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

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

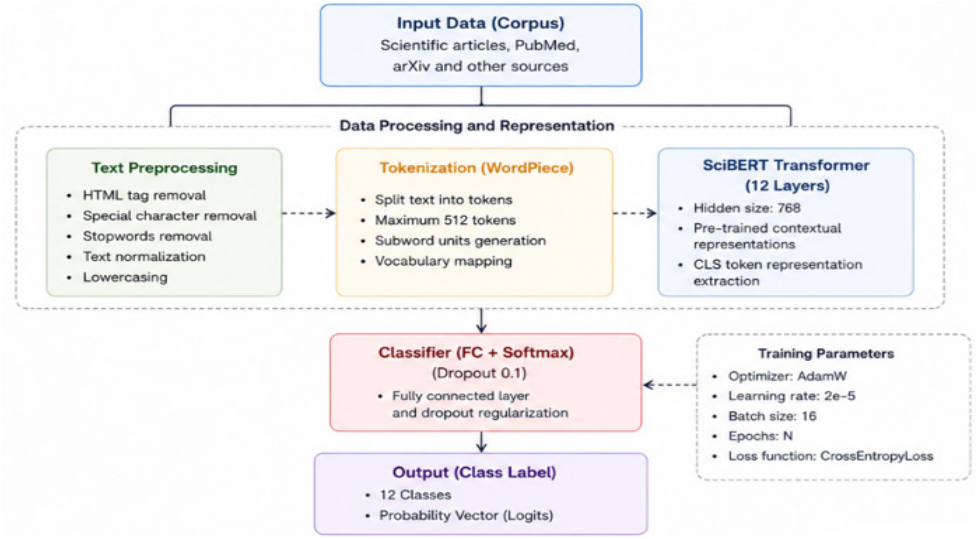


Fig. 1. Architectural diagram of the proposed AI classification pipeline.

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: $R^2 = 0.9983$.

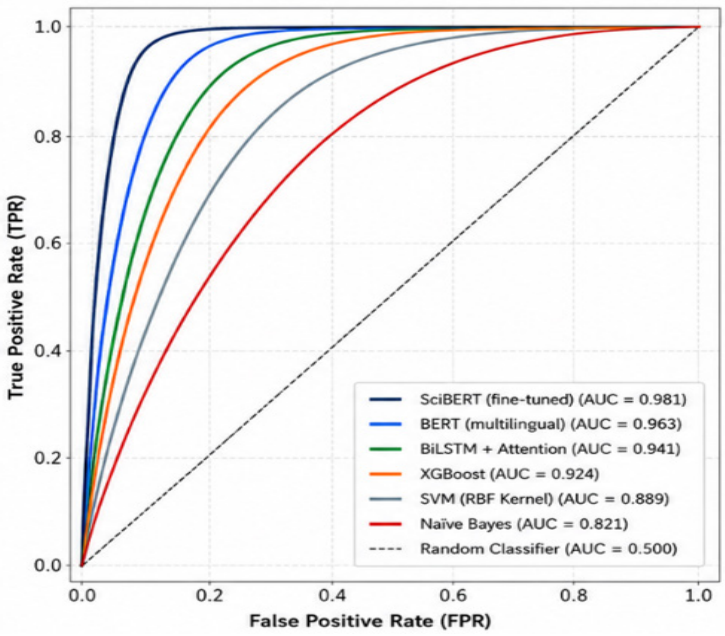


Fig. 2. ROC curves for six classification algorithms. Dashed diagonal: random classifier (AUC = 0.500) Shaded area: AUC gain of SciBERT over the random baseline.

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.

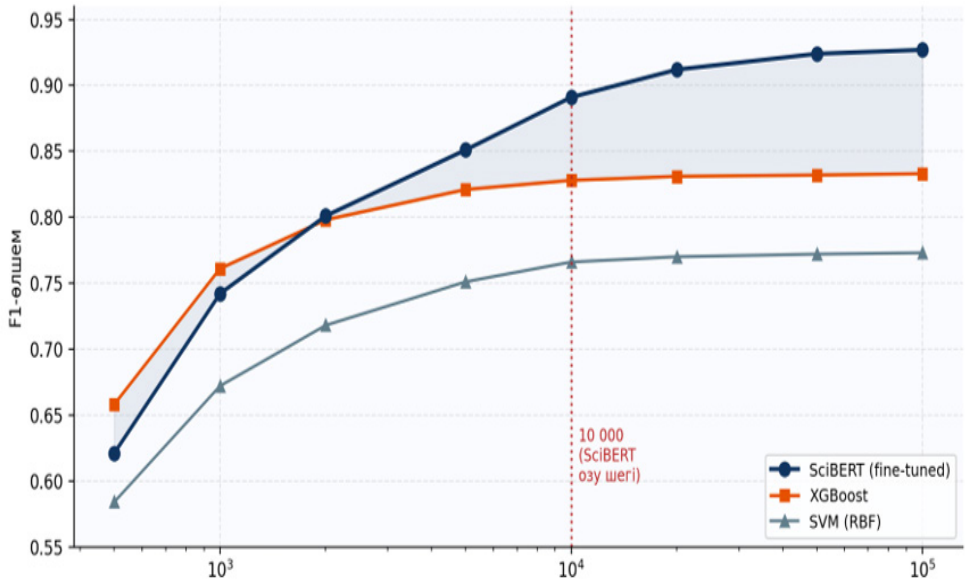


Fig. 3. Learning curves: F_1-score vs. training set size (logarithmic scale). Dotted vertical line: 10,000-article threshold beyond which SciBERT consistently outperforms XGBoost.

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.

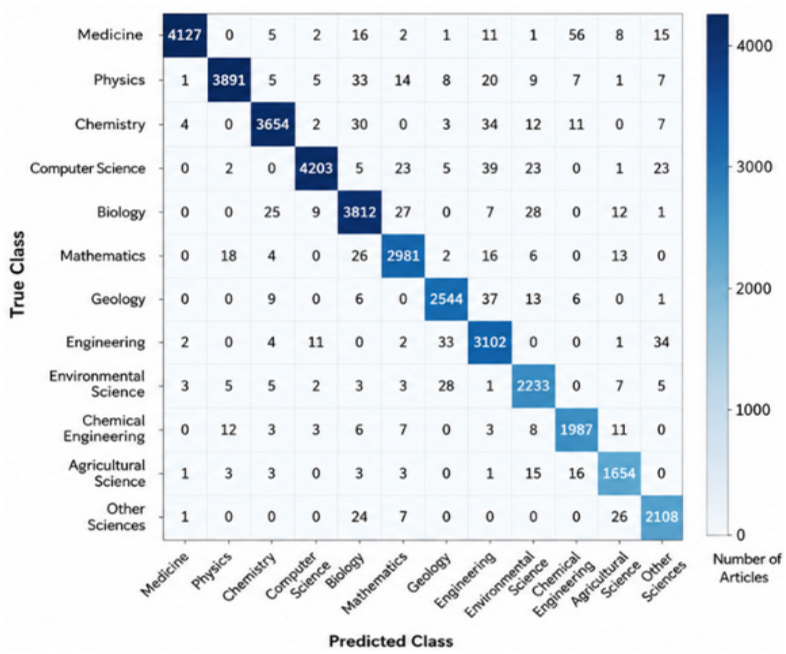


Fig. 4. Confusion matrix for SciBERT (fine-tuned), test set, 12 scientific domains. Diagonal cells (correct predictions) are highlighted. Colour intensity is proportional to article count.

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

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"