The object of this research is to develop a data-efficient pipeline for the detection of atrial septal defects (ASDs) using echocardiographic images. ASDs are common congenital heart defects that can lead to serious health issues if not diagnosed early. Rising mortality rates due to undetected ASDs highlight the urgent need for improved diagnostic methods. To address the problem of limited annotated medical data hindering accurate detection models, this study fine-tuned the SegFormer model for precise segmentation of cardiac structures in echocardiography images, focusing on the four-chamber heart view essential for ASD detection. By integrating SegFormer with the YOLOv7 detection model, known for real-time object detection, the ASD regions within the segmented heart structures were accurately identified. This cross-referencing ensures anatomically accurate diagnoses and reduces false positives. The study results demonstrate that despite limited data, the integrated method achieves high accuracy and speed, outperforming traditional models. This improvement is explained by the synergy between SegFormer's transformer-based segmentation and YOLOv7's efficient detection capabilities. The distinctive feature of our approach is the successful integration of these models in a data-efficient manner, enabling effective ASD detection even with scarce data. The scope of practical use includes deployment in clinical settings with limited resources, requiring only echocardiographic equipment and basic computational resources. By providing clinicians with a reliable tool for ASD detection, the study supports timely interventions in pediatric cardiology, ultimately improving patient outcomes and enhancing care consistency

Keywords: deep learning, SegFormer, echocardiography, image segmentation, YOLOv7, data augmentation

UDC 004.93
DOI: 10.15587/1729-4061.2024.312621

DEVELOPMENT OF DATA-EFFICIENT TRAINING TECHNIQUES FOR DETECTION AND SEGMENTATION MODELS IN ATRIAL SEPTUM DEFECT ANALYSIS

Sabina Rakhmetulayeva
Corresponding author
PhD, Professor
Department of Cybersecurity,
Information Processing and Storage
Satbayev University
Satbayev str., 22, Almaty, Republic of Kazakhstan, 050013
E-mail: ssrakhmetulayeva@gmail.com

Baubek Ukibassov
MSc, Senior Lecturer
School of Digital Technologies
Narxoz University
Zhandossov str., 55, Almaty, Republic of Kazakhstan, 050035

Zhandos Zhanabekov
MSc, Senior Lecturer
School of Informational Technologies and Engineering
Kazakh-British Technical University
Tole-bi str., 59, Almaty, Republic of Kazakhstan, 050000

Aigerim Bolshibayeva
PhD, Assistant Professor
Department of Information Systems
International IT University
Manas str., 34/1, Almaty, Republic of Kazakhstan, 050000

Received date 05.08.2024
Accepted date 17.10.2024
Published date 30.10.2024

***How to Cite**: Rakhmetulayeva, S., Ukibassov, B., Zhanabekov, Z., Bolshibayeva, A. (2024). Development of data-efficient training techniques for detection and segmentation models in atrial septum defect analysis. Eastern-European Journal of Enterprise Technologies, 5 (2 (131)), 13–23. https://doi.org/10.15587/1729-4061.2024.312621*

1. Introduction

Atrial septal defects (ASDs) are one of the most prevalent congenital heart defects, affecting approximately 1–2 kids per 1 000 live births and having a profound impact on children's health during critical early developmental stages, and represent approximately 10–15 % of all congenital heart diseases [1]. If left undiagnosed, these defects can lead to severe complications such as heart failure, pulmonary hypertension, and stroke, underscoring the necessity of early detection and accurate diagnosis to mitigate these risks and improve long-term health outcomes [2].

Due to its non-invasive nature and ability to provide detailed anatomical and hemodynamic information, echocardiography (echo-CG) is widely recognized as the primary and most effective diagnostic tool for visualizing heart structures and diagnosing ASDs in pediatric patients [3]. Transthoracic echocardiography (TTE), in particular, is preferred for its safety and comprehensive diagnostic capabilities [4]. Echocardiography is indispensable not only for initial diagnosis but also for the ongoing management and post-treatment monitoring of these patients, ensuring that complications are promptly detected and treated [5].

However, the effectiveness of echocardiography is highly dependent on the expertise of the clinician performing the examination. Research consistently shows [6, 7] that the accuracy of echocardiographic assessments is heavily influenced by the clinician's expertise, with studies indicating that diagnostic variability can lead to errors in up to 20 % of cases, depending on the complexity of the defect [8]. This variability underscores the critical need for specialized training and highlights the potential role of artificial intelligence (AI) and computer vision (CV) methods in providing more consistent and objective assessments. Although AI has demonstrated promise in assisting the clinicians in decision making, expert oversight remains essential to ensure the reliability of diagnoses.

Therefore, studies that are devoted to enhancing the diagnostic accuracy of echocardiography in ASDs through the integration of artificial intelligence and computer vision methods while ensuring expert clinical oversight are of significant scientific relevance. These studies not only aim to reduce diagnostic variability but also strive to improve early detection and treatment outcomes for pediatric patients with ASDs. By combining advanced technological tools with clinical expertise, it is possible to move toward more consistent, objective, and reliable assessments, ultimately improving patient care and long-term health outcomes.

2. Literature review and problem statement

Echocardiography (echo-CG) imaging is commonly represented as two-dimensional images, allowing for the application of extensive computer vision and image processing techniques.

In [9], a UNet model was employed to segment echocardiography images into four critical views – A4c, A2c, PLAX, and PSAX – facilitating the identification of weak areas and aiding in the detection of diseases such as hypertrophic cardiomyopathy and cardiac amyloidosis. While this study showcased the potential of CNNs in echo-CG segmentation, it primarily focused on a limited set of views and did not address generalization to more diverse datasets.

In a subsequent study, [10] applied CNNs to short video clips recorded during echocardiography to determine which parts of the heart were being diagnosed, such as the subcostal four-chamber or subcostal inferior vena cava views. This approach improved automation in view classification of echocardiography videos. However, the study faced challenges in accurately capturing global contextual information, a common limitation of CNN architectures that focus on local features.

In the paper [11] proposed ResNet, introducing residual learning to ease the training of deeper networks. While ResNet has achieved remarkable success in various image recognition tasks, its application in echocardiography has been limited by the complex and variable nature of medical images, which require models to capture both local and global features. The inability to effectively capture global context can hinder the model's performance in accurately interpreting echo-CG images.

Study [12] developed the R-CNN architecture for object detection, which has been adapted for detecting specific structures in medical images. Despite its effectiveness, R-CNN-based approaches in echocardiography have been hindered by the need for extensive labeled data and significant computational resources, making them less practical in clinical settings.

In [13], a deep regression neural network was proposed to identify heart anomalies from echocardiography images. While the model showed promising results in anomaly detection, it required extensive preprocessing and manual annotation, which can be time-consuming and introduce variability in clinical environments. This reliance on preprocessing steps limits the scalability and robustness of the model.

The EchoNet-Dynamic model introduced by [14] segments echocardiography images into structural parts and predicts ejection fraction and cardiomyopathy through video-based echocardiograms. Although this model advanced the integration of segmentation and diagnosis tasks, it still relied heavily on CNN architectures. The inherent limitations of CNNs in capturing global dependencies may affect the model's ability to generalize across diverse patient populations.

Despite these advancements, there are still unresolved questions related to the limitations of CNN-based approaches in echocardiography image segmentation. One significant challenge is the need for extensive preprocessing [15], which can be both time-consuming and prone to variability due to differences in imaging protocols and equipment. Additionally, these models often struggle with generalization in heterogeneous clinical environments, as documented in studies showing decreased performance on diverse datasets [16]. This lack of robustness hampers the widespread adoption of CNN-based models in clinical practice.

