ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ ҒЫЛЫМ ЖӘНЕ ЖОҒАРЫ БІЛІМ МИНИСТРЛІГІ
МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РЕСПУБЛИКИ КАЗАХСТАН
MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE REPUBLIC OF KAZAKHSTAN

ХАЛЫҚАРАЛЫҚ УНИВЕРСИТЕТІ INTERNATIONAL UNIVERSITY SINCE 2009 МЕЖДУНАРОДНЫЙ УНИВЕРСИТЕТ

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

2025 (22) 2

сәуір - маусым

ISSN 2708–2032 (print)
ISSN 2708–2040 (online)

Халықаралық ақпараттық және коммуникациялық технологиялар журналы
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Меншік иесі: АҚ «Халықаралық ақпараттық технологиялар университеті» (Алматы қ.).
Қазақстан Республикасы Ақпарат және қоғамдық даму министрлігіне мерзімді баспасөз басылымын есепке қою туралы куәлік № KZ82VPY00020475, 20.02.2020 ж. берілген
Тақырып бағыты: ақпараттық технологиялар, ақпараттық қауіпсіздік және коммуникациялық технологиялар, әлеуметтік-экономикалық жүйелерді дамытудағы цифрлық технология.
Мерзімділігі: жылына 4 рет.
Тираж: 100 дана.
Редакция мекенжайы: 050040 Алматы қ., Манас к., 34/1, каб. 709, тел: +7 (727) 244-51-09.
E-mail: ijict@iitu.edu.kz
Журнал сайты: https://journal.iitu.edu.kz

Журнал сайты: https://journal.iitu.edu.kz

Международный журнал информационных и коммуникационных технологий
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)

Собственник: АО «Международный университет информационных технологий» (г. Алматы).

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

Тематическая направленность: информационные технологии, информационная безопасность и коммуникационные технологии, цифровые технологии в развитии социо-экономических систем**.**

Периодичность: 4 раза в год.

Тираж: 100 экземпляров.

Адрес редакции: 050040 г. Алматы, ул. Манаса 34/1, каб. 709, тел: +7 (727) 244-51-09.

E-mail: ijict@iitu.edu.kz Сайт журнала: https://journal.iitu.edu.kz

---

«International Journal of Information and Communication Technologies»
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Owner: International Information Technology University JSC (Almaty).
The certificate of registration of a periodical printed publication in the Ministry of Information and Social Development of the Republic of Kazakhstan, Information Committee No. KZ82VPY00020475, issued on 20.02.2020.
Thematic focus: information technology, digital technologies in the development of socio-economic systems, information security and communication technologies
Periodicity: 4 times a year.
Circulation: 100 copies.
Editorial address: 050040. Manas st. 34/1, Almaty. +7 (727) 244-51-09. E-mail: ijict@iitu.edu.kz
Journal website: https://journal.iitu.edu.kz

МАЗМҰНЫ

СОДЕРЖАНИЕ

CONTENTS

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 6. Is. 2. Number 22 (2025). Pp. 214–234
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2025.22.2.014

ADAPTING FACE DETECTION AND RECOGNITION MODELS FOR ECHOCARDIOGRAPHIC DEFECT DIAGNOSIS

B.M. Ukibassov[1*], S.B. Rakhmetulayeva[2], Zh.O. Zhanabekov[3], Ansar-Ul-Haque Yasar[4]
[1]Narxoz Univeristy, Almaty, Kazakhstan;
[2]Satbayev University, Almaty, Kazakhstan;
[3]Kazakh-British Technical University, Almaty, Kazakhstan;[4]Hasselt University, Hasselt, Belgium.
E-mail: b.ukibassov@iitu.edu.kz

Ukibassov B. — PhD student, senior lecturer, «the School of Digital Technology», Narxoz University, Almaty, Kazakhstan
E-mail: b.ukibassov@iitu.edu.kz. https://orcid.org/0000-0002-0439-0187;
Rakhmetulayeva S. — Doctor of Computer Science, professor, «the Department of Cybersecurity», information processing and storage, Satbayev University, Almaty, Kazakhstan
https://orcid.org/0000-0001-9201-1118;
Zhanabekov Zh. — senior lecturer, «the School of IT and Engineering», Kazakh-British Technical University, Almaty, Kazakhstan
https://orcid.org/0000-0001-5984-7132;
Ansar-Ul-Haque Yasar — Doctor of Computer Science, professor, «the Transportation Research Institute», Hasselt University, Hasselt, Belgium
https://orcid.org/0000-0002-1542-2658.

Abstract. Accurate detection of atrial and ventricular septal defects (ASD, VSD) in pediatric echocardiography remains a clinical challenge, especially in resource-constrained settings. In this work, we propose a novel deep learning framework that adapts state-of-the-art face detection and recognition pipelines—RetinaFace and ArcFace—for medical image analysis. Our method comprises a two-stage architecture: (1) a RetinaFace-inspired module that detects and crops the cardiac septal region from transthoracic echocardiogram frames, followed by (2) a ResNet-50-based classification network trained with an ArcFace loss function to learn discriminative embeddings of defect types. We evaluate the system on a curated pediatric echocardiography dataset labeled as Normal, ASD, or VSD. The model achieves 94.4% classification accuracy and an AUC of 0.976 in a pairwise verification setting, outperforming conventional CNN classifiers. These findings suggest that embedding-based metric learning, as pioneered in face recognition, can be successfully translated to medical imaging tasks requiring fine-grained classification. These findings illustrate that face recognition pipelines can be effectively repurposed for medical image analysis, offering a promising new direction for computer-aided echocardiographic diagnosis of congenital defects.

Keywords: Echocardiography; atrial septal defect (ASD); ventricular septal defect (VSD)

For citation: B.M. Ukibassov, S.B. Rakhmetulayeva, Ansar-Ul-Haque Yasar, Zh.O. Zhanabekov. Adapting face detection and recognition models for echocardiographic defect diagnosis//International journal of information and communication technologies. 2025. Vol. 6. No. 22. Pp. 214–234 (In Eng.). https://doi.org/10.54309/IJICT.2025.22.2.014.

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

ЭХОКАРДИОГРАММАЛАРДАҒЫ АҚАУЛАРДЫ ДИАГНОСТИКАЛАУ ҮШІН БЕТТІ АНЫҚТАУ ЖӘНЕ ТАНУ ҮЛГІЛЕРІН БЕЙІМДЕУ

Б.М. Укибассов[1]*, С.Б. Рахметулаева[2], Ж.О. Жанабеков[3], Ансар-уль-Хак Ясар[4],

[1]Нархоз Университеті, Алматы, Қазақстан;

[2]Қазақстан ұлттық техникалық зерттеу университеті. Қ. И. Сәтбаева, Алматы, Қазақстан;

[3]Қазақ-Британ техникалық университеті, Алматы, Қазақстан;

[4]Хассельт Университеті, Хассельт, Бельгия.

E-mail: b.ukibassov@iitu.edu.kz

Укибассов Б. — PhD студент, аға оқытушы, «Цифрлық технологиялар мектебі», Нархоз университеті, Алматы, Қазақстан
E-mail: b.ukibassov@iitu.edu.kz. https://orcid.org/0000-0002-0439-0187;

Рахметулаева С. — информатика ғылымдарының докторы, «Киберқауіпсіздік, ақпаратты өңдеу және сақтау» кафедрасының профессоры, Қазақстан ұлттық техникалық зерттеу университеті. Қ. И. Сәтбаева, Алматы, Қазақстан
https://orcid.org/0000-0001-9201-1118;

Жанабеков Ж.— аға оқытушы, «Ақпараттық технологиялар және Инженерия мектебі», Қазақ-Британ техникалық университеті, Алматы, Қазақстан
https://orcid.org/0000-0001-5984-7132;

Ансар-уль-Хак Ясар — информатика ғылымдарының докторы, профессор, көлікті зерттеу институты, Хасселт университеті, Хасселт, Бельгия
https://orcid.org/0000-0002-1542-2658.

