I.J. Image, Graphics and Signal Processing, **2024, 5, 105-120**
Published Online on October 8, 2024 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2024.05.08

Intelligent Infective Endocarditis Diagnostic System Based on Echocardiography

Victor Sineglazov*
Department of Aviation Computer Integrated Complexes, National Aviation University, Kyiv, Ukraine
E-mail: svm@nau.edu.ua
ORCID iD: https://orcid.org/0000-0002-3297-9060
*Corresponding Author

Kirill Ryazanovskiy
Department of Artificial Intelligence, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
E-mail: k.riazanovskyi@kpi.ua
ORCID iD: https://orcid.org/0000-0002-8771-8060

Andrew Sheruda
Department of Artificial Intelligence, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
E-mail: sheruda.andrew@lll.kpi.ua

Received: 08 May, 2024; Revised: 10 July, 2024; Accepted: 22 August, 2024; Published: 08 October, 2024

Abstract: In this paper, we developed a new approach to solve the problem of infective endocarditis (IE) diagnostics based on intelligent analysis of patients' echocardiography images. The approach is based on echocardiography segmentation results and detection of valvular anomalies (namely vegetations). In this article for the first time investigates CNNs and Visual Transformers (ViT) based segmentation methods within the framework of the vegetation segmentation task on echocardiography images. Additionally, ensemble methods for combining segmentation models using a new method of models competition for data points were proposed. Furthermore, we investigated methods for aggregating the results of the ensemble based on a new meta-model, pointwise weighted aggregation, which weighs the results of each model pixel by pixel. The last proposed step was to automatically calculate the volume of segmented vegetation to determine the degree of disease and the need for urgent surgical intervention. For the studied and proposed methods, the following ensemble segmentation accuracy was achieved on the test dataset: iou 0.7822, dice score 0.886. The proposed empirical algorithm for calculating the volume of vegetations provided the basis for further improvements of the studied approach. The results obtained indicate the great potential of the developed approaches in clinical practice.

Index Terms: CNN, ViT, semantic segmentation, echocardiography, infective endocarditis, vegetation, hybrid models

1. Introduction

1.1. Features of infectious endocarditis

Endocarditis is a disease of the cardiovascular system that manifests itself in inflammation of the endocardial surface of the heart. Depending on the cause, there are infectious and non-infectious forms. As the name implies, infectious endocarditis is caused by inflammation as a result of bacteria or fungi entering the bloodstream, which subsequently settle on the heart valves or other parts of the endocardium, while non-infectious endocarditis can be caused by autoimmune reactions, prolonged inflammation, blood clots, or other non-infectious processes.

Infective endocarditis (IE) usually occurs along the edges of the heart valves. The lesions, called vegetations, are masses of fibrin, platelets and infectious microorganisms that are held together by agglutinating antibodies produced by bacteria. As the inflammation continues, the ulcer can lead to erosion or perforation of the valve leaflets, resulting in valve failure, conduction pathway damage (in the septal region), or rupture of the sinus of Valsalva (in the aortic region) [1].

Infective endocarditis is a life-threatening cardiac infection that is prone to occur in some individuals with multiple heart valve disease, with an annual incidence of 4.6 per 100,000 people with a definite linear progression [2]. Increasing rates of antibiotic resistance, atypical microbes and even fungi now affect a significant proportion of all patients with IE.

Groups of people with an existing artificial valve or stent are particularly vulnerable to the disease. The combination of these unfavourable host, source, and pathogen factors leads to a disproportionately high mortality rate (30%), usually sudden and short-lived, which contrasts with the cure rate in other cardiovascular diseases [3].

1.2. Diagnosis of endocarditis

Modern methods for identifying IE include modified Duke criteria, which are based on echocardiography - the use of ultrasound examination of the heart. This is a type of medical imaging that uses standard ultrasound or Doppler [4]. The visual image created with this method is called an echocardiogram or cardiac echo.

When diagnosing endocarditis, three echocardiographic findings are distinguished: (1) vegetations, defined as mobile echo-dense masses implanted in the valvular or murine endocardium in the path of the regurgitant jet or implanted in prosthetic material without an alternative anatomical explanation; (2) abscesses; or (3) dissection of the new prosthetic valve. Abnormal echocardiographic findings not meeting these definitions were considered secondary criteria [5].