The reasons for these limitations can be attributed to the intrinsic design of CNNs. While they are adept at capturing local features through convolutional operations, they may not effectively capture global context and long-range dependencies across the image [17]. This limitation makes it difficult for CNN-based models to generalize well to new, unseen data, especially in complex medical imaging tasks where global context is crucial for accurate interpretation.

An option to overcome these relevant difficulties is to leverage transformer architectures, which offer superior capability in capturing global context and dependencies. The SegFormer model, introduced by [18], combines transformer architectures with multilayer perceptron (MLP) decoders to address the generalization and preprocessing challenges faced by CNNs. SegFormer does not require complex preprocessing steps and has demonstrated strong performance on standard segmentation benchmarks without the need for positional embeddings, making it more adaptable to various image types.

All this allows to argue that it is appropriate to conduct a study devoted to the application of transformer-based models like SegFormer in echocardiography image segmentation. By exploring this avenue, let's aim to address the limitations of CNN-based approaches and improve the accuracy and generalization of echocardiography image analysis. Such a study could pave the way for more robust and efficient diagnostic tools in cardiology.

3. The aim and objectives of the study

The aim of this study is to integrate a machine learning-based solution into the diagnostic process, providing medical professionals with a tool to improve the detection and diagnosis of atrial septal defects (ASD) using echocardiographic images. This approach is designed to enhance early diagnosis and mitigate the severe health risks associated with undiagnosed ASDs in pediatric patients.

To achieve this aim, the following objectives are pursued:

– to develop an efficient segmentation model using the SegFormer architecture, optimized for identifying key cardiac structures in echocardiography images;

– to implement and integrate the YOLOv7 detection model to identify ASD regions within the segmented heart structures.

4. Materials and methods of research

4. 1. Object and hypothesis of the study

The object of this study is the detection of atrial septal defects (ASDs) in echocardiographic images. The subject of this study is to develop and validate a data-efficient diagnostic pipeline for the detection of atrial septal defects (ASDs) in echocardiographic images. This pipeline integrates the SegFormer transformer-based segmentation model with the YOLOv7 object detection model to enhance the accuracy and efficiency of ASD diagnosis in resource-constrained clinical settings.

The main hypothesis of the study is that combining the SegFormer model for precise segmentation of cardiac structures with the YOLOv7 model for efficient ASD detection will significantly improve the accuracy and speed of ASD diagnosis in echocardiographic images, even when trained on a limited dataset. This integrated approach is expected to outperform traditional convolutional neural network (CNN)-based models in both segmentation and detection tasks.

In conducting this research, it is possible to assume that the limited dataset of 331 echocardiographic images, enhanced through data augmentation techniques, is sufficiently representative to train models that generalize well to new, unseen data. Let's also presume that the echocardiographic images are of adequate quality and consistency, allowing the models to effectively perform segmentation and detection tasks. The accuracy and reliability of the expert annotations provided by cardiologists and cardiac surgeons are taken as a given, providing a solid foundation for supervised learning. Furthermore, it is possible to assume that integrating the SegFormer and YOLOv7 models will leverage their individual strengths synergistically, resulting in improved performance over existing methods.

To focus on the core objectives, several simplifications were adopted in this study. Standardized imaging conditions and equipment settings across all echocardiographic images were assumed, which may not fully reflect the variability present in different clinical environments. Collecting the dataset from a single medical center simplified the variability that might arise from multi-center data but may limit the generalizability of the findings. Let's use data augmentation as a substitute for diversity in the dataset, acknowledging that it might not entirely capture the full range of real-world clinical data variability. Additionally, it is possible to evaluate the models using splits from the same dataset for both training and validation, which simplifies the evaluation process but may not account for variations across different patient populations or imaging devices.

4. 2. SegFormer model for echocardiographic image segmentation

In this study, let's utilize SegFormer, a transformer-based model designed for semantic segmentation tasks, to segment echocardiographic images focusing on the four-chamber view of the heart, which is crucial for detecting atrial septal defects (ASDs) [19, 20]. SegFormer integrates hierarchical transformers with lightweight multilayer perceptron (MLP) decoders, making it suitable for handling variable resolutions and complexities typical of echocardiographic imaging [21].

Unlike conventional convolutional neural networks (CNNs), SegFormer does not rely on positional encodings, allowing it to adapt to fluctuations in image quality caused by patient movement, equipment variation, or operator expertise [18]. The model's architecture divides the input echocardiographic image into overlapping patches using a hierarchical transformer encoder, as illustrated in Fig. 1. This enables local self-attention mechanisms to capture important features across multiple scales.

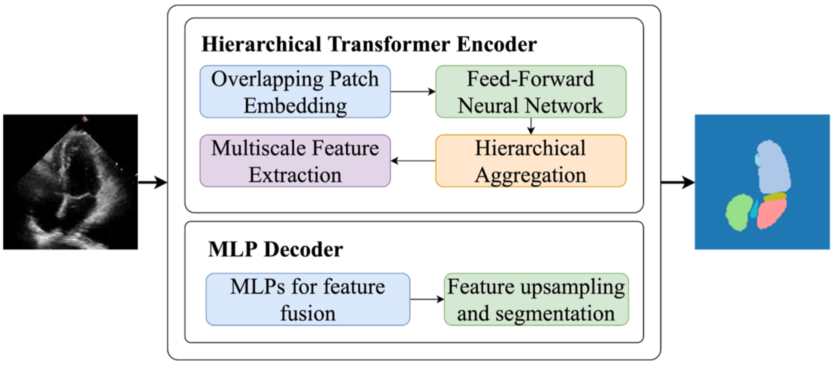

Hierarchical Transformer Encoder
Overlapping Patch Embedding
Feed-Forward Neural Network
Multiscale Feature Extraction
Hierarchical Aggregation
MLP Decoder
MLPs for feature fusion
Feature upsampling and segmentation

Fig. 1. SegFormer model scheme

Key components of the SegFormer architecture include:

– Hierarchical Feature Aggregation: Combines information from various levels of the image for accurate segmentation.

– Multiscale Feature Extraction: Operates at different resolutions to enhance feature representation.

– MLP Decoder: Fuses multiscale features and performs efficient upsampling to generate detailed segmentation maps representing distinct anatomical structures within the echocardiographic image.

For experiments, it is possible to consider multiple SegFormer variants (B0 to B5), each differing in complexity and capacity. The models were implemented using the NVIDIA/mit framework and fine-tuned on our echocardiography dataset.

4. 3. YOLOv7

Complementing the SegFormer model, let's incorporate YOLOv7, the latest iteration in the YOLO (You Only Look Once) series, known for its real-time object detection capabilities. YOLOv7 introduces advanced features such as trainable bag-of-freebies and re-parameterization techniques, which enhance detection speed and accuracy – critical factors for identifying subtle ASD-related anomalies in high-resolution echocardiographic images.

YOLOv7 employs a Cross-Stage Partial Network (CSPNet) strategy to partition the feature map of the base layer into two parts and then merges them through a cross-stage hierarchy. This split-and-merge approach allows for more efficient gradient flow through the network, improving learning efficiency and detection performance.