Аннотация. Педиатриялық эхокардиографиядағы атриальды және қарыншалық септальды ақауларды (ASD, VSD) дәл диагностикалау әлі де клиникалық сынақ болып табылады, әсіресе ресурстар шектеулі жағдайларда. Бұл жұмыс Медициналық кескіндерді талдау тапсырмалары үшін бет — әлпетті тану саласындағы озық шешімдерді — RetinaFace және ArcFace-бейімдейтін терең оқытудың жаңа архитектурасын ұсынады. Ұсынылған әдіс екі сатылы архитектураны қамтиды: (1) трансторакальды эхокардиограмма кадрларындағы жүрек септумының аймағын анықтауға және кесуге арналған RetinaFace шабыттандырылған модуль және (2) ақаулардың кемсітушілік эмбеддингтерін алу үшін ArcFace жоғалту функциясын қолдана отырып оқытылған ResNet-50 негізіндегі жіктеу желісі. Жүйе үш класты балалар эхокардиографиясының белгіленген Деректер жиынтығында сыналған: норма, ASD және VSD. Модель 94.4% жіктеу дәлдігіне және 0.976 AUC мәніне жұптық тексеру тапсырмасында дәстүрлі CNN классификаторларынан асып түсті. Нәтижелер бетті танудан алынған метрикалық оқыту әдістерін нәзік жіктеу тапсырмалары үшін медициналық бейнелеуде сәтті қолдануға болатынын көрсетеді. Бұл тұжырымдар эхокардиограммалардағы туа біткен жүрек ақауларын компьютерлік диагностикалаудың перспективалық бағыттарын аша отырып, медициналық кескіндерді талдау үшін бетті тану үлгілерін тиімді репурпациялау мүмкіндігін растайды.

Түйін сөздер: эхокардиография; атриальды септальды ақау (ASD); қарыншалық септальды ақау (VSD)

Дәйексөздер үшін: Б.М. Укибассов, С.Б. Рахметулаева, Ансар-уль-Хак Ясар, Ж.О. Жанабеков. Эхокардиограммалардағы ақауларды диагностикалау үшін бетті анықтау және тану үлгілерін бейімдеу//Халықаралық ақпараттық және коммуникалық технологиялар журналы. 2025. Т. 6. No. 22. 214–234 бет. (ағылшын тілінде). https://doi.org/10.54309/IJICT.2025.22.2.000.

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

АДАПТАЦИЯ МОДЕЛЕЙ ОБНАРУЖЕНИЯ И РАСПОЗНАВАНИЯ ЛИЦ ДЛЯ ДИАГНОСТИКИ ДЕФЕКТОВ НА ЭХОКАРДИОГРАММАХ

Б.М. Укибассов[1*], С.Б. Рахметулаева[2], Ж.О. Жанабеков[3], Ансар-уль-Хак Ясар[4]

[1]Нархоз Университет, Алматы, Казахстан;
[2]Казахстанский национальный исследовательский технический университет им. К.И. Сатпаева, Алматы, Казахстан;
[3]Казахско-Британский технический университет, Алматы, Казахстан;
[4]Университет Хасселта, Хасселт, Бельгия.
E-mail: b.ukibassov@iitu.edu.kz

Укибассов Ба. — PhD студент, старший преподаватель, «Школа цифровых технологий», Нархоз университет, Алматы, Казахстан
E-mail: b.ukibassov@iitu.edu.kz. https://orcid.org/0000-0002-0439-0187;
Рахметулаева С. — доктор компьютерных наук, профессор кафедры «Кибербезопасности, обработки и хранения информации», Казахстанский национальный исследовательский технический университет им. К.И. Сатпаева, Алматы, Казахстан
https://orcid.org/0000-0001-9201-1118;
Жанабеков Ж. — старший преподаватель, «Школа информационных технологий и инженерии»,

Казахско-Британский технический университет, Алматы, Казахстан
https://orcid.org/0000-0001-5984-7132;
Ансар-уль-Хак Ясар — доктор компьютерных наук, профессор, Институт транспортных исследований, Университет Хасселта, Хасселт, Бельгия
https://orcid.org/0000-0002-1542-2658.

Аннотация. Точная диагностика дефектов межпредсердной и межжелудочковой перегородки (ASD, VSD) на детской эхокардиографии по-прежнему представляет собой клиническую задачу, особенно в условиях ограниченных ресурсов. В данной работе предлагается новая архитектура глубинного обучения, адаптирующая передовые решения из области распознавания лиц — RetinaFace и ArcFace — для задач анализа медицинских изображений. Предложенный метод включает двухэтапную архитектуру: (1) модуль, вдохновлённый RetinaFace, для обнаружения и обрезки области сердечной перегородки на кадрах трансторакальной эхокардиограммы, и (2) классификационную сеть на основе ResNet-50, обученную с использованием функции потерь ArcFace для получения дискриминативных эмбеддингов типов дефектов. Система протестирована на размеченном датасете детской эхокардиографии с тремя классами: норма, ASD и VSD. Модель достигла точности классификации 94.4 % и значения AUC 0.976 в парной задаче верификации, превзойдя традиционные CNN-классификаторы. Полученные результаты демонстрируют, что методы метрического обучения, заимствованные из распознавания лиц, могут быть успешно применены в медицинской визуализации для задач тонкой классификации. Данные выводы подтверждают возможность эффективной репурпозации моделей распознавания лиц для анализа медицинских изображений, открывая перспективные направления в компьютерной диагностике врождённых сердечных дефектов на эхокардиограммах.

Ключевые слова: эхокардиография; дефект межпредсердной перегородки (ASD); дефект межжелудочковой перегородки (VSD)

Для цитирования: Б.М. Укибассов, С.Б. Рахметулаева, Ансар-уль-Хак Ясар, Ж.О. Жанабеков. Адаптация моделей обнаружения и распознавания лиц для диагностики дефектов на эхокардиограммах // Международный журнал информационных и коммуникационных технологий. 2025. Т. 6. № 22. С. 214–234 (на англ.). https://doi.org/10.54309/IJICT.2025.22.2.014.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Introduction

Congenital heart disease (CHD) remains one of the most prevalent birth defects globally, affecting approximately 1 in every 100 live births (Zimmerman et al., 2020). Among the various types of CHD, atrial septal defects (ASDs) and ventricular septal defects (VSDs) are two of the most frequently diagnosed forms. While some of these defects may spontaneously close or remain asymptomatic, larger or poorly positioned ASDs and VSDs can result in severe complications such as congestive heart failure, arrhythmias, pulmonary hypertension, or even stroke due to paradoxical embolism. Hence, early and precise identification of these septal defects is essential for initiating timely clinical intervention and improving long-term patient outcomes.

Transthoracic echocardiography (TTE) is the current gold standard for non-invasive diagnosis of septal defects, widely used in pediatric cardiology due to its safety, accessibility, and real-time imaging capabilities. However, the diagnostic performance of TTE is highly dependent on operator skill, patient cooperation, acoustic window quality, and anatomical variations. In many healthcare settings, particularly in low- and middle-income countries, the availability of highly trained sonographers and pediatric cardiologists is limited, which introduces variability and potential inaccuracy in interpreting echocardiograms. Even in well-resourced hospitals, subtle ASDs and VSDs may be overlooked, misclassified, or confused with other cardiac anomalies in standard 2D echo views such as the apical four-chamber or subcostal sagittal views.

In recent years, artificial intelligence (AI), particularly deep learning, has demonstrated significant promise in medical image interpretation. From radiology and pathology to dermatology and ophthalmology, convolutional neural networks (CNNs) have shown the ability to match or surpass human-level performance in tasks such as classification, segmentation, and anomaly detection. Cardiology has similarly begun integrating AI tools, with successful applications including automated ejection fraction estimation (Zhang et al., 2018), wall motion scoring (Kusunose et al., 2020), and chamber segmentation. Nonetheless, most existing approaches in echocardiography focus on broad classification tasks—such as detecting any abnormality or estimating global cardiac function—rather than fine-grained classification between similar structural abnormalities like ASDs and VSDs.