Echocardiography is commonly used to diagnose, treat, and monitor patients with any suspected or known heart dis- ease. It is one of the most commonly used imaging modalities in cardiology. It can provide a wealth of useful information, including the size and shape of the heart (a quantification of the size of the internal chamber), pump performance, the location and extent of any tissue damage, and a valve assessment. The echocardiogram image is a time-lapse recording of the work of the heart muscle in its various phases.

2. Problem statement

2.1. Problem description

Although the gold standard for diagnosing infective endocarditis is the Duke criterion, it does not determine the need for emergency surgery. And therefore, a visual assessment of the size of the vegetation can play a key role in organizing a treatment plan. Vegetation size $10mm^3$ - is a marker for urgent surgery. However, based on visual data it is difficult to estimate its actual size, that is, no one has made such attempts before. The possible decision to operate was made on the basis of indirect indicators, taking into account the experience of the surgeon, which creates significant risks. Artificial intelligence, on the one hand, reduces the workload of the doctor who processes echocardiography images, and on the other hand, allows you to make the right decision regarding the urgency of the operation and reduce patient mortality.

According to a research [6] from the journal Cardiovascular Ultrasound, the frequency of false negative checks of im- ages for valve abnormalities reaches 14.5%. The main reasons were the atypical position and small size of the vegetations. The authors of [7] note that although the classical clinical classification into acute or subacute endocarditis syndromes has not completely lost its usefulness, modern clinical standards have increased in line with the profound epidemiological changes observed in developed countries. The usual human error can play a role during any visual checks, so the neces- sary solution is to integrate computer vision models into the diagnostic process, which can help to eliminate the human factor and improve the quality of diagnosis.

Timely diagnosis of the disease can significantly reduce patient mortality. According to [8], the overall mortality rate was 24%, 42%, 50% and 56% in 1, 5, 10 and 20 years, respectively. It is not only about the urgent, immediate help that a patient delivered with an attack needs, but also about the diagnosis of people at risk in general. Intelligent signal and image processing is already used in various technical fields, such as for classifying and recognizing anomalies in radar data [9], diagnosing radio electronic systems [10] or detection of singularities of electric field distribution [11]. The development of AI tools for echocardiogram analysis and accurate disease diagnostics can radically change and simplify the entire process of disease detection.

2.2. Scope of the paper

The scope of this work is to calculate the volume of vegetation based on its images. Formally, it can be divided into several steps: (1) Training and selection of a segmentation model. In this step, we obtain a basic segmentation model from the baseline pool, open to improvement. (2) Model improvement through the use of ensembles. At this stage, ensembles of several models are built in order to improve their accuracy. (3) Improvement of the model selection algorithm for the ensemble. Since existing model selection algorithms are poorly adapted to the segmentation task, a version suitable for use is required. (4) Improvement of the ensemble aggregation method. Basic aggregation methods have little effect on the productivity of the model and require more powerful adaptation for this task. (5) Creation of an algorithm for calculating the volume of vegetation. Using the vegetation mask generated by the final model, it is necessary to calculate the volume of formation.

3. Literature Review

First of all, it is worth pointing out that the task of segmentation of vegetations in echocardiography images is completely new, that is, there are no similar works with the same problem at the time of publication.

The paper [12] from the Journal of Imaging investigated the role of artificial intelligence in echocardiography analysis. The authors confirm that advances in the application of AI in cardiology and echocardiography are rapidly expanding, leading to revolutionary changes in patient care. AI algorithms help to detect, classify, diagnose, and predict cardiac pathologies. They promise increased workflow efficiency, improved reproducibility, and higher diagnostic accuracy, and can be a cost-effective tool to meet the growing demand for cardiac imaging. There are many obstacles that need to be overcome to enable the use of AI in clinical practice, including a lack of data and clinical outcomes. Therefore, as the authors conclude, further research is needed to determine the accuracy and effectiveness of AI-based solutions, as well as how AI can affect clinical outcomes.

The developments of the last 3 years in the field of deep learning and echocardiography have been focused generally on segmentation and identification of the cardiac structure by deep pyramidal neural networks of local attention [13], deep EchoNet networks [14], detection of heart rhythm phases by deep convolutional networks [15], classification of video echocardiograms by time-distributed convolutional networks and dual-flow networks [16], creation of 3D heart models based on segmentation by convolutional networks [17]. The review works [18] also generally focused on the general tasks of segmentation and classification of the internal part of the heart by conventional CNNs.

