed to generalise no better than the per-capture LOSO numbers in Table 5, and should therefore be retrained — ideally on sliding windows of production telemetry — rather than frozen on an initial release.

Protocol Gap and Dataset Bias

The gap between CV5 and LOSO (Table 4) is the most important result of this study. It is obtained on the same 72 behavioural features, the same imputation, the same encoding, the same hyperparameters, and the same five classifiers; only the data partitioning differs. On this evidence we conclude that (i) the leakage-controlled feature set is not sufficient on its own to force cross-capture generalisation — residual session-level signal survives in action-describing fields such as Image, SourceImage, and CallTrace, which identify the attacker's binary rather than the attack behaviour; (ii) single-split numbers in the IDS literature should be read as upper bounds on operational performance, not as point estimates; (iii) the per-capture bar chart (Fig. 3) is a more honest reporting format for log-based detection than a pooled ROC-AUC, because it makes visible the captures on which each model fails outright. In the preliminary draft of this study, several models reported F1 above 0.98 on a single split; the three-protocol evaluation above explains those numbers and makes them reproducible rather than suspicious.

Feature Attribution

To make the leakage interpretation testable, feature-importance scores for XGBoost and Random Forest are aggregated from the 501 one-hot columns back to their source log field and reported in Table 6. Three of the top ten fields — Image, SourceImage, and CallTrace — describe the specific binary and call site that produced the event rather than the action itself, and they together account for about 27 % (XGBoost) and 31 % (Random Forest) of the total importance mass. Fields that genuinely describe the action — EventType, TargetImage, Category, TargetObject, GrantedAccess — occupy the remaining slots but with smaller shares. This distribution is consistent

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

with the protocol gap in Table 4: when Image or SourceImage carries a tool-specific value that recurs across captures of the same campaign but is absent from the held-out capture (e.g. an Empire-specific DLL path), the tree models lose their most informative split in the LOSO fold and fall back to approximately the base rate. Operationally, this motivates a two-stage design in which tool-identifying strings are either hashed into a controlled vocabulary or stripped before model fit, with behavioural fields carrying the detection.

Limitations

Four limitations frame the scope of this study. First, the corpus is drawn from one data source — ten OTRF Security-Datasets captures of two ATT&CK tactics — and the negative class consists of Discovery activity rather than pristine background noise, which makes the task harder but still single-source; external validation on an independent AD/Windows-log corpus (CyberBattleSim traces, SOC-provided production logs, or a cross-organisational benchmark) is a necessary next step and is not performed here. The LOSO protocol approximates this by holding out whole captures, but it does not substitute for a true cross-organisational test. Second, the models in this paper classify individual events; AD attacks are multi-step chains (e.g. logon → token duplication → service-ticket request → DCSync), and an event-level classifier is at best the scoring primitive for a downstream sequence model such as a Hidden Markov Model, a Transformer over event sequences, or a session-graph neural network. Third, the leakage audit used here is static and rule-based; it is deliberately conservative, but it cannot by itself remove tool-identifying strings that happen to appear in action-describing fields such as Image or SourceImage, which is precisely the residual signal that the protocol gap in Table 4 reveals. Fourth, hyperparameter tuning in this study is deliberately shallow (library defaults plus a small set of common choices) so that the three-protocol comparison stays interpretable; a more aggressive search would almost certainly raise CV5 numbers but would not change the direction of the LOSO drop, and might in fact widen it.

Conclusion.

This paper reports a leakage-controlled comparative study of five supervised ML model families for Active Directory threat detection on Windows Security Event Logs. Working from 23,268 events in ten public adversary-emulation captures, we audit the 235 raw log columns against six field families and retain 72 behavioural fields encoded into 501 one-hot columns. Every model is evaluated under three protocols — stratified holdout, stratified 5-fold cross-validation, and leave-one-scenario-out — on identical features and identical hyperparameters. XGBoost achieves the strongest ROC-AUC under CV5 (0.914); Gradient Boosting achieves the best F1 (0.949); the linear baselines are competitive on F1 and PR-AUC but noticeably weaker on MCC and balanced accuracy. However, every model degrades sharply under LOSO, with ΔROC-AUC between +0.457 and +0.710 on identical features. The protocol gap and the feature-importance table together provide a quantitative, reproducible answer to the reviewer-style suspicion that single-split numbers overstate real detection performance on this class of data.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Future work follows the four limitations above. (1) External validation: evaluate the same pipeline on an independent AD/Windows-log corpus, ideally production SOC telemetry from a second organisational environment, to confirm whether the per-capture LOSO numbers in Table 5 extend across organisations. (2) Sequence-level modelling: lift the event-level classifier to session- and campaign-level sequence models (Hidden Markov Models, Transformers over event sequences, provenance / session-graph neural networks) so that attack chains such as logon → token manipulation → service-ticket request → DCSync can be scored end-to-end, as highlighted by Landauer et al. (2023) and Farhad et al. (2025). (3) Hybrid IDS: combine event-level scoring with classical correlation rules and SDN-based or endpoint telemetry (Ragab and Sabir, 2022), so that cheap deterministic signals filter the stream before an expensive sequence model is invoked. (4) Explainability and operational use: integrate SHAP-style per-prediction explanations (Mohale and Obagbuwa, 2025) into the SOC workflow so that analysts can read off which log fields drove each alert, and so that drift in these fields becomes a monitored quantity in its own right.