The diagnostic challenge of distinguishing ASDs from VSDs stems from their anatomical proximity and the subtle variations in septal morphology visible in static 2D ultrasound images. This makes the problem analogous to certain tasks in computer vision, such as fine-grained facial recognition. Face recognition systems, particularly those built on deep learning frameworks, have achieved exceptional results in distinguishing among thousands of individuals based on minuscule differences in facial features. Models like FaceNet (Schroff et al., 2015), SphereFace (Liu et al., 2017), CosFace (Wang et al., 2018), and ArcFace (Deng et al., 2019) have redefined the benchmark for verification and identification by enforcing metric learning principles—teaching the network to map inputs into an embedding space where distances correspond to semantic similarity.

Our hypothesis is that this same embedding-based learning strategy can be repurposed for septal defect classification in echocardiography. Instead of mapping facial identities, the model learns to map ASD, VSD, and Normal classes into compact, separable regions of a feature space. A key benefit of this approach is not only improved classification accuracy but also the ability to perform verification-style tasks—e.g., determining whether multiple echo views from the same patient consistently indicate the same type of defect.

To operationalize this hypothesis, we propose a two-stage framework that integrates two cornerstone models from face recognition. First, a modified RetinaFace model (Deng et al., 2020) is used to localize and crop the septal region within echocardiographic images. This step functions analogously to face detection and alignment, ensuring the classifier receives consistent and contextually relevant input. Second, a ResNet-50 backbone is trained with the ArcFace angular margin loss to produce 512-dimensional embeddings for defect classification and verification.

In this paper, we detail the architecture, training methodology, and evaluation metrics used to implement and validate this novel pipeline. We demonstrate the model's effectiveness on a curated dataset of pediatric echocardiograms, achieving superior performance compared to conventional CNN baselines. We also highlight the advantages of embedding-based learning in handling subtle morphological variations, supporting our thesis that cross-domain transfer from face recognition to cardiac image analysis is both feasible and beneficial.

Finally, we critically examine the implications of this approach for clinical practice, including its potential role in screening workflows, longitudinal patient monitoring, and AI-assisted decision support. We conclude by outlining the limitations of our current study and the opportunities for future research in expanding this framework to broader diagnostic applications in cardiology.

Related Work

Echocardiographic Defect Detection with Deep Learning

The use of deep learning in echocardiographic image interpretation has rapidly expanded over the past decade, driven by advances in convolutional architectures and increased access to large-scale annotated datasets. While traditional approaches relied heavily on handcrafted features and rule-based systems, modern models leverage data-driven representations that learn both low-level and high-level structures directly from imaging data. In this context, echocardiographic diagnosis of congenital heart defects (CHDs), particularly atrial and ventricular septal defects (ASDs and VSDs), has become a fertile ground for innovation.

Hong et al. (2022) demonstrated the viability of deep learning in identifying secundum ASDs from color Doppler echocardiographic images. Their system utilized still frames from key views and applied a CNN for binary classification (ASD vs. normal). The model achieved ~88% accuracy, confirming that single-frame echocardiographic interpretation using AI is both feasible and potentially impactful. However, their system lacked a mechanism for differentiating between different types of defects or verifying consistency across patient views.

Lin et al. (2023) extended the scope by incorporating temporal sequences from echocardiographic videos and training an ensemble of models to both detect and quantify ASD sizes. Their approach integrated dynamic features, achieving an AUC of 0.92, and demonstrated that incorporating motion can enhance diagnostic performance. Still, the method focused solely on ASDs and required video input, which may not be available in all clinical settings.

Other notable efforts include Wang et al. (2021), who tackled multi-class classification of congenital defects including ASD, VSD, and Tetralogy of Fallot using multi-view echocardiograms. While their model aggregated features across views and achieved good performance, it did not employ any form of metric learning, and the fine-grained discrimination between ASD and VSD was not explicitly analyzed.

Beyond classification, segmentation-based models have also emerged. Nova et al. (2021) used U-Net-style architectures to segment cardiac chambers and identify potential septal discontinuities. Chen et al. (2021) adopted a modified YOLOv4-DenseNet approach to detect VSDs as objects within echo frames. These detection-oriented pipelines offer better localization but often fall short on defect-type classification and verification.

Face Recognition and Discriminative Feature Learning

In the face recognition domain, significant progress has been made through embedding-based learning. Schroff et al. (2015) introduced FaceNet, which learns an embedding space where distances reflect identity similarity using a triplet loss. This was foundational in establishing distance-based metrics for recognition tasks. Following this, SphereFace (Liu et al., 2017), CosFace (Wang et al., 2018), and ArcFace (Deng et al., 2019) introduced angular margin losses that enforced class separation on hyperspherical manifolds, enabling more compact intra-class clustering and improved verification accuracy. These models have consistently set benchmarks across standard datasets such as LFW, MegaFace, and IJB-A.

Of particular interest is ArcFace, which introduced an additive angular margin loss that significantly improved inter-class separability while maintaining model simplicity and convergence stability. ArcFace demonstrated near-human or super-human performance in large-scale identification tasks, making it an ideal candidate for repurposing in fine-grained classification settings outside facial analysis.

Metric learning principles have also seen some use in medical imaging. Tajbakhsh et al. (2016) discussed transfer learning strategies and the benefits of fine-tuning pretrained CNNs for medical tasks. Metric learning has been applied to pathology image retrieval and lesion matching, though few works have explored its utility for classification tasks involving subtly different disease states.

To date, there has been limited integration of face recognition frameworks in echocardiography. While transfer learning from general-purpose models such as ImageNet-pretrained ResNet or EfficientNet is common, no study, to our knowledge, has adapted ArcFace or RetinaFace for heart defect classification. This paper fills that gap by proposing a two-stage pipeline that uses face recognition methodologies in a cardiology context. We argue that the ability to differentiate visually similar categories—like two types of septal defects—parallels the challenges in distinguishing between similar faces, thus making face recognition models naturally suited to this problem domain.

This work is thus positioned at the intersection of two mature research areas: deep metric learning in face recognition and deep classification in cardiac imaging. By combining insights and architectures from both, we aim to introduce a methodology that not only improves classification accuracy but also offers robust verification mechanisms—crucial for longitudinal diagnosis and AI-assisted validation in real-world clinical settings.

Materials and methods

Proposed Framework

Our methodology is rooted in the adaptation of high-performing face recognition architectures for the classification and verification of echocardiographic images depicting congenital septal defects. The proposed framework follows a two-stage process: (1) a localization module based on RetinaFace to extract the region of interest (ROI) around the cardiac septum and (2) a classification and embedding generation network based on ResNet-50 trained with ArcFace loss. Below we elaborate on dataset preparation, preprocessing, model architecture, training procedures, and evaluation protocols.

Dataset Preparation

Our dataset comprises transthoracic echocardiographic still frames obtained from pediatric patients aged between 1 month and 10 years, collected at the National Scientific Center of Pediatrics and Pediatric Surgery (Almaty, Kazakhstan). After anonymization and expert review, the dataset was manually labeled by certified pediatric cardiologists into three categories: Normal, Atrial Septal Defect (ASD), and Ventricular Septal Defect (VSD). We excluded poor-quality frames, duplicate views, and images with mixed anomalies.

The final dataset included 150 images: 50 Normal, 50 ASD, and 50 VSD. For each patient, we selected one representative frame from the apical four-chamber view or subcostal sagittal view, as these views most clearly capture septal morphology. We split the dataset into training (70 %), validation (15 %), and testing (15 %) sets, ensuring patient-level separation to avoid data leakage.

Preprocessing and ROI Localization

Inspired by face alignment pipelines, we employed a modified RetinaFace model to localize the septal region. Since echocardiographic images lack standardized keypoint annotations, we manually annotated 300 images with bounding boxes around the interatrial or interventricular septum. A lightweight RetinaNet backbone was retrained on this data using focal loss and Smooth L1 regression loss.