In [19], the authors tested the hypothesis that neural networks and logistic regression analysis would provide better prediction of IE compared to the modified Duke's criterion. The results on an independent sample of 261 patients showed that logistic regression analysis and neural networks provide better prediction of IE compared to the clinically established modification of the Duke criterion. This result confirmed the feasibility and power of using artificial intelligence to diagnose IE and the need for further development and modification of such algorithms. However, the peculiarity and disadvantage of this work is the use of only clinical indicators of patients, while the echocardiograph image was evaluated manually by a cardiologist in the laboratory.

Due to the high mortality rate (~30%) of IE, the authors of [20] developed a robust machine learning-based risk assessment model to predict early mortality after IE surgery, which can help clinical decision-making and improve outcomes. A total of 476 consecutive patients with IE who were operated on at 2 centres were included. The training dataset consisted of 276 patients. Out of 89 potential predictors, eight variables were selected to train the XGBoost model and became the input to the prediction model. The authors also included vegetation features in their model, although they did not propose automatic calculation and segmentation. Nevertheless, on the internal test dataset, the AUC (Area Under the Curve) was 0.813 (95% CI, 0.670-0.933), and on the external test dataset, the AUC was 0.812 (95% CI, 0.606-0.956). The authors' AUC was significantly higher than that of other ensemble learning models, logistic regression models, and the European Cardiac Operative Risk Assessment System II.

On the basis of this work, an intelligent system was created; materials from [21] were used in its implementation. The conceptual approach for the detection system was inspired by papers [22, 23].

At present, we have not found any works that would aim to segment and analyse vegetation on echocardiography images for the diagnosis of IE, which confirms the relevance of the study. Therefore, the proposed work is unique and critically important for the automatic identification and calculation of vegetation volume for their further use in risk assessment and diagnosis.

4. Materials

The dataset used for the study consists of 668 cardiac images with a specialist-recorded abnormality (vegetation and abscess) and 632 "clean" images, a total of 1300 in parasternal and apical projections. Images were extracted from 17 echocardiograms in DICOM format. The dataset was provided by the Heart Institute of the Ministry of Health of Ukraine and registered on the VIVIDS70-100096 machine. The data was marked up by a PhD with more than 10 years of experience.

The original image format was 708 by 1016 pixels with three RGB channels, respectively, and was subsequently processed to a single-channel grey image of 512 by 512 pixels with a fixed target area.

The images were pre-processed using cropping, normalisation and contrast enhancement techniques. No augmentation was used, given that the images were taken from the same angles and under the same conditions.

An example of images from the dataset is shown in Fig. 1. Fig. 2 depicts the marked training images with segmented vegetations. The vegetations are typically seen attached to the heart valves and appear as irregular masses. The size ranges from a few millimeters to over a centimeter in length. Some vegetations appear more rounded or oval, while others have more elongated, amorphous, or lobulated shapes. The irregular shapes and varied appearances are typical due to the nature of the vegetations comprising clumps of infectious material and host tissue responses.