In this study, let's incorporate the YOLOv7 model due to its state-of-the-art object detection capabilities, particularly in real-time applications. Although ASDs can be subtle and challenging to detect on echocardiographic images, YOLOv7 has proven to be efficient in detecting small and difficult-to-distinguish features. However, by itself, YOLOv7 may not capture the fine details required for ASD detection, as its primary strength lies in detecting larger and more prominent objects. Therefore, it is possible to integrate YOLOv7 with the SegFormer model, which provides precise segmentation of heart structures. YOLOv7 is applied to the segmented images to focus its detection capabilities on potential ASD regions, improving its ability to localize subtle defects. This combination ensures a more reliable identification of ASDs by narrowing the detection to anatomically relevant areas within the segmented heart structures (Fig. 2). The integration between SegFormer and YOLOv7 involved:

Segmentation Stage: Using SegFormer to generate precise segmentation masks of the heart's anatomical structures from echocardiographic images:

– detection stage: applying YOLOv7 to identify potential ASD regions within the segmented images;

– cross-referencing: aligning detected ASD regions with the segmented heart chambers to ensure anatomical relevance.

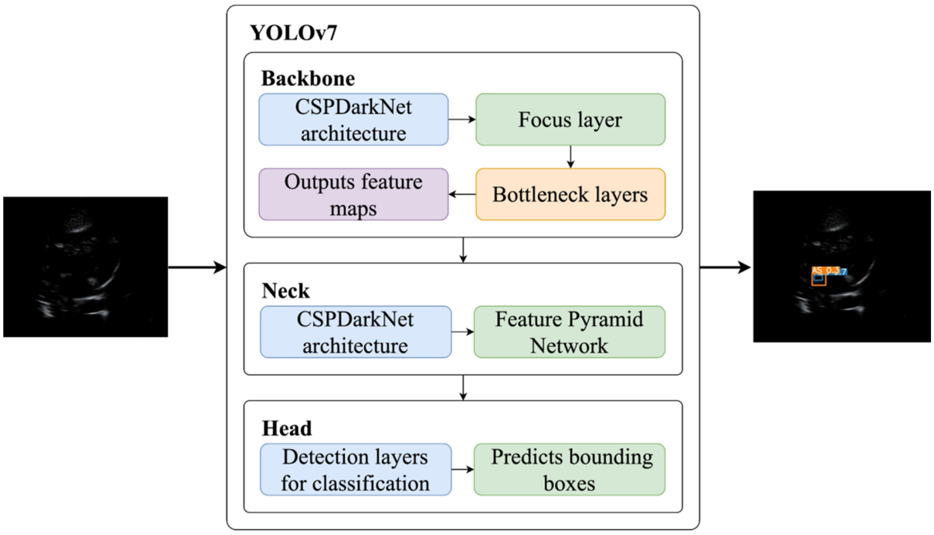


Fig. 2. YOLOv7 model scheme

This integrated pipeline was designed to enhance the overall diagnostic process by combining the strengths of both models.

4. 4. Dataset and preprocessing

The dataset used in this study consists of 331 echocardiographic images collected using a portable ultrasound device. All patients were examined by echocardiography on M5 Diagnostics ultrasound systems from MindRay at the Center for Perinatology and Pediatric Cardiac Surgery in Almaty, Kazakhstan. Each echocardiographic study was performed following standard techniques outlined in the American Society of Echocardiography (ASE) guidelines [22, 23].

The study was approved by the Republican State Enterprise on the right of economic management "Institute of Genetics and Physiology" of the Committee of Science of the Ministry of Education and Science of the Republic of Kazakhstan (approval number #13-364 from 17.11.2023). An exemption from mandatory patient consent was applied due to the de-identified nature of the data.

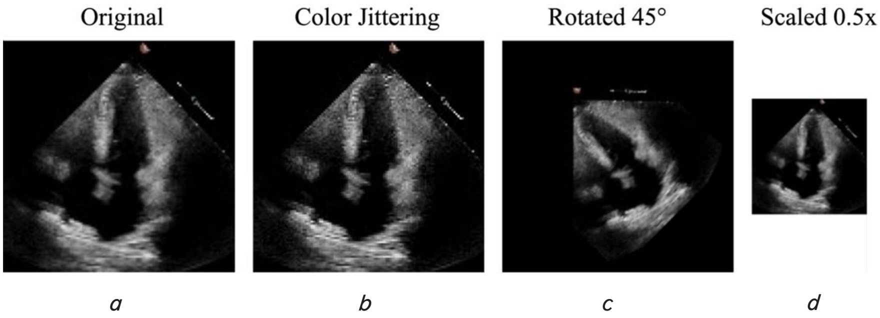


Fig. 3. Sample of image preprocessing steps: *a* – original; *b* – color jittering; *c* – rotated 45°; *d* – scaled 0,5x

The images were labeled by cardiologists and cardiac surgeons into nine classes:

– Left Ventricle (LV);

– Right Ventricle (RV);

– Right Atrium (RA);

– Left Atrium (LA);

– Atrial Septal Defect (ASD);

– Ventricular Septal Defect (VSD);

– Mitral Valve (MK);

– Tricuspid Valve (TK);

– Aortic Stenosis (AS).

To prepare the dataset for training the models, several preprocessing steps were applied:

– data anonymization: all data were strictly de-identified to protect patient confidentiality;

– image resizing: all images were resized to a consistent size of 256×256 pixels to standardize the input for the models, ensuring uniformity throughout the training process;

– pixel normalization: pixel values were normalized to a range of 0 to 1 by dividing by the maximum pixel value (255). This standardization helps the models learn more effectively by keeping input values within a consistent range.

To enhance the dataset and improve model robustness, various data augmentation techniques were employed (Fig. 3):

– color jittering: random adjustments to brightness, contrast, saturation, and hue to help the models generalize to different lighting conditions and color variations [24];

– image rotation and flipping: random rotations between −45 to +45 degrees and horizontal flipping to make the models invariant to the orientation of the heart in echocardiographic images;

– scaling: random scaling within a range of 0.8 to 1.2 of the original size to handle variations in image size and resolution from different ultrasound devices.

4. 5. Software and hardware resources

The implementation of the SegFormer and YOLOv7 models was carried out using specific software and hardware resources to ensure efficient development and execution. For programming frameworks, let's utilize PyTorch for model development and training due to its dynamic computational graph and widespread use in deep learning research. OpenCV and Pillow (PIL) were employed for image processing and augmentation tasks, providing a comprehensive suite of tools for handling image data.

Our hardware setup included an NVIDIA GeForce RTX 3080 TI GPU, which provided accelerated training capabilities essential for handling the computational demands of transformer-based models like SegFormer and real-time object detection models like YOLOv7. The system was also equipped with 32 GB of RAM to manage data loading and preprocessing efficiently, preventing bottlenecks during training.

4. 6. Experimental procedure

The experimental procedure was meticulously designed to develop and validate the integrated diagnostic pipeline comprising SegFormer and YOLOv7 models. The process began with data preparation, where it is possible to collect and anonymize echocardiographic images to protect patient confidentiality. Expert cardiologists provided annotations for the images, labeling key cardiac structures and anomalies, which were crucial for supervised learning. Let's apply preprocessing techniques such as resizing images to 256×256 pixels and normalizing pixel values to a range of 0 to 1. Data augmentation methods, including color jittering, rotation, flipping, and scaling, were employed to enhance the dataset's diversity and help the models generalize better.