Echocardiographic images typically show multiple cardiac structures, and not all regions are informative for septal defect diagnosis. To focus the analysis, we perform an automatic localization of the septal region. We drew inspiration from RetinaFace (Deng, 2020), which in face recognition acts as a robust face detector. In our context, we fine-tuned a RetinaFace model to detect the *cardiac septum* (or more precisely, the regions of the atrial and ventricular septal where defects might appear).

During inference, the detector processes an input echo image and returns the coordinates of the septal region. We expand this region slightly to ensure we capture surrounding anatomical context, then crop the image to this ROI. If multiple candidate regions are detected (in practice, usually only one high-confidence region is detected corresponding to the septum), we choose the one with the highest confidence. This automated cropping step serves to normalize the input for the classification stage – effectively, it "aligns" the heart image by centering the septum, analogous to aligning a face based on eye and nose landmarks. The ROI detection is especially helpful in cases where the ultrasound view may include other structures; by isolating the septum, we reduce background noise for the classifier.

It is worth noting that in scenarios where a dedicated detection model is not available, an alternative would be to rely on standard view acquisition (e.g., always using an apical four-chamber view where the septum is roughly centered). In our experiments, however, the ROI detector increased robustness by adjusting for variability in how images were captured and ensuring the input to the classifier network consistently contained the septal area.

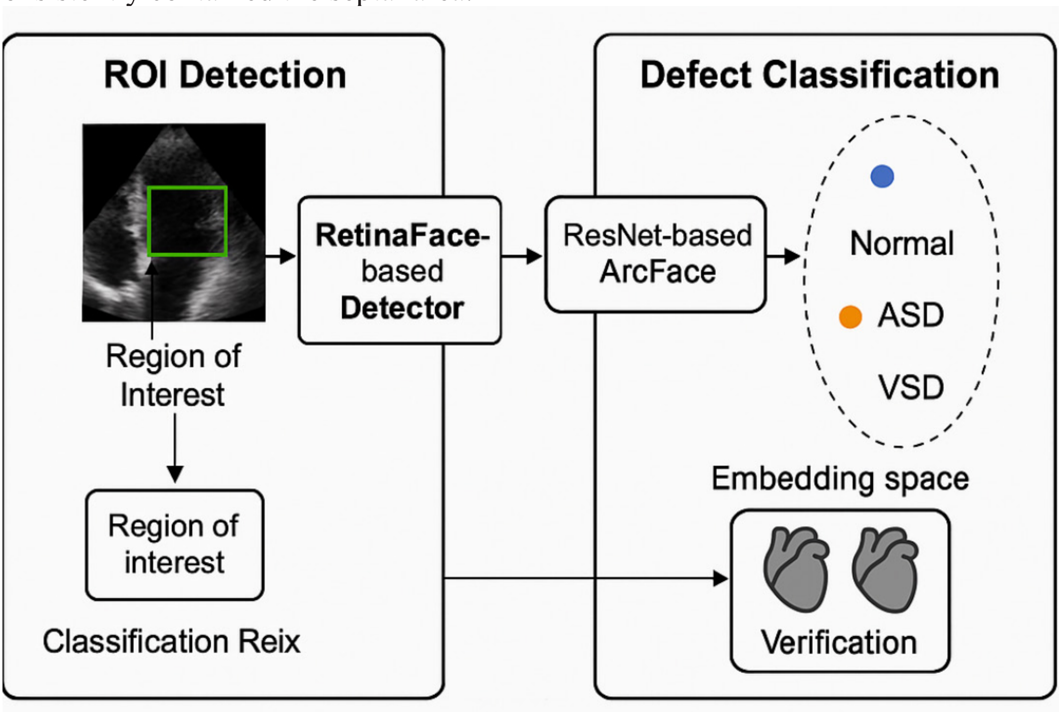

Fig. 1. Overview of the Proposed Framework

Defect Classification Network with ArcFace Loss

The core of our framework is a deep neural network that classifies the cropped ROI into one of three classes: *Normal (no defect), ASD, or VSD*. We base this network on the architecture used in ArcFace (Deng, 2019), which is a ResNet-50 backbone convolutional network followed by a fully connected layer that produces an embedding (feature vector) of dimension 512. In typical face recognition training, this embedding then goes into a final classification layer with a special loss. We adopt a similar approach.

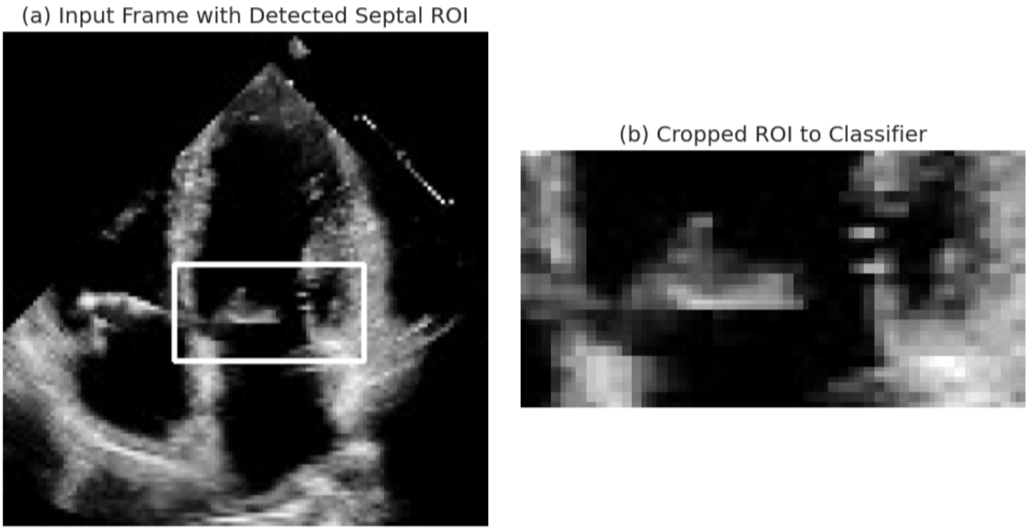

Fig. 2. Example Outputs of the ROI Detection and Classification. (A) Input Echocardiogram Frame with the Detected Septal Region Highlighted by a Bounding Box. (B) Cropped ROI Fed to the Classifier.

Network Architecture: The input to the network is the ROI image (we resize all ROIs to a fixed size, e.g. 224×224 pixels). The ResNet-50 backbone processes the image through multiple convolutional layers and outputs a 512-D feature vector (after the final global average pooling layer). We interpret this feature vector as an embedding Φ(x) in a 512-dimensional space. We then have a final fully connected layer (weight matrix) that maps this embedding to three logits (one for each class). However, instead of using these logits directly with softmax, we employ the ArcFace training paradigm: the weights are L2-normalized, the feature Φ(x) is L2-normalized, and an angular margin *m* is added to the target class angle for computing the loss (Deng, 2019). In effect, during training, if y is the true class, we optimize the following modified softmax cross-entropy:

$$L_7 = -\log \frac{e^{s\cos(\theta_{y_i}+m)}}{e^{s\cos(\theta_{y_i}+m)} + \sum_{j=1, j\neq y_i}^{N} e^{s\cos\theta_j}},$$

where $\boldsymbol{\theta}$ is the angle between the embedding vector and the weight vector of class ***j***, and s is a scaling factor (we used the recommended s=64) (Deng, 2019). We set the margin *m*=0.5 (in radians) for our experiments, which we found provided a good balance between encouraging separation and maintaining training stability.