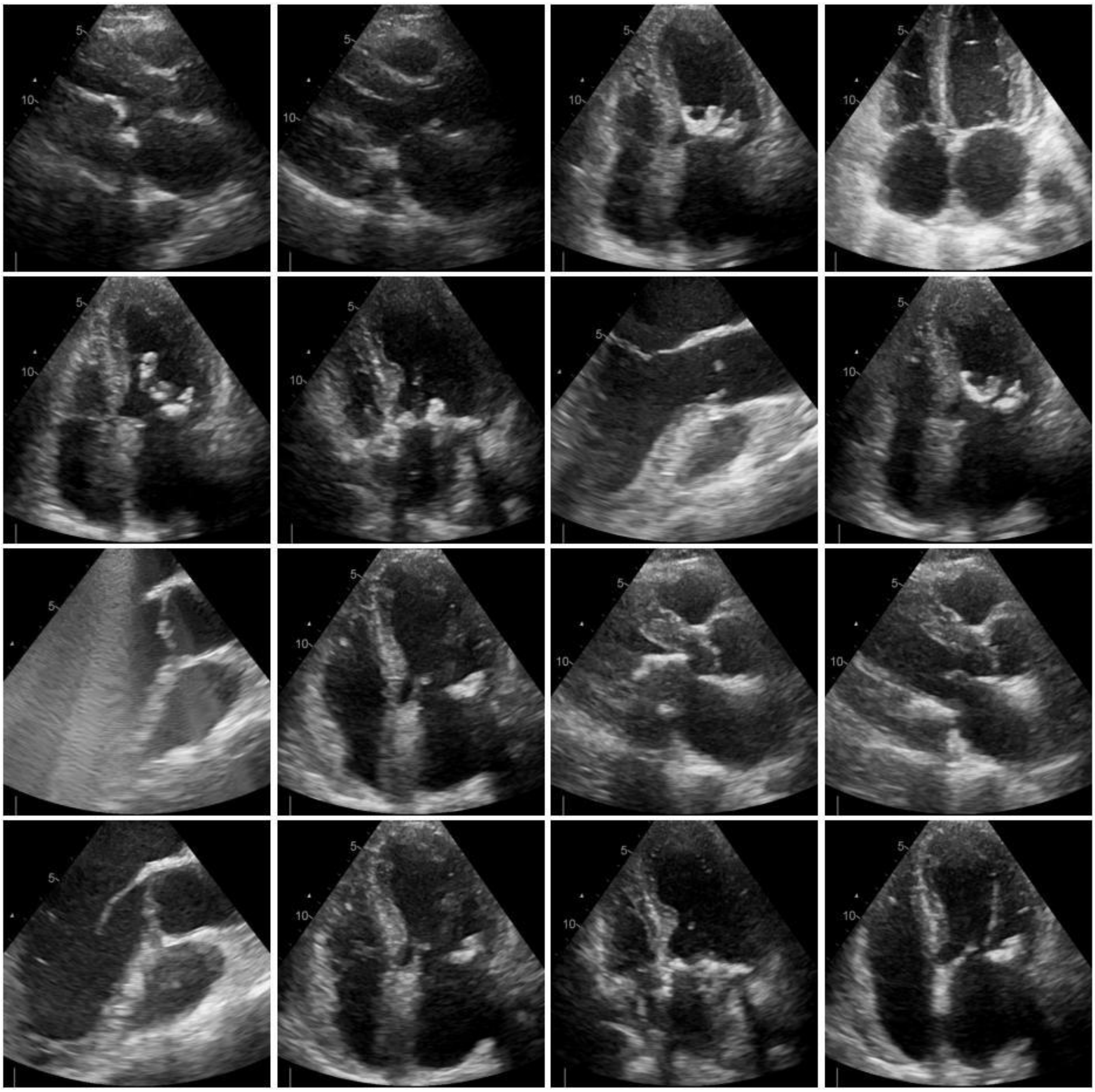

Fig. 1. Examples of echocardiograph images

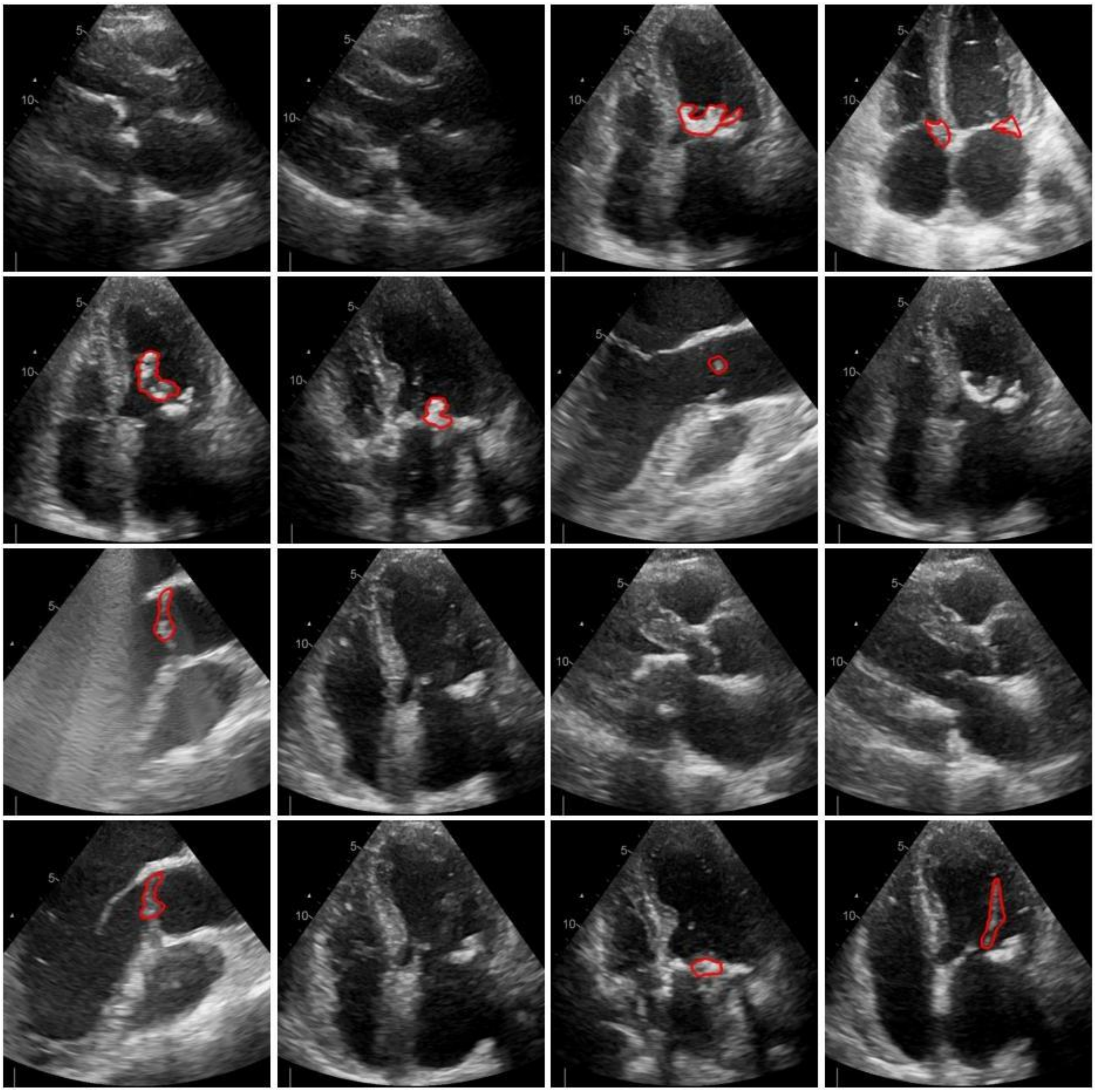

Fig. 2. Anomalies (vegetations) detected by specialists on the samples (marked in red)

5. Methods

5.1. Problem formulation

The problem of diagnosing IE is to find active foci (vegetations) in echocardiography images of patients' hearts. The main task is to semantically segment images into two classes: vegetation and background.

For a dataset $(X_1, Y_1), \ldots, (X_l, Y_l)$, where $X_i \in R^{n\times n}$ is the image matrix and $Y_i \in \{0,1\}^{n\times n}$ is a binary mask of the same size as the original image, where the matrix element 0 means the absence of vegetation, and 1 means the presence of vegetation on a given pixel of the image, to create a classifier $f(X)$ that correctly predicts the binary matrix $Y \in \{0,1\}^{n\times n}$ of the new image matrix $X \in R^{n\times n}$.

Referring to the result, it is possible to estimate how significant the set of abnormal pixels is and make a prediction about the health of the tissues in the image (Fig. 3).

The main problem for training the model on this dataset is the blurring on the echocardiogram when the heart muscle moves. It is also important to consider the complexity of segmentation on data in the low range of the color spectrum - target areas with vegetation merge strongly with the background and are easily confused with an artifact in the image.