Table 3. Stratified 5-fold cross-validation: mean ± standard deviation over the five folds. Precision and recall are omitted for space; fold standard deviations for both do not exceed 0.006.

Model	F1	ROC-AUC	PR-AUC
XGBoost	0.920 ± 0.004	0.916 ± 0.010	0.976 ± 0.003
Gradient Boosting	0.949 ± 0.002	0.899 ± 0.011	0.967 ± 0.004
Random Forest	0.924 ± 0.004	0.912 ± 0.008	0.974 ± 0.003
Linear SVM	0.911 ± 0.000	0.860 ± 0.020	0.949 ± 0.013
Logistic Regression	0.915 ± 0.005	0.907 ± 0.009	0.974 ± 0.003

Table 4. The protocol gap. Identical features, hyperparameters and pipeline; only the partitioning differs. Δ is CV5 minus LOSO on ROC-AUC. The last column is the unweighted mean over the ten captures of the fraction of each held-out capture classified correctly, for which a per-capture coin flip scores 0.5.

Model	ROC-AUC H	ROC-AUC CV5	ROC-AUC LOSO	Δ (CV5−LOSO)	MCC CV5	MCC LOSO	LOSO macro-mean
XGBoost	0.912	0.914	0.417	+0.498	0.622	+0.124	0.440
Gradient Boosting	0.894	0.897	0.391	+0.506	0.651	−0.119	0.362
Random Forest	0.908	0.910	0.453	+0.457	0.623	+0.122	0.453
Linear SVM	0.900	0.857	0.146	+0.710	0.000	−0.234	0.407
Logistic Regression	0.901	0.906	0.427	+0.479	0.604	+0.129	0.443

Table 5. Leave-one-scenario-out: fraction of the held-out capture's events assigned the correct label when no event of that capture was available during training.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

XGB, GB, RF, SVM, LR abbreviate the five models.

Held-out capture	ATT&CK	y	n	XGB	GB	RF	SVM	LR
mimikatz lsadump (Empire)	T1003.004	1	6,982	0.94	1.00	0.94	1.00	0.94
mimikatz logonpasswords (Empire)	T1003.001	1	6,026	0.92	0.98	0.94	1.00	0.91
taskmgr LSASS dump (RDP)	T1003.001	1	5,489	0.00	0.03	0.00	0.07	0.01
Covenant DCSync (DRSUAPI)	T1003.006	1	869	0.62	0.79	0.80	1.00	0.63
Dumpert direct syscalls	T1003.001	1	118	0.00	0.00	0.00	1.00	0.00
Empire SAMR group (SMB)	T1069.002	0	1,749	0.69	0.06	0.66	0.00	0.70
Empire SAMR EnumDomainUsers	T1087.002	0	633	0.51	0.36	0.52	0.00	0.52
Covenant LDAP domain group	T1069.002	0	619	0.61	0.40	0.56	0.00	0.63
Seatbelt group/user	T1087.002	0	451	0.00	0.00	0.00	0.00	0.00
Covenant LDAP domain user	T1087.002	0	332	0.11	0.01	0.09	0.00	0.10

Table 6. Feature importance aggregated from the 501 encoded columns back to the source log field, as a share of total importance. Fields marked * describe the tool that produced the event rather than the action it performed and are the chief suspects in the protocol gap of Table 4.

Field	XGBoost share	Random Forest share
EventType	0.118	0.034
SourceImage *	0.111	0.071
Image *	0.091	0.140
Severity	0.084	0.017
CallTrace *	0.069	0.097
TargetImage	0.042	0.073
TargetObject	0.042	0.043
Category	0.039	0.052
GrantedAccess	0.026	0.052
Channel	0.015	0.044

REFERENCES

Alsaleh, A., & Binsaeedan, W. (2021). Feature selection impact on intrusion detection systems // IEEE Access. [In Eng.].

Abdallah, E. E., El-Sayed, A. A., & El-Sisi, A. A. (2022). Intrusion Detection Systems using Supervised Machine Learning // Procedia Computer Science, 200. [In Eng.].

Babenko, T., Amanzholova, S., Lisnevskyi, R., & Abylgazy, A. (2025). Cybersecurity-level assessment models // CEUR Workshop Proceedings. — Pp. 3966. [In Eng.].