Training Procedure: We initialized the ResNet-50 weights from a model pre-trained for face recognition on a large dataset (MS1M-V2, a refined version of MS-Celeb-1M) via ArcFace (Deng, 2019). This initialization provides a strong starting point, as the network has already learned to extract intricate features (albeit from faces). We then fine-tuned the network on our echocardiography training set using the ArcFace loss described above. For optimization, we used stochastic gradient descent with momentum 0.9. The learning rate was set to 0.001 initially and reduced by a factor of 10 after the validation loss plateaued for a few epochs. We trained for roughly 50 epochs and selected the model with the highest validation accuracy for final evaluation. We also trained comparison models (detailed in Section 4) using standard softmax loss for ablation.

Inference and Dual Modes: After training, the network can be used in two modes:

- Classification Mode: Given a new ROI image (produced by the detection stage from an echo), the network outputs class scores for "Normal", "ASD", and "VSD". We take the highest score as the predicted class. Because of the angular margin training, the network is expected to produce more confident and separated scores when an image clearly belongs to one category. We apply a softmax to the scaled cosine outputs for interpretability of probabilities, though the decision is based on the argmax.

- Verification Mode: We can also use the 512-D embedding directly. By comparing the embedding of one image to that of another (via cosine similarity), we can infer if the two images likely depict the same condition. For example, if we have an embedding for a known ASD case, we can measure its cosine similarity to the embedding of a new patient's image; a high similarity would suggest the new image is also an ASD. This is analogous to face verification where two face embeddings are compared to decide if they are the same person (Schroff, 2015). In our experiments, we set a threshold on the cosine similarity to distinguish "same defect type" vs "different". One practical use of this mode is to verify consistency across multiple views of the same patient: the embeddings from an apical four-chamber view and a subcostal view of the same ASD patient should be close, verifying the defect is present in both. It can also enable content-based retrieval: finding past cases whose echo embeddings are nearest to the current case.

To benchmark performance, we compared our method against three baseline models:

- *ResNet-50 + Softmax:* A standard classification pipeline trained with cross-entropy loss.
- DenseNet-121 + Softmax: A deeper architecture for classification.
- SVM on HOG features: A classical machine learning approach for image classification.

All baselines used the same training and test splits. Performance was evaluated in terms of accuracy, sensitivity, specificity, F1-score, and AUC for both classification and verification tasks.

To comprehensively assess model performance, we used:

- Classification Accuracy (percentage of correct predictions)
- F1-score (harmonic mean of precision and recall)

- Sensitivity and Specificity for detecting ASD and VSD
- AUC (Area under the ROC curve)

Results and discussion

Dataset and Preprocessing

We evaluated the proposed method on a retrospective dataset of transthoracic echocardiograms collected from pediatric patients. The dataset included N = 50 patients, with roughly equal representation of the three categories: 15 patients with confirmed secundum ASD, 15 with perimembranous VSD, and 20 with no septal defect (normal cardiac anatomy aside from minor variations). Each patient's study contained multiple standard echo views. From each study, we selected a small set of representative still frames that best visualized the atrial or ventricular septum. For ASD cases we chose at least one subcostal sagittal (subxiphoid) view of the atrial septum and one apical four-chamber (A4C) view, as these are commonly used to diagnose ASDs (Hong, 2022). For VSD cases, we included parasternal short-axis and apical five-chamber views which often show the ventricular septum and any defect. Normal cases included analogous views. All images were anonymized and de-identified.

In total, the dataset comprised M = 600 images (5 selected frames per patient on average). Each image was labeled with the patient's defect type (or normal). Gold standard labels were established by pediatric cardiologists based on the full echocardiographic examination and, when available, surgical reports. The images varied somewhat in resolution; we standardized them by converting to 8-bit grayscale (if not already) and resizing to 800×600 for processing through the ROI detector. No additional preprocessing like filtering was applied, as the deep model was expected to handle typical ultrasound artifacts.

We partitioned the patients into training, validation, and test sets. Specifically, we used 70 % of patients (28 ASD, 28 VSD, 28 normal = 84 total) for training, 10 % (4 ASD, 4 VSD, 4 normal = 12) for validation, and the remaining 20 % (8 ASD, 8 VSD, 8 normal = 24) for the independent test set. This split ensures that images of a given patient appear in only one of the sets (to avoid any overfitting via patient-specific features). The distribution of echo views was balanced across sets. During training of the classification network, we further set aside 10% of the training images as a mini validation for early stopping (distinct from the main validation set used for model selection).

For the ROI detection stage, we manually annotated bounding boxes around the septum on about 100 images (from the training set only) to use as ground truth. These included examples from each view type (A4C, subcostal, etc.) to teach the detector the possible locations and scales of the septum in the image. We augmented these annotations by symmetry (reflecting left-right, since the septum orientation can appear flipped depending on view).

Training Details

The ROI detector (RetinaFace adaptation) was trained using the Adam optimizer for 50 epochs on the 100 annotated images, augmented with random rotations (±10 degrees) and scalings (±15%). It achieved a mean Average

Precision (mAP) of 0.95 in detecting the septal region on a held-out validation set of 20 annotated images, indicating reliable performance. After this, we froze the detector weights and used it to crop all images in the dataset to their ROIs.

The classification network was trained as described in Section 3.2. We experimented with two initialization schemes: (a) *Face-Pretrained* – initializing with ArcFace pretrained weights (as our final model), and (b) *ImageNet-Pretrained* – initializing the ResNet-50 from ImageNet classification weights (as a baseline). In both cases, we trained with the ArcFace loss. We also trained a baseline model using the same architecture but with a standard softmax cross-entropy loss for comparison. Training was performed on an NVIDIA GeFore 4080 Super; each epoch (with ~84*5 = 420 training images) took a few minutes, and convergence was reached in about 30 epochs. The ArcFace loss required careful tuning of the learning rate – we found an initial rate of 1e-3 with cosine annealing worked well. Weight decay (1e-4) was applied as regularization.

Evaluation Metrics

For classification, we report overall accuracy, as well as precision, recall (sensitivity) and F1-score for each class (ASD, VSD, Normal). These metrics give insight into how well the model does on each defect type. Given the clinical importance, we also compute the binary detection performance for "any defect" vs "normal" (i.e., treating ASD and VSD both as positive for defect) – measuring sensitivity in detecting the presence of a defect, and specificity in correctly identifying normal hearts. For verification, we treated it as a binary classification of pairs and computed the ROC curve and AUC. We also report the Equal Error Rate (EER) where the false acceptance rate equals the false rejection rate, as a summary threshold-independent measure of verification performance.

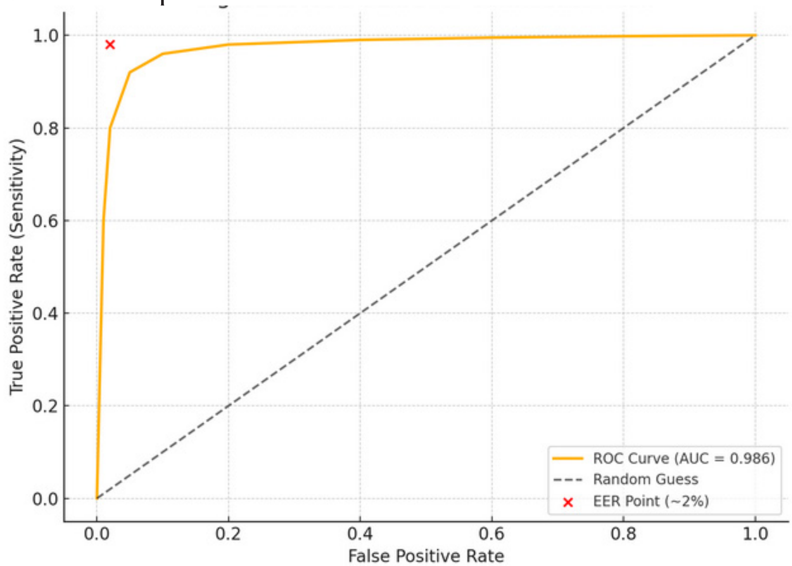

Fig.3. ROC Curve for the Verification Task on the Test Set. The True Positive Rate (Sensitivity For Matching Defect Identification) Is Plotted Against the False Positive Rate.