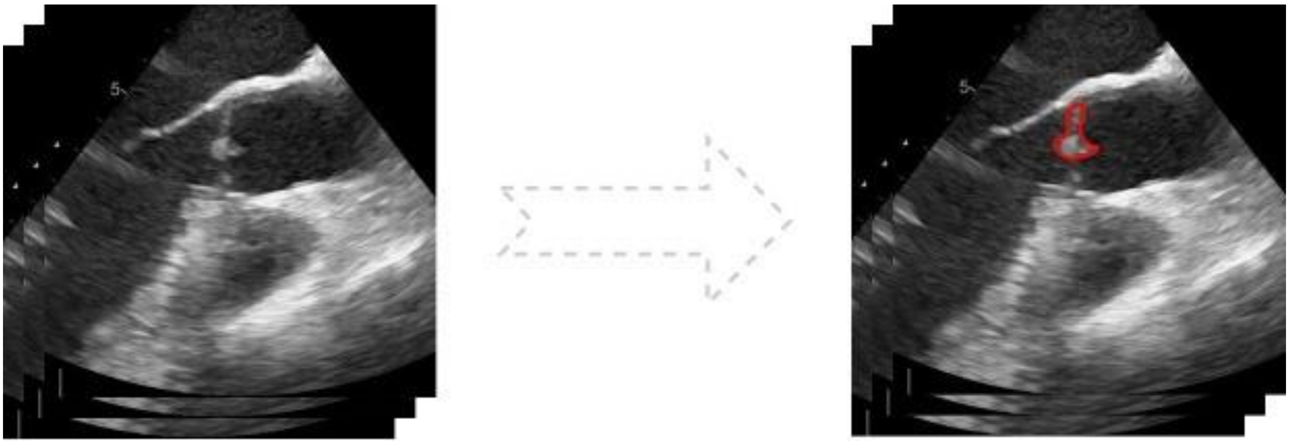

Fig. 3. Detection of endocardial tissue abnormalities

5.2. Modern approaches. CNN and ViT

The task in the field of computer vision is classified as semantic segmentation. Unlike instance segmentation, this type of task is distinguished by its extremely high accuracy and equally high demands on computing resources. Even the simplest models require gigabytes of GPU memory, and a certain infrastructure is needed to study such models, so currently, AI-based visual information analysis systems in medicine are quite rare.

Modern classical methods for solving the semantic segmentation problem are deep convolutional models of the encoder-decoder architecture (e.g., UNet [24], DeepLabV3+ [25]). However, studies over the past two years have shown that new architectures based on transformers, or rather Visual Transformers (ViT) (SegFormer is a special case for the semantic segmentation problem), can achieve comparatively better results than classical convolutional models [26, 27], including the semantic segmentation task [28]. This is mainly explained by the ability of transformers to study image fragments in more detail, due to the mechanism of self-attention.

Despite that ViTs face several challenges when applied to medical image segmentation. They often require large datasets to achieve good performance, which is problematic in the medical field where annotated data is scarce. Additionally, ViTs are computationally demanding, needing substantial GPU resources and high memory capacity, making them less accessible for institutions with limited computational infrastructure. Their reliance on pretraining on large-scale datasets like ImageNet, which may not capture the specific nuances of medical images, further complicates their application in this domain.

5.3. Proposed approach

5.3.1 Ensemble

Among the currently existing arrays of segmentation models, this paper proposes to use only CNN-based, Transformer- based and their hybrid methods, given the need for accuracy of predictions [29]. Transformer and convolutional architec- tural configurations can be too specialised for certain classes of images and perform poorly in feature extraction for others. The convolutional operation takes into account only informative features from a certain neighbourhood and aggregates them at a high level in the process of deep learning. Transformers on each Multihead-Attention block build a relevance map for the entire array of pixels in relation to each other [30].

An efficient way to combine the calculations and generalization abilities of several models is to use ensembles. In [31, 32], ensemble methods proved to be the best for the task of medical image segmentation. However, classical ensembles cannot always cover and fairly take into account the results of the models, so it was decided to use an ensemble based on a meta-model aggregation, aggregation of the mean and maximum [21, 33].

The general structure of the ensemble is shown in Fig. 4.