In the model configuration phase, it is possible to initialize the SegFormer models (B0 to B5) with pre-trained weights, adapting them for our specific segmentation task involving echocardiographic images. The YOLOv7 model was configured with default settings optimized for small object detection, which is particularly suitable for identifying subtle anomalies like atrial septal defects.

During the training setup, let's split the dataset into training (80 %) and validation (20 %) subsets to facilitate unbiased evaluation of model performance. It is possible to define appropriate loss functions for both segmentation (e.g., cross-entropy loss) and detection tasks and selected optimization algorithms like AdamW for efficient training. Hyperparameters such as learning rate, batch size, and the number of epochs were set based on preliminary experiments and existing literature, aiming to optimize model convergence and prevent overfitting.

The model training phase involved training the SegFormer models on the segmentation task using the prepared dataset. The models learned to generate precise segmentation masks of cardiac structures, which are essential for accurate ASD detection. Subsequently, the YOLOv7 model was trained on the detection task, utilizing the segmented images from SegFormer to focus on potential ASD regions. This sequential training ensured that the detection model benefited from the enhanced input provided by the segmentation model.

Following training, it is possible to proceed with the integration of models to form the complete diagnostic pipeline. Let's combine the outputs of SegFormer and YOLOv7 by overlaying the detection results onto the segmentation masks. Cross-referencing mechanisms were implemented to align detected ASD regions with the corresponding anatomical structures identified by the segmentation model. This integration ensured that the detected anomalies were anatomically relevant, reducing the likelihood of false positives.

Finally, in the validation of proposed solutions, it is possible to evaluate the performance of the integrated pipeline on the held-out validation dataset. Let's use evaluation metrics such as Intersection over Union (IoU) to assess segmentation accuracy and precision-recall curves to evaluate detection performance. These metrics provided quantitative measures of the models' effectiveness and adequacy for the intended diagnostic application. The validation process helped fine-tune the models and confirmed the viability of our approach in a clinical context.

5. Segmentation and classification models development results

5. 1. Development of an efficient segmentation model using SegFormer

5. 1. 1. Model setup and training

SegFormer models (B0 to B5) pre-trained on the NVIDIA/mit framework were fine-tuned for this task. The experiments were conducted on an NVIDIA GeForce RTX 3080 TI with 32 GB of RAM, and wandb.ai was used to store and visualize experiment results. Each configuration was tested with multiple random seeds to ensure consistency and robustness. To ensure the consistency and robustness of the results, each configuration was tested using multiple random seeds.

The fine-tuning of the models involved varying several key parameters, including the number of epochs, which ranged from 25 to 150, and the batch size, which was tested with values of 8 and 16. Additionally, the learning rate was fine-tuned with values of 1e-6, 1e-5, 1e-4, and 1e-3, to determine the optimal setting for each model configuration. Through this process, let's aim to identify the most effective setup for fine-tuning the SegFormer models in this specific task.

The dataset was split into training (80 %) and testing (20 %) sets, ensuring representative samples from each class. The training process involved evaluating the models at regular intervals and saving the best- performing models based on validation metrics. The following metrics were used to evaluate model performance:

– mean intersection over union (IoU): measures the average overlap between predicted and true segmentations. the values of the of the overlap could vary from 0 % to 100 %;

– accuracy: overall and per-category accuracy. In the given context, the categories are segmentations from b0 to b5. The average accuracy per category varies from;

– standard deviation: measures the variation in accuracy and IoU across different runs.

5. 1. 2. Performance comparison of other architectures

A series of experiments with different configurations were conducted, and the key results are summarized in Table 1.

The graph in Fig. 4 illustrates the generally positive relationship between Mean IoU and Accuracy across the SegFormer models: as accuracy increases, Mean IoU tends to increase as well. For instance, Model B1, with the highest accuracy, also has the highest IoU, indicating that better overall classification leads to better segmentation. However, exceptions like B3, which has a higher IoU but lower accuracy, suggest that a model can perform well in segmenting regions (IoU) but still struggle with overall pixel classification.

The graph in Fig. 5 graph compares all metrics for different SegFormer models. Model B1 achieves the highest accuracy and IoU, but with higher variability in predictions, while B5 shows a good balance between IoU and accuracy with moderate consistency. B3, despite being the most consistent, performs poorly in both accuracy and IoU, making it less effective. Overall, B1 seems to be the best-performing model, offering superior segmentation performance at the cost of some prediction variability.

The Fig. 6 illustrates examples of segmented images for models. These visualizations highlight the strengths and weaknesses of each model in segmenting echocardiography images.

Table 1

Performance of SegFormer models

Model	Epochs	Batch Size	Learning Rate	Mean IoU	Accuracy	Std Dev
B0	150	8	1e-6	60.13	64.93	8.04
B1	150	8	1e-5	72.72	77.93	7.78
B2	50	8	1e-5	64.62	68.71	4.84
B3	50	16	1e-3	66.66	59.99	4.36
B4	50	8	1e-6	57.79	63.30	5.72
B5	50	16	1e-3	65.05	67.37	3.83