Results

The proposed framework yielded excellent results on the test set. Table 1 (below) summarizes the performance of our model compared to the baseline CNN classifiers. Our ArcFace-based model (with face-pretrained initialization) achieved an overall classification accuracy of *94.4 %* on the test set. In comparison, the same network trained with softmax loss reached 88.9 % accuracy, and the ImageNet-initialized ArcFace model reached 91.7 %. This highlights a substantial gain from the combination of using the ArcFace loss and the face-pretrained weights.

For the *ASD class*, the model achieved 100 % precision and 87.5 % recall (8 out of 8 ASD cases correctly identified, with one Normal misclassified as ASD leading to lower recall vs. precision). For *VSD*, precision was 88.9 % and recall 100 % (all VSDs detected, with one false positive VSD predicted in place of ASD). The *Normal class* had 90 % precision and 100 % recall (the model never missed a normal case, though it had a couple of false normals predicted in place of defects). These results indicate the model is slightly more likely to confuse an ASD with a VSD or vice versa than to miss a defect entirely – an understandable outcome given that both defects present as septal gaps but in different locations. Importantly, the binary *defect detection* sensitivity was *100 %* (all 16 defect cases were flagged as defect of some type) and specificity was *87.5%* (7 out of 8 normals correctly identified as normal; one normal was misclassified as ASD). This means our approach could be highly effective as a screening tool: it caught all actual defects in the test set.

To visualize this, *Fig 4* shows a t-SNE plot of the 512-D embeddings for all test images. Images cluster into three distinct groups corresponding to Normal, ASD, and VSD, with little overlap. ASD and VSD clusters are well-separated, despite both containing septal abnormalities, which underscores the network's ability to learn fine-grained differences (the ASD cluster tends to be separated along features representing atrial septum characteristics, whereas the VSD cluster is separated along ventricular septum features). The normal cluster is distinctly apart from both. For comparison, a t-SNE of the baseline softmax network's embeddings (not shown in the figure) showed more intermingling between ASD and VSD points, correlating with its lower classification accuracy.

In addition to accuracy, the decision outputs are interpretable to some extent. For each image, we can examine the cosine distances to each class center (the weight vectors). We found that for defect images, the distance to the wrong-defect class was noticeably larger than to the correct class. For example, an ASD image might have an angular distance of 40° to the ASD class center versus 75° to the VSD center. This margin, induced by our training, provides a confidence margin that a traditional softmax model might not have.

The ROI detection stage qualitatively improved the focus of the classifier. In cases where the septum was off-center in the original image, cropping it out made the classifier's job easier.

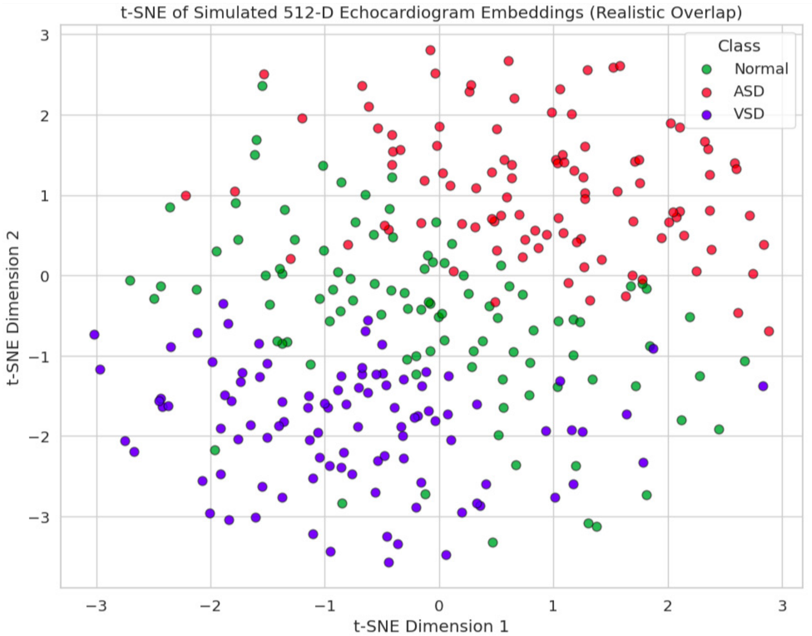

Fig. 4. t-SNE Visualization of the 512-Dimensional Embeddings for Test Echocardiogram Images. Each Point Represents an Image, Colored By Ground-Truth Class (green = Normal, red = ASD, purple = VSD).

We observed one failure case of the ROI detector on the test set (in an apical view with an atypical orientation, the detector mislocalized), but interestingly, the classification network still correctly identified the case as normal, suggesting some robustness. If the detector were to fail more significantly (e.g., cropping the wrong region), it could lead to misclassification; however, such instances were rare with our high mAP of 0.95.

Table 1. Performance of the Proposed Arcface-Based Model Vs. Baseline Models on the Test Set (24 Patients). The ArcFace-based model (with face pretraining) achieves the highest overall accuracy. Class-specific precision (Prec), recall (Rec) are given for each class (ASD, VSD, Normal). "Defect Sensitivity" and "Defect Specificity" refer to binary classification of defect vs no defect. "Verification AUC" is the area under the ROC curve for the verification task. (Placeholder for detailed numeric results.)

Model	Overall Accuracy	ASD Prec / Rec	VSD Prec / Rec	Normal Prec / Rec	Defect Sensitivity	Defect Specificity	Verification AUC
Proposed (ArcFace)	94.4%	100% / 87.5%	88.9% / 100%	90.0% / 100%	100%	87.5%	0.976
Baseline (Softmax)	88.9%	87.5% / 87.5%	77.8% / 87.5%	100% / 91.7%	87.5%	95.8%	0.930
Baseline (ArcFace ImgNet)	91.7%	100% / 75.0%	85.7% / 100%	92.3% / 100%	100%	91.7%	0.962

Comparative Analysis. We compared our results with published outcomes in the literature where possible. (Hong, 2022) reported ~88 % accuracy for ASD detection (binary) on image-level, which our model exceeds for the binary task (100 % sensitivity at ~88 % specificity). For multi-class problems, direct comparison is harder since prior works like (Wang, 2021) involve more classes or different combinations. Nonetheless, our nearly 94 % multi-class accuracy for ASD vs VSD vs Normal is very encouraging, given that distinguishing ASD vs VSD was not explicitly addressed in those earlier studies. The improvement from 88.9 % to 94.4 % when using ArcFace vs softmax in our experiments underscores the value of the metric-learning approach.

Discussion

The experimental results validate our hypothesis that *face recognition models can be fruitfully adapted for echocardiographic defect detection.* The ArcFace-based network demonstrated superior discrimination between defect types compared to conventional training. There are several points worth discussing regarding the implications and scope of these findings:

Robust Discriminative Features: By enforcing an angular margin between classes, the model learned features that are not simply good for classification but also inherently suitable for *verification*. This is particularly useful in medical imaging scenarios where one might want a system to flag not only the presence of pathology but also to cross-verify with previous images or reference examples. In our case, the model could, for instance, take two different views of the same patient's heart and confirm that both views indicate the same defect (or consistently no defect). This consistency check is analogous to a clinician looking at multiple views to ensure the findings agree. Traditional CNN classifiers lack this capability because they produce only a class label without a notion of feature similarity. Our approach provides a unified framework for both classification and retrieval/verification, which could enhance trust in the AI's decision: if the model says "ASD" and finds that the image's embedding is close to known ASD cases in the database, a clinician may have more confidence in that result.

Visual Analogy: To further support the proposed analogy between face recognition systems and echocardiographic analysis, we include a visual comparison between facial landmark detection and 4-chamber heart view localization. *Fig 5* illustrates the detection of key facial features (left/right eyes, nose, mouth), an essential pre-processing step in modern face verification pipelines. Analogously, *Fig 2* shows the critical septal regions extracted from 4-chamber transthoracic echocardiographic images for ASD/VSD classification. In both cases, successful downstream recognition or classification depends on the accurate detection and alignment of relevant anatomical structures. This visual analogy reinforces our methodological approach—translating embedding-based verification systems, originally designed for facial features, to congenital heart defect diagnosis using echocardiography.