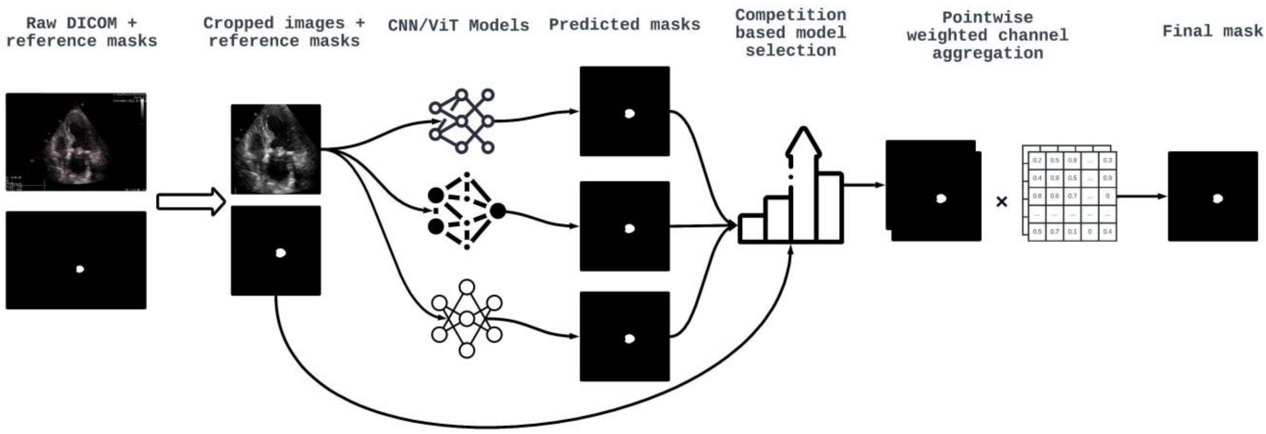

Fig. 4. Ensemble method results aggregation pipeline

5.3.2 Meta-model aggregation

To improve the capability of the meta-model in aggregating segments from multiple networks, we propose a trainable meta-model named pointwise weighted channel aggregation (PWCA). The proposed model consists of single-layer ma- chine learning model, which contains an $W \times H \times C$ tensor of learnable weights, where W and H are the width and height of the input masks, and C is the number of models in the ensemble (i.e. the number of channels). It receives channel- concatenated mask-probabilities from each individual model-participant of the ensemble as inputes (Eq. 1), multiplies them with the pixel-level weights and aggregates its own result (Eq. 2).

Let $S_i \in \{S_1, \ldots, S_N\}$ is a set of segmenters, $P_i \in \{P_1, \ldots, P_K\}$ is a set of pixels;

$$M_k^{(i)} = S_k(P_i),\ M_k^{(i)} \in M_k \tag{1}$$

where $M_k^{(i)}$ is the mask value of the *k-th* segmentor in pixel *i*;

$$PWCA_j\left(M_j^{(1)}, .., M_j^{(n)}\right) = \sum_{i=1}^{n} w_{ij} \cdot M_j^{(i)} \tag{2}$$

That is, in this configuration, we set our own weights for each pixel of the output of each network. The result is obtained by performing dynamic pixel-by-pixel weighting of the results of several different models. Fig. 5 shows the structure of such an aggregation.

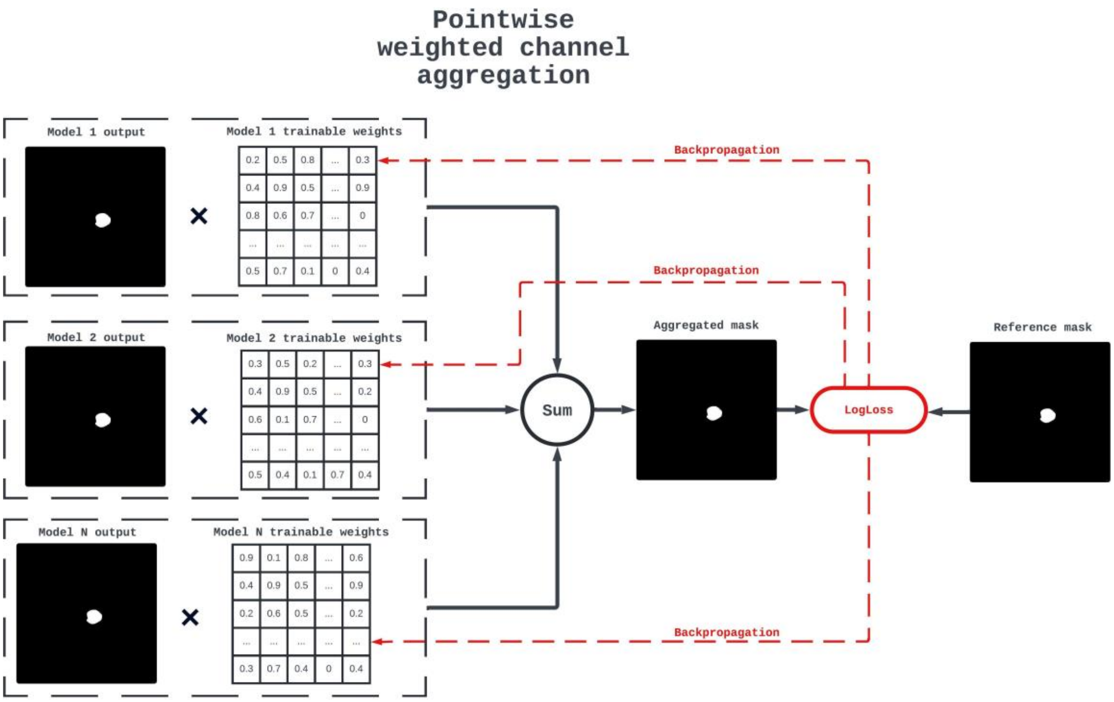

Fig. 5. Pointwise weighted channel aggregation structure

The method requires training of a meta-model based on the comparison of the calculated weighted mask from Eq. 2 with the reference masks. The training is performed using an error backpropagation algorithm. As a loss function Log Loss can be used.