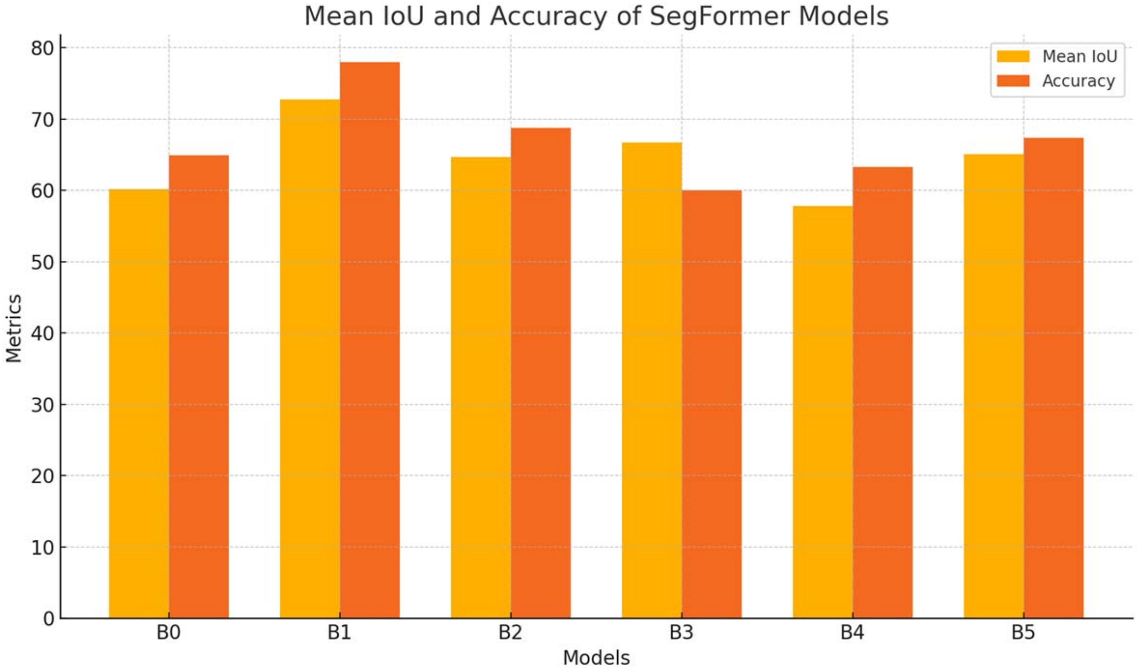


Fig. 4. Mean IoU and accuracy of SegFormer models

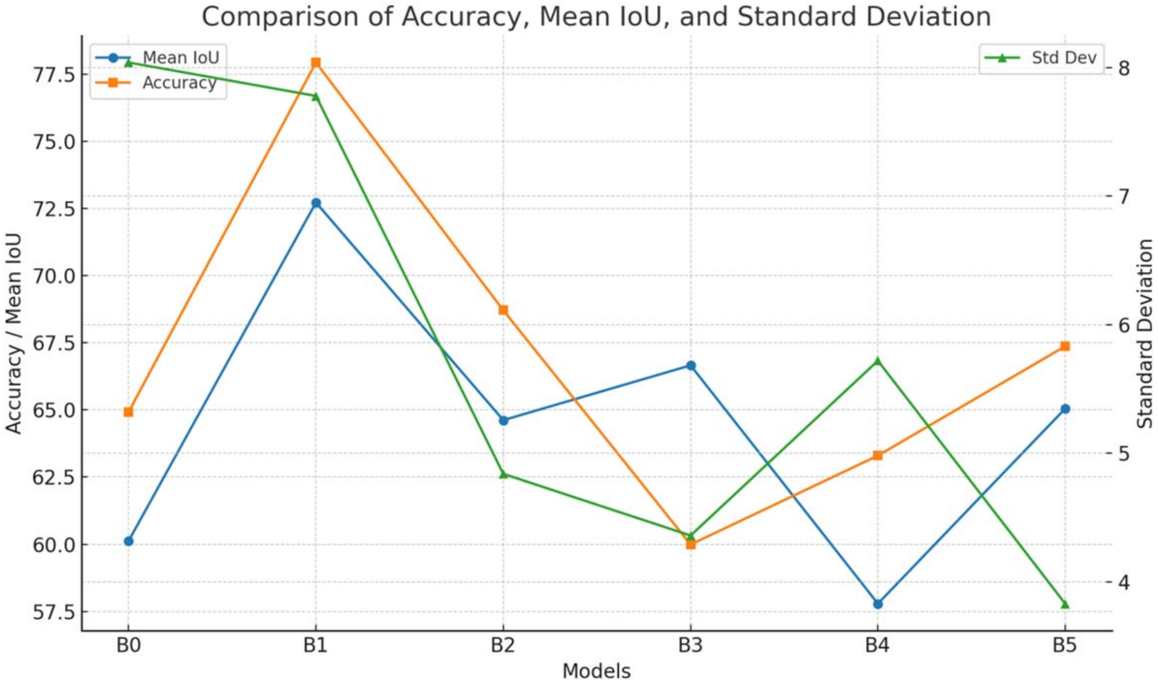


Fig. 5. Comparison of accuracy, mean IoU, and standard deviation for SegFormer models

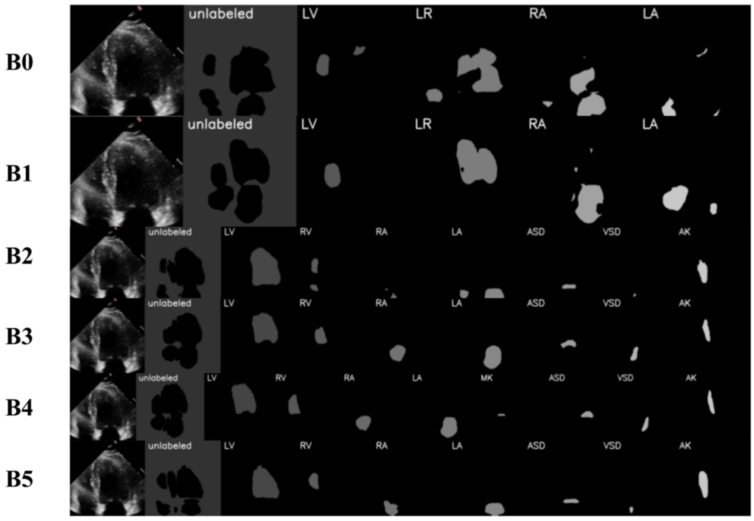


Fig. 6. Sample segmented images for Model B0—B5

The Fig. 6 illustrates the quality of segmentation across the SegFormer models B0 to B5. Each row corresponds to a different model, with improvements in segmentation quality generally observed as to move from B0 to B5. The segmentation maps highlight how effectively each model identifies and separates different regions within the echocardiographic images.

Model B0 shows the lowest quality segmentation, with many regions poorly defined and a significant amount of overlap or misclassification of anatomical areas. As it is possible to progress to Model B1, there is a noticeable improvement in segmentation clarity, with structures like the left ventricle (LV) and right atrium (RA) becoming more distinctly outlined.

Models B2 and B3 further enhance this segmentation quality, with better delineation of cardiac structures and less ambiguity in the regions identified. However, Model B3, despite its improved segmentation, occasionally misclassified smaller structures, indicating that while it can capture the general shape of the regions well, it may lack precision in differentiating finer details.

Models B4 and B5 demonstrate the most refined segmentation performance, with clearly separated regions and accurate boundaries for even the smaller anatomical features. These models consistently highlight structures such as the left atrium (LA) and ventricular septal defect (VSD) with greater precision, indicating a higher capability in distinguishing between different areas within the echocardiographic images.

5. 2. Implementation and integration of the YOLOv7 detection model

The main goal of this experiment is to train YoloV7 models for the task of binary classification of cardiac pathology (ASD – negative, AS – positive) in echocardiography images.

5 2. 1. Model setup for ASD detection

The YoloV7 experiments were conducted on an NVIDIA GeForce RTX 3060 with 6GB of RAM, and wandb.ai was used to store and visualize experiment results. The training configurations included the number of epochs of 500 and batch size of 32. The dataset was split into training (66 %), testing (10 %) and validation (34 %), ensuring representative samples from each class.

5. 2. 2. Classification results

The output is represented as rectangles of regina detection superimposed on the original medical image in grayscale.

As shown in Fig. 7, each rectangle in the map corresponds to one of the two classes (ASD, AS) and is their region of interest that was identified by the model.

The images were generated using Weights&Biases (wandb), which doesn't currently allow changing text parameters like font size in diagrams. This limitation is due to the tool's focus on experiment tracking and data visualization, and it doesn't offer advanced options for customizing graphic elements.

The left part of Fig. 8 shows the precision-confidence relationship for two classes, along with overall performance across all classes. The class AS demonstrates higher and more stable precision across different confidence levels, while ASD struggles with lower precision, especially at higher confidence thresholds, indicating more challenges in accurately predicting this class. The overall model reaches perfect precision at a confidence level of 0.934, but significant variability, especially for ASD, suggests potential issues with class imbalance or feature variability affecting performance.