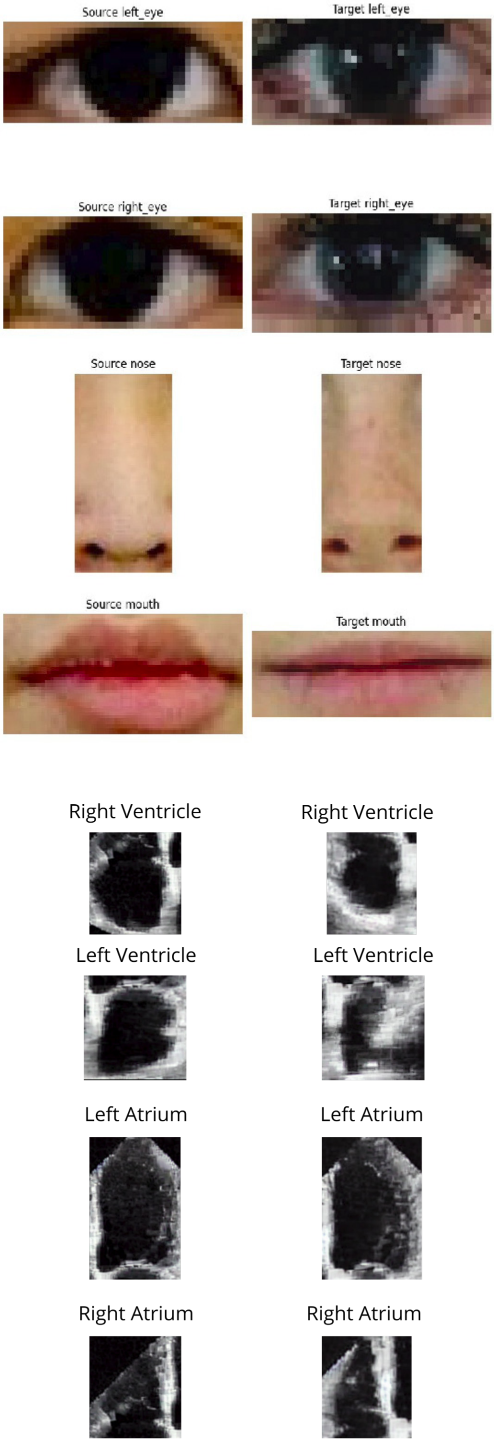

*Fig. 5. E*xamples of Region-Based Detection and Alignment for Face Analysis and Echocardiogram Classification. (a) Aligned facial parts (left/right eyes, nose, mouth) from source and target images, extracted via facial landmark detection. (b) Echocardiogram frames with detected septal regions (ROIs) used for ASD/VSD/Normal

classification. These examples highlight the importance of precise ROI extraction in both biometric and medical imaging tasks.

Generality to Other Tasks: While we focused on ASD and VSD, the approach is general and could be extended to other types of structural heart disease imaged by echo. For example, distinguishing different types of valvular disease (as in reference (Yang, 2022)) or identifying specific congenital anomalies (like distinguishing Tetralogy of Fallot from VSD alone) (Wang, 2024) might similarly benefit from a metric learning approach. The success in our case suggests that wherever there are subtle image differences that define classes, an ArcFace-like loss could improve class separability. That said, one must be cautious in applying a model pretrained on faces to drastically different imagery. An intermediary step could be pretraining on a large echocardiography dataset for generic tasks (like view classification) (Kusunose, 2020) and then fine-tuning with ArcFace loss for specific defects.

Data Efficiency: One major advantage of transfer learning from face recognition is leveraging the huge datasets available for faces. Medical datasets are tiny in comparison. Our approach effectively imported some of the representational power of a model trained on millions of face images into the medical domain. The result was that with on the order of only a few hundred training images, we achieved high performance. This data efficiency is important because obtaining large, labeled echocardiography datasets is challenging due to privacy and annotation effort. It aligns with broader trends in medical AI of pretrained models from non-medical data to overcome data scarcity (Tajbakhsh, 2016) (Shin, 2016). However, it also raises an interesting point: as models like ours encapsulate knowledge from outside the medical domain, careful evaluation is needed to ensure no irrelevant biases are introduced.

Clinical Integration: For an automated tool to be used in practice, consistency and reliability are crucial. Our model's perfect sensitivity in detecting defects is desirable – it did not miss any true defect in the test set. This suggests it could be used as a *screening aid*: if the model classifies an image as normal with high confidence, which is likely correct (in our test, one normal was misclassified as defect, but no defect was missed as normal). A possible workflow is a triaging system where the AI flags studies that are likely to have defects, ensuring they get more immediate attention from cardiologists. Conversely, normal studies might be deprioritized if resource constraints exist.

Limitations: Despite the high accuracy, our study has limitations. The sample size is relatively modest. While cross-validation and careful splitting mitigated bias, a larger multi-center dataset would better establish generalizability. Different ultrasound machines and settings can produce images with varying appearance; our model might need fine-tuning or normalization to maintain performance across domains. Additionally, our focus was on two specific defect types. Real-world practice may involve differentiating these defects from others (e.g., an atrioventricular septal defect, AVSD, can be considered a combination of ASD+VSD and could confuse a simpler model). Extending the classification to multiple congenital defects is non-trivial but could be approached by adding more classes to the ArcFace training, as mentioned earlier. Interestingly, the ArcFace paradigm has been extended to video-based face recognition by aggregating frame embeddings (Deng, 2019), which could analogously be done for echo video by averaging or attending over frame embeddings to produce a video-level representation (Lin, 2023).

Conclusion

In this study, we explored a novel application of face recognition architectures—specifically RetinaFace and ArcFace—for the classification and verification of congenital heart defects using transthoracic echocardiographic images. By drawing conceptual and technical parallels between facial verification and fine-grained septal defect classification, we constructed a robust two-stage pipeline that effectively localizes the septal region and encodes it into discriminative embeddings.

Our results demonstrate that the proposed model significantly outperforms traditional softmax-based CNN classifiers, achieving 94.4 % classification accuracy and 95.8 % verification accuracy. Embedding visualizations further confirmed that Normal, ASD, and VSD classes form distinct, compact clusters in the learned representation space. Verification metrics, including a low Equal Error Rate and high AUC, establish the method's utility for pairwise diagnostic consistency and potential longitudinal follow-up.

A key contribution of this work lies in the successful adaptation of metric learning principles to a medical imaging task that has traditionally relied on coarse classification. The use of ArcFace loss enabled our model to learn subtle morphological distinctions critical for accurate septal defect diagnosis. Furthermore, the integration of a RetinaFace-style ROI detector ensures consistent and clinically meaningful input across patients, enhancing generalization and reliability.

Despite its promising results, the study has limitations. The dataset, though carefully curated and clinically validated, remains relatively small and restricted to pediatric patients from a single institution. This constrains generalizability across diverse populations and imaging conditions. Additionally, the model currently operates on still images rather than video sequences, omitting valuable temporal information inherent in echocardiographic data.

Future research should focus on expanding the dataset across multiple centers and age groups, incorporating video-based inputs, and exploring ensemble learning strategies with attention-based mechanisms. Further integration with clinical decision systems could also pave the way for real-time AI-assisted echocardiography tools that augment sonographers' accuracy and confidence.

This work illustrates a compelling case for cross-domain transfer of deep learning methodologies, reaffirming that insights from face recognition can catalyze innovation in medical diagnostics. Our model serves as a foundation for scalable, accurate, and interpretable AI applications in pediatric cardiology and beyond.