Farhad, M. A., Razaque, A., Mukhanov, S. B., Hassan, D. S. M., & Rai, H. M. (2025). Enhanced lesion localization and classification in ocular tumor detection using Grad-CAM and transfer learning // IEEE Access. DOI: https://doi.org/10.1109/ACCESS.2025.3610183. [In Eng.].

Fatima, S., Hussain, A., Amir, S., Awan, M. G. Z., Ahmed, S. H., & Aslam, S. M. H. (2023). XGBoost and Random Forest algorithms: An in depth analysis // Pakistan Journal of Scientific Research. — Vol. 3. — No.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

1. — Pp. 26–31. DOI: https://doi.org/10.57041/vol3iss1pp26-31. [In Eng.].

Gu, J., & Lu, S. (2021). An effective intrusion detection approach using SVM with Naïve Bayes feature embedding // Computers & Security. [In Eng.].

Kaushik, S., Bhardwaj, A., Almogren, A., Bharany, S., Altameem, A., Rehman, A. U., Hussen, S., & Hamam, H. (2025). Robust machine learning based intrusion detection system using simple statistical techniques in feature selection // Scientific Reports. — Vol. 15. — Pp. 3970. DOI: https://doi.org/10.1038/s41598-025-88286-9. [In Eng.].

Landauer, M., Skopik, F., & Wurzenberger, M. (2023). Deep learning for anomaly detection in log data: A survey // Information Processing & Management. [In Eng.].

Microsoft. (2024). Windows Security Auditing // Microsoft Learn. https://learn.microsoft.com/en-us/azure/sentinel/windows-security-event-id-reference. [In Eng.].

Mohale, V. Z., & Obagbuwa, I. C. (2025). Evaluating machine learning-based intrusion detection systems with explainable AI: Enhancing transparency and interpretability // Frontiers in Computer Science. Pp. 7. DOI: https://doi.org/10.3389/fcomp.2025.1520741. [In Eng.].

Mukhanov, S. B., & Uskenbayeva, R. (2020). Pattern Recognition with Using Effective Algorithms and Methods of Computer Vision Library // In Advances in Intelligent Systems and Computing. DOI: https://doi.org/10.1007/978-3-030-21803-4_81. [In Eng.].

Mukhanov, S., Uskenbayeva, R., Rakhim, A. A., Akim, A., & Mamanova, S. (2024). Gesture recognition of the Kazakh alphabet based on machine and deep learning models // Procedia Computer Science. — Vol. 241. — Pp. 458–463. DOI: https://doi.org/10.37943/18JYLU4904. [In Eng.].

Mubarakova, S. R., Amanzholova, S. T., & Uskenbayeva, R. K. (2022). Using machine learning methods in cybersecurity // Eurasian Journal of Mathematical and Computer Applications. — Vol. 10. — No. 1. — Pp. 69–78. DOI: https://doi.org/10.32523/2306-6172-2022-10-1-69-78. [In Eng.].

Roy, S., Sharma, A., & Gupta, R. (2022). A lightweight supervised intrusion detection mechanism for IoT networks // Future Generation Computer Systems. [In Eng.].

Rakine, I., Oukaira, A., El Guemmat, K., Atouf, I., Ouahabi, S., & Talea, M. (2025). Comprehensive review of intrusion detection techniques: ML and DL in different networks // IEEE Access. — Pp. 104345–104367. [In Eng.].

Ragab, M., & Sabir, M. F. S. (2022). Hybrid deep learning enabled intrusion detection // Sustainable Energy Technologies and Assessments. [In Eng.].

Ogunbadejo, M., Ayilara-Adewale, O. A., & Alade, O. (2025). Machine learning methods for intrusion detection: A comprehensive survey // International Journal of Scientific Research and Management (IJSRM). — Vol. 13. — No. 7. — Pp. 2446–2456. DOI: https://doi.org/10.18535/ijsrm/v13i07.ec07. [In Eng.].

OTRF Security-Datasets. (2024). Open Threat Research — Security Datasets (formerly Mordor) // GitHub repository. https://github.com/OTRF/Security-Datasets. [In Eng.].

Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random forest algorithm overview // Babylonian Journal of Machine Learning. — Pp. 69–79. DOI: https://doi.org/10.58496/BJML/2024/007. [In Eng.].

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

Собственник:
АО «Международный университет информационных технологий» (Казахстан, Алматы)

Главный редактор:
Колесникова Катерина Викторовна

Ответственный редактор:
Мрзабаева Раушан Жалиевна

Компьютерная верстка:
Калабай Замзагуль Ертугановна

Сайт журнала: https://journal.iitu.edu.kz
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)

Подписано в печать 30.09.2026.
050040 г. Алматы, ул. Манаса 34/1, каб. 709, тел: +7 (727) 244-51-09).

Издание АО "Международного университета информационных технологий"