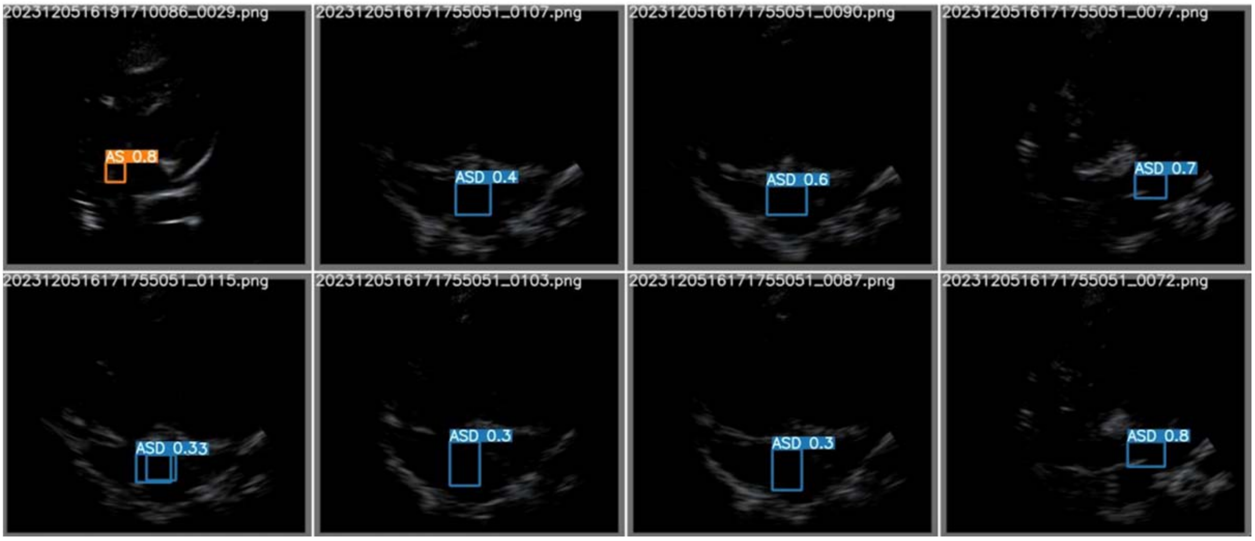
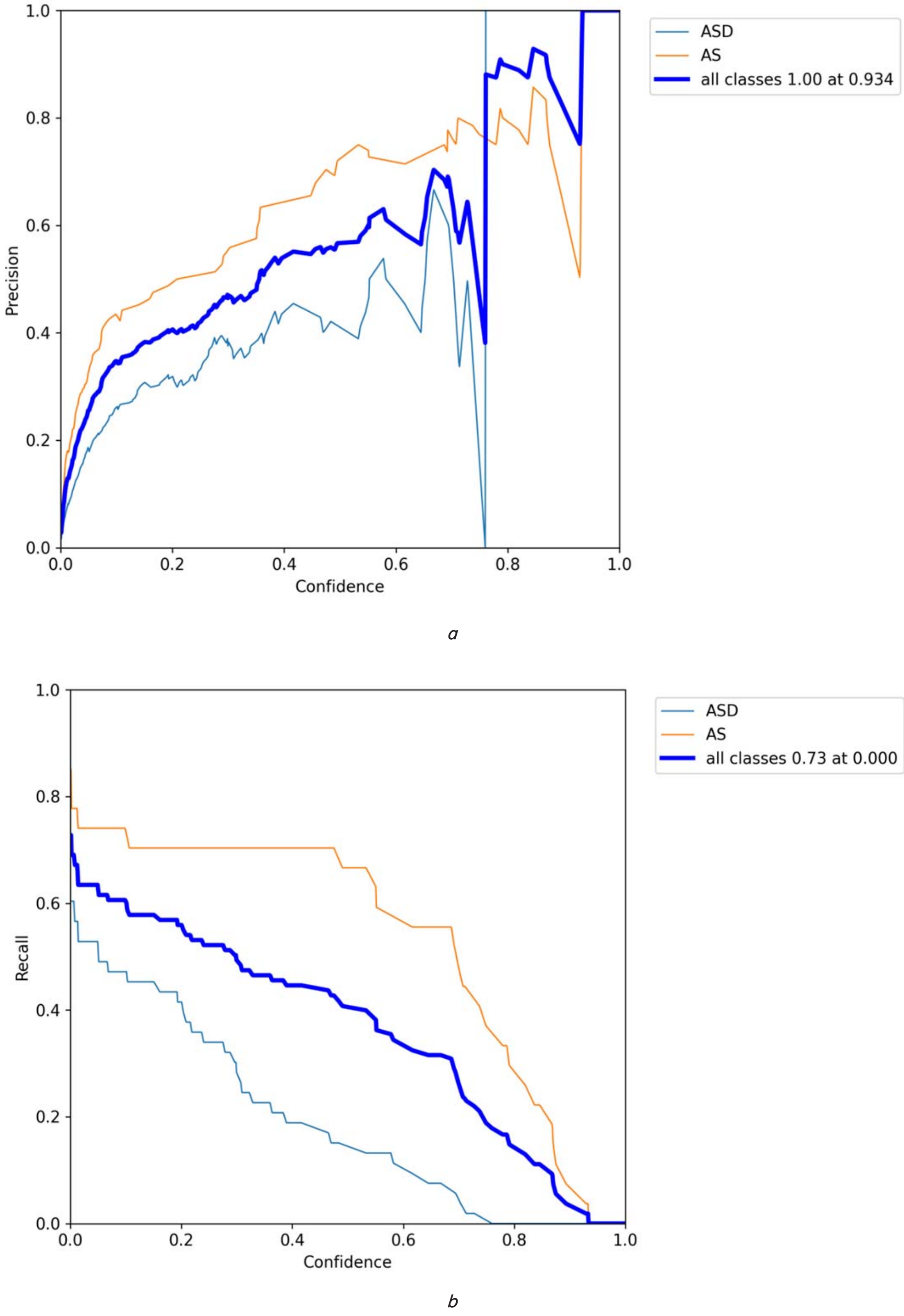

20231205161917100086_0029.png
AS 0.8
20231205161717555051_0107.png
ASD 0.4
20231205161717555051_0090.png
ASD 0.6
20231205161717555051_0077.png
ASD 0.7
20231205161717555051_0115.png
ASD 0.33
20231205161717555051_0103.png
ASD 0.3
20231205161717555051_0087.png
ASD 0.3
20231205161717555051_0072.png
ASD 0.8

Fig. 7. Atrial septal defects and aortic stenosis classification samples

The right part of Fig. 8 depicts how recall changes with confidence for ASD, AS, and overall. ASD starts with moderate recall, which declines steadily as confidence increases, showing that higher confidence levels lead to fewer true positives being detected. AS maintains higher recall initially but also drops sharply at higher confidence, indicating a similar trend. The combined curve shows that while recall decreases with increasing confidence, the optimal trade-off zones occur at lower to mid-confidence levels.

The Fig. 9 shows the F1 score against varying confidence levels, which balances precision and recall for each class.



a

b

Fig. 8. Precision and recall analysis for ASD and AS detection: *a* – precision vs. confidence curve; *b* – recall vs. confidence curve