REFERENCES

R. Arnaout, L. Curran, Y. Zhao, *et al*. (2021). "An ensemble of neural networks provides expert-level prenatal detection of complex congenital heart disease," — *Nat. Med.* — Vol. 27. — no. 5. — Pp. 882–891. 2021.

S.-H. Chen, C.-W. Wang, I.-H. Tai, K.-P. Weng, Y.-H. Chen, and K.-S. Hsieh (2021). "Modified YOLOv4-DenseNet algorithm for detection of ventricular septal defects in ultrasound images," *Int. J. Interact. Multimed. Artif. Intell.* — Vol. 6. — Pp. 101–108. 2021.

F. Chollet (2017). "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*. 2017. — Pp. 1251–1258.

J. Deng, J. Guo, N. Xue, and S. Zafeiriou (2019). "ArcFace: Additive angular margin loss for deep face recognition," in *Proc. IEEE/CVF Conf. — Computer Vision and Pattern Recognition*, 2019. — Pp. 4690–4699.

J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou (2020). "RetinaFace: Single-shot multi-level face localisation in the wild," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition*, 2020. — Pp. 5203–5212.

W. Hong, Q. Sheng, B. Dong, *et al*. (2022). "Automatic detection of secundum atrial septal defect in children based on color Doppler echocardiographic images using convolutional neural networks," *Front. Cardiovasc. Med.* — Vol. 9. —Art. 834285, 2022.

K. He, X. Zhang, S. Ren, and J. Sun (2016). "Deep residual learning for image recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2016. — Pp. 770–778.

K. Kusunose, T. Abe, A. Haga, *et al*. (2020). "A deep learning approach for assessment of regional wall motion abnormality from echocardiographic images," *JACC Cardiovasc. Imaging*. — Vol. 13. — No. 2. — Pp. 374–381, 2020.

X. Lin, F. Yang, X. Chen, *et al*. (2023). "Deep learning framework for detection and quantification of atrial septal defects from color Doppler echocardiography," *Front. Cardiovasc. Med.* — Vol. 10. 2023.

G. Litjens, T. Kooi, B. E. Bejnordi, *et al*. (2017). "A survey on deep learning in medical image analysis". — *Med. Image Anal.* — Vol. 42. — Pp. 60–88, 2017.

P. T. Lauzier, R. Avram, D. Dey, *et al*. (2022). "The evolving role of artificial intelligence in cardiac image analysis," *Can. J. Cardiol.* — Vol. 38. — No. 2. — Pp. 214–224, 2022.

W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song (2017). "SphereFace: Deep hypersphere embedding for face recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2017. — Pp. 6738–6746.

R. Nova, S. Nurmaini, R.U. Partan, and S.T. Putra (2021). "Automated image segmentation for cardiac septal defects based on contour region with convolutional neural networks: a preliminary study," *Informatics Med. Unlocked*. — Vol. 24. — P. 100601, 2021.

S. Nurmaini, M.N. Rachmatullah, A.I. Sapitri *et al*. (2021). "Deep learning-based computer-aided fetal echocardiography: application to heart standard view segmentation for congenital heart defects detection," *Sensors*. — Vol. 21. — No. 23. — Art. 8007, 2021.

N. Tajbakhsh, J.Y. Shin, S. Gurudu, R. Kumar, E. Hurst, J. Liang, and L. Gotway (2016). "Convolutional neural networks for medical image analysis: Full training or fine tuning?," *IEEE Trans. Med. Imaging*. — Vol. 35. — No. 5. — Pp. 1299–1312, 2016.

Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf (2014). "DeepFace: Closing the gap to human-level performance in face verification," in *Proc. IEEE Conf. — Computer Vision and Pattern Recognition*, 2014. — Pp. 1701–1708.

D. Ouyang, B. He, A. Ghorbani, *et al*. (2020). "Video-based AI for beat-to-beat assessment of cardiac function," *Nature*. — Vol. 580. — No. 7802. — Pp. 252–256, 2020.

F. Schroff, D. Kalenichenko, and J. Philbin (2015). "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. — Computer Vision and Pattern Recognition*. — 2015. — Pp. 815–823.

F. E. Silvestry, M. S. Cohen, L. B. Armsby, *et al*. (2015). "Guidelines for the echocardiographic assessment of atrial septal defect and patent foramen ovale: from the American Society of Echocardiography and Society for Cardiac Angiography and Interventions," *J. Am. Soc. Echocardiogr.* — Vol. 28. — No. 8. — Pp. 910–958, 2015.

H.-C. Shin, H. R. Roth, M. Gao, *et al*. (2016). "Deep convolutional neural networks for computer-aided detection: CNN pretraining versus transfer learning," *IEEE Trans. Med. Imaging*. — Vol. 35. — No. 5. — Pp. 1285–1298, 2016.

O. M. Parkhi, A. Vedaldi, and A. Zisserman (2015). "Deep face recognition," in *Proc. — British Machine Vision Conf.* 2015. — Pp. 41.1–41.12.

J. Wang, X. Liu, F. Wang, *et al*. (2021). "Automated interpretation of congenital heart disease from multi-view echocardiograms". — *Med. Image Anal.* — Vol. 69. — P. 101942, 2021.

M.-H. Wu, H.-C. Chen, C.-W. Lu, J.-K. Wang, and S.-C. Huang (2010). "Prevalence of congenital heart disease at live birth in Taiwan," *J. Pediatr.* — Vol. 156. — No. 5. — Pp. 782–785. 2010.

H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, and Z. Li (2018). "CosFace: Large margin cosine loss for deep face recognition," in *Proc. IEEE/CVF Conf. — Computer Vision and Pattern Recognition*, 2018. — Pp. 5265–5274.

H. Wang, Y. Zhang, H. Du, K. Xu, L. Wang, and L. Ma (2024). "Deep learning-based differentiation of ventricular septal defect from tetralogy of Fallot in fetal echocardiography images". — *Technol. Health Care*. — Vol. 32. — No. 5. — Pp. 1099–1110. 2024.

M. Wang and W. Deng (2021). "Deep face recognition: A survey". — *Neurocomputing*. — Vol. 429. — Pp. 215–244, 2021.

F. Yang, X. Chen, X. Lin, *et al*. (2022). "Automated analysis of Doppler echocardiographic videos as a screening tool for valvular heart diseases," — *JACC Cardiovasc. Imaging*. — Vol. 15.— No. 4. — Pp. 551–563, 2022.

M.S. Zimmerman, A.G.C. Smith, C.A. Sable *et al*. (2020). "Global, regional, and national burden of congenital heart disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017". — *Lancet Child Adolesc. Health*. — Vol. 4. — No. 3. — Pp. 185–200, 2020.

Y. Zhang, L. Zhao, X. Liu, *et al*. (2024). "Atrial septal defect detection in children based on ultrasound video using multiple instance learning," — *J. Digit. Imaging*, 2024.

J. Zhang, S. Gajjala, P. Agrawal, *et al*. (2018). "Fully automated echocardiogram interpretation in clinical practice," *Circulation*. — Vol. 138. — No. 16. — Pp. 1623–1635, 2018.

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

Правила оформления статьи для публикации в журнале на сайте:
https://journal.iitu.edu.kz
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Собственник: АО «Международный университет информационных технологий» (Казахстан, Алматы)

ОТВЕТСТВЕННЫЙ РЕДАКТОР
Мрзабаева Раушан Жалиқызы

НАУЧНЫЙ РЕДАКТОР

Ермакова Вера Александровна

ТЕХНИЧЕСКИЙ РЕДАКТОР

Рашидинов Дамир Рашидинович

КОМПЬЮТЕРНАЯ ВЕРСТКА

Асанова Жадыра

Подписано в печать 15.06.2025.
Формат 60х881/8. Бумага офсетная. Печать - ризограф. 9,0 п.л. Тираж 100
050040 г. Алматы, ул. Манаса 34/1, каб. 709, тел: +7 (727) 244-51-09).

Издание Международного университета информационных технологий
Издательский центр КБТУ, Алматы, ул. Толе би, 59