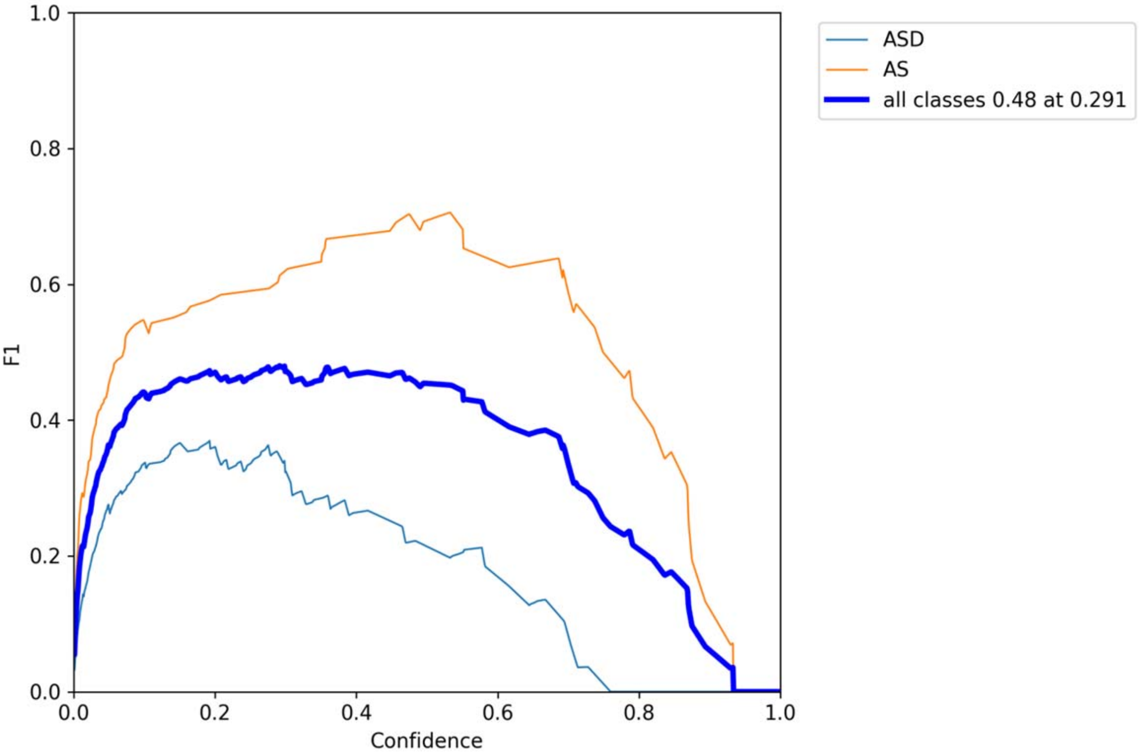


Fig. 9. F1 vs. confidence curve

The ASD curve fluctuates with lower F1 scores, suggesting instability in the model's predictions, particularly for ASD. The AS curve performs significantly better, with smoother and higher F1 peaks, indicating a stable balance between precision and recall. The combined performance of all classes (bold blue) highlights a more balanced F1 score, peaking around moderate confidence levels, which are the optimal points for the model's operation.

6. Discussion: efficacy and implications of the SegFormer-YOLOv7 pipeline for ASD detection

In this paper, let's propose a novel approach for assisting in the detection of ASD. To achieve efficient detection, let's employ a combination of methods: a recently introduced transformer-based segmentation network (SegFormer) is used to identify cardiac structures, which are then cross-referenced with a detection network (YOLOv7) that has been trained independently. This dual scheme allows to leverage the strengths of both models: the consistent, high-accuracy segmentation provided by SegFormer is complemented by the fast and precise detection capabilities of YOLOv7, effectively mitigating the limitations of each approach. In summary, let's make the following contributions:

– rigorously evaluate the SegFormer model, both pretrained and trained from scratch, emphasizing its segmentation accuracy and computational efficiency when applied to echocardiography images;

– reveal that SegFormer, when combined with YOLOv7, not only matches but, in key aspects, surpasses existing CNN-based models in medical image segmentation and object detection, thereby establishing its potential as a robust alternative for clinical deployment in pediatric cardiology.

Our results demonstrate that the SegFormer models with higher complexity, such as B4 and B5, achieved superior accuracy and mean Intersection over Union (IoU) metrics compared to simpler models (Table 1). These models were more effective at precisely segmenting echocardiographic images, as evidenced by the improvement in segmentation quality shown in Fig. 6. The segmentation maps illustrate the ability of models B4 and B5 to clearly delineate cardiac structures, such as the left atrium (LA) and ventricular septal defect (VSD), making them more suitable for ASD detection. However, this improvement comes at the cost of longer training times and higher computational demands, as seen in Fig. 5, where these models demonstrate higher accuracy but also greater variability in predictions.

In terms of the detection model, YOLOv7 showed strong performance in identifying small and subtle anomalies characteristic of ASD, particularly when integrated with SegFormer. This is highlighted in the results depicted in Fig. 7, where the precision-confidence relationship for ASD detection reveals challenges in maintaining high precision at higher confidence levels, as reflected in the precision-confidence curve (Fig. 8, *a*). This issue, along with a decline in recall for both ASD and AS classes at higher confidence thresholds (Fig. 8, *b*), indicates the need for balancing confidence thresholds to optimize the model's performance.

YOLOv7's performance in detecting small objects is further emphasized by the comparison of F1 scores across different confidence levels in Fig. 9, where the model struggles with lower precision and recall for ASD compared to other classes. This highlights the sensitivity of the model to parameter adjustments, and further refinement is necessary to improve the balance between precision and recall for ASD detection.

The integration of SegFormer and YOLOv7 models allowed for cross-referencing the detected ASD regions with the segmented heart chambers, improving diagnostic accuracy. This is evident in the enhanced identification of ASD regions with fewer false positives, as demonstrated in the performance metrics presented in Table 1 and the visual examples in Fig. 6. The cross-referencing mechanism ensured anatomical relevance, which is crucial for reducing diagnostic errors.

Overall, all results show that the SegFormer-YOLOv7 pipeline offers a balanced approach for ASD detection, leveraging the segmentation accuracy of SegFormer and the detection capabilities of YOLOv7. However, the variability in the model's predictions, particularly for ASD detection, indicates that further improvements are needed in handling diverse image quality and patient variability.

7. Conclusions

1. An efficient segmentation model utilizing the SegFormer architecture, optimized for identifying key cardiac structures in echocardiography images was successfully developed. Specifically, the SegFormer-B1 and B2 models achieved high segmentation accuracy (77,93 %+/–7.78, 68.71 %+/–4.84 respectively). These models maintained lower computational demands compared to their more complex counterparts, making them suitable for practical deployment. The peculiarity of this result lies in the ability of SegFormer to capture both local and global features without extensive preprocessing, which differs from traditional CNN-based models that often require larger datasets and more computational resources. This effectiveness is explained by SegFormer's transformer-based architecture, which enhances generalization and efficiency even with limited data. The practical impact of this model's efficiency is reflected in a reduced inference time on hardware typical of resource-constrained environments.

2. YOLOv7 detection model was implemented and integrated to identify ASD regions within the segmented heart structures. YOLOv7 demonstrated strong performance in detecting small and subtle anomalies characteristic of ASDs. The model achieved high precision for aortic stenosis (AS) detection, reaching peak performance at a confidence level of 0.934. However, it faced challenges with ASD detection, showing lower precision and fluctuating F1 scores, indicating the need for further optimization in detecting smaller anomalies. This result differs from earlier object detection models that struggled with small defect detection, and it is explained by YOLOv7's advanced features like trainable bag-of-freebies and re-parameterization techniques, which enhance its capability to detect small objects efficiently. This capability plays a crucial role in early diagnosis and more accurate treatment planning, with the potential to significantly reduce the occurrence of false negatives in ASD detection. This improvement not only enhances diagnostic confidence but also contributes to timely and effective patient care, minimizing the risk of missed or delayed diagnoses.

Conflicts of Interest

The authors declare that they have no conflict of interest in relation to this research, whether financial, personal, authorship or otherwise, that could affect the research and its results presented in this paper.

Financing

The study was performed without financial support.

Data availability

Data cannot be made available for reasons disclosed in the data availability statement.

Use of artificial intelligence

The authors have used artificial intelligence technologies within acceptable limits to provide their own verified data, which is described in the research methodology section.

Acknowledgements

This research has been funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP13068032 – Development of methods and algorithms for machine learning for predicting pathologies of the cardiovascular system based on echocardiography and electrocardiography).

References

1. Williams, M. R., Perry, J. C. (2018). Arrhythmias and conduction disorders associated with atrial septal defects. Journal of Thoracic Disease, 10 (S24), S2940–S2944. https://doi.org/10.21037/jtd.2018.08.27
2. Liu, Y., Huang, Q., Han, X., Liang, T., Zhang, Z., Lu, X. et al. (2024). Atrial Septal Defect Detection in Children Based on Ultrasound Video Using Multiple Instances Learning. Journal of Imaging Informatics in Medicine, 37 (3), 965–975. https://doi.org/10.1007/s10278-024-00987-1
3. Mertens, L., Friedberg, M. K. (2009). The gold standard for noninvasive imaging in congenital heart disease: echocardiography. Current Opinion in Cardiology, 24 (2), 119–124. https://doi.org/10.1097/hco.0b013e328323d86f
4. Sadeghpour, A., Alizadehasl, A. (2018). Echocardiography in the Critical Care Unit. Case-Based Textbook of Echocardiography, 423–430. https://doi.org/10.1007/978-3-319-67691-3_32
5. Lancellotti, P., Price, S., Edvardsen, T., Cosyns, B., Neskovic, A. N., Dulgheru, R. et al. (2014). The use of echocardiography in acute cardiovascular care: Recommendations of the European Association of Cardiovascular Imaging and the Acute Cardiovascular Care Association. European Heart Journal - Cardiovascular Imaging, 16 (2), 119–146. https://doi.org/10.1093/ehjci/jeu210
6. Tiver, K. D., Horsfall, M., Swan, A., De Pasquale, C., Horsfall, E., Chew, D. P., De Pasquale, C. G. (2022). Accuracy of Highly Limited Echocardiographic Screening Images for Determining a Structurally Normal Heart: The Quick-Six Study. Heart, Lung and Circulation, 31 (4), 462–468. https://doi.org/10.1016/j.hlc.2021.08.021

7. Sicari, R., Gargani, L., Wiecek, A., Covic, A., Goldsmith, D., Suleymanlar, G. et al. (2012). The use of echocardiography in observational clinical trials: the EURECA-m registry. Nephrology Dialysis Transplantation, 28 (1), 19–23. https://doi.org/10.1093/ndt/gfs399
8. Grenon, V., Szymonifka, J., Adler-Milstein, J., Ross, J., Sarkar, U. (2023). Factors Associated With Diagnostic Error: An Analysis of Closed Medical Malpractice Claims. Journal of Patient Safety, 19 (3), 211–215. https://doi.org/10.1097/pts.0000000000001105
9. Li, Y., Liu, Z., Lai, Q., Li, S., Guo, Y., Wang, Y. et al. (2022). ESA-UNet for assisted diagnosis of cardiac magnetic resonance image based on the semantic segmentation of the heart. Frontiers in Cardiovascular Medicine, 9. https://doi.org/10.3389/fcvm.2022.1012450
10. Madani, A., Arnaout, R., Mofrad, M., Arnaout, R. (2018). Fast and accurate view classification of echocardiograms using deep learning. Npj Digital Medicine, 1 (1). https://doi.org/10.1038/s41746-017-0013-1
11. Shamir, O. (2018). Are resnets provably better than linear predictors. arXiv. https://doi.org/10.48550/arXiv.1804.06739
12. Li, Y.-Z., Wang, Y., Huang, Y.-H., Xiang, P., Liu, W.-X., Lai, Q.-Q. et al. (2023). RSU-Net: U-net based on residual and self-attention mechanism in the segmentation of cardiac magnetic resonance images. Computer Methods and Programs in Biomedicine, 231, 107437. https://doi.org/10.1016/j.cmpb.2023.107437
13. Ghorbani, A., Ouyang, D., Abid, A., He, B., Chen, J. H., Harrington, R. A. et al. (2020). Deep learning interpretation of echocardiograms. Npj Digital Medicine, 3 (1). https://doi.org/10.1038/s41746-019-0216-8
14. Lim, G. B. (2020). Estimating ejection fraction by video-based AI. Nature Reviews Cardiology, 17 (6), 320–320. https://doi.org/10.1038/s41569-020-0375-y
15. Guo, Z., Zhang, Y., Qiu, Z., Dong, S., He, S., Gao, H. et al. (2023). An improved contrastive learning network for semi-supervised multi-structure segmentation in echocardiography. Frontiers in Cardiovascular Medicine, 10. https://doi.org/10.3389/fcvm.2023.1266260
16. Vakanski, A., Xian, M. (2021). Evaluation of Complexity Measures for Deep Learning Generalization in Medical Image Analysis. 2021 IEEE 31st International Workshop on Machine Learning for Signal Processing (MLSP), abs 2012 4115, 1–6. https://doi.org/10.1109/mlsp52302.2021.9596501
17. Moradi, S., Oghli, M. G., Alizadehasl, A., Shiri, I., Oveisi, N., Oveisi, M. et al. (2019). MFP-Unet: A novel deep learning based approach for left ventricle segmentation in echocardiography. Physica Medica, 67, 58–69. https://doi.org/10.1016/j.ejmp.2019.10.001
18. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., Luo, P. (2021). Segformer: Simple and efficient design for semantic segmentation with transformers. arXiv. https://doi.org/10.48550/arXiv.2105.15203
19. Liu, X., Yang, X. (2008). Automatic acquisition of the four-chamber view for 3D echocardiography. IEICE Electronics Express, 5 (9), 316–320. https://doi.org/10.1587/elex.5.316
20. Ukibassov, B. M., Rakhmetulayeva, S. B., Zhanabekov, Zh. O., Bolshibayeva, A. K., Yasar, A.-U.-H. (2024). Implementation of Anatomy Constrained Contrastive Learning for Heart Chamber Segmentation. Procedia Computer Science, 238, 536–543. https://doi.org/10.1016/j.procs.2024.06.057
21. Zhou, Q., Sun, Z., Wang, L., Kang, B., Zhang, S., Wu, X. (2023). Mixture lightweight transformer for scene understanding. Computers and Electrical Engineering, 108, 108698. https://doi.org/10.1016/j.compeleceng.2023.108698
22. Silvestry, F. E., Cohen, M. S., Armsby, L. B., Burkule, N. J., Fleishman, C. E., Hijazi, Z. M., Lang, R. M. 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. Journal of the American Society of Echocardiography, 28 (8), 910–958. https://doi.org/10.1016/j.echo.2015.05.015
23. Rakhmetulayeva, S. B., Bolshibayeva, A. K., Mukasheva, A. K., Ukibassov, B. M., Zhanabekov, Zh. O., Diaz, D. (2023). Machine learning methods and algorithms for predicting congenital heart pathologies. 2023 IEEE 17th International Conference on Application of Information and Communication Technologies (AICT). https://doi.org/10.1109/aict59525.2023.10313184
24. Jwaid, W. M., Al-Husseini, Z. S. M., Sabry, A. H. (2021). Development of brain tumor segmentation of magnetic resonance imaging (MRI) using U-Net deep learning. Eastern-European Journal of Enterprise Technologies, 4 (9 (112)), 23–31. https://doi.org/10.15587/1729-4061.2021.238957