5.4. Baseline models for comparison

Among the existing architectures of segmentation models, there are two main approaches - based on convolutional networks and based on the attention mechanism (i.e. based on visual transformers - VIT). In addition, there are also variants of hybrid networks. Both CNN and ViT-based models were used as models for comparing the results. A total of 6 base models were used:

1) Unet++ is an architecture based on U-Net. Due to the use of densely connected nested decoder subnets, it improves the processing of selected features and, as its authors have studied [34], outperforms U-Net in the tasks of medical segmentation of electron microscopy (EM) images, cells, nuclei, brain, liver, and lung tumors.
2) DeepLabv3+ is based on DeepLabv3 and outperforms it by having a simple but effective decoding module [25].
3) SegTransformer is based on a transformer, which is commonly used for sequence processing tasks such as machine translation or text modelling. In the case of segmentation, input images are treated as sequences of pixels, and the transformer's self-attention unit evaluates the relevance of pixels in relation to each other [35]. To improve computational efficiency, SegTransformer uses parallel image processing. Instead of processing each pixel sequentially, it divides the image into blocks and calculates the interactions within each block and between blocks in parallel. One of the key ideas is to use an attention mechanism to model the interaction between objects in the image. This helps the model to determine the context of objects and their relationships for more accurate segmentation. In fact, it is an adaptation of the visual transformer (VIT) adapted to the task of semantic segmentation. It is worth noting that for SegFormer, as for all architectures based on the attention mechanism, there is one common problem - the model is worse trained on small samples than classical approaches based on the CNN. This problem is especially significant for medical data [36], because there is often little of it.
4) UNETR, or UNet Transformer, is a transformer-based architecture for medical image segmentation that uses a pure transformer as an encoder to learn sequence representations of an input dataset [37]. The transformer encoder is connected directly to the decoder via a skip connection of different dimensions like U-Net to compute the final semantic segmentation output.
5) MANET is a multiple attention mechanism network based on the use of multiple attention mechanisms, which allows the network to process important areas of the image more efficiently, thereby improving segmentation accuracy [38].
6) PAN - combines an attention mechanism and a spatial pyramid to obtain accurate and concise features for pixel labelling instead of complex extended convolution and artificially designed decoder networks [39].

5.5. Metrics used

The following metrics were used to test the accuracy of binary segmentation models. The metrics are standard for the segmentation task.

1. Mean intersection over union (mIoU)
2. F1 (Dice) score
3. F2 score
4. Precision
5. Recall

The formula for IoU:

$$IoU = \frac{TP}{TP + FP + FN}$$

Precision and Recall were calculated as follows:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

The F1 score was used in the following form with $\beta = 1$:

$$F1 = \frac{(1 + \beta^2) \cdot TP + \epsilon}{(1 + \beta^2) \cdot TP + \beta^2 \cdot FN + FP + \epsilon}$$

where TP is the number of true-positive labels (pixels in the case of segmentation), FN is the number of false-negative labels, FP is the number of false-positive labels, ϵ is a very small constant to avoid division by zero, β is the importance coefficient of Recall over Precision. For the binary segmentation task, TP, FN, FP were calculated on a pixel-by-pixel basis.

Accordingly, the F2 score was calculated for a coefficient of $\beta = 2$.

5.6. Justification of the model choice for the ensemble. Contribution based on competition

According to the classical approach, the choice of models to form an ensemble is based on the combination of their accuracy and computational cost [40]. That is, it can be represented as a multi-criteria optimisation problem, where the value of each criterion is different [41]. The problem with this approach is that there is no assessment of the interaction of models with each other, so an additional criterion needs to be introduced - the diversity of models in the ensemble. Their inequality is explained by the fact that for models, diversity is more important than their overall accuracy, which can be manifested, for example, only on a part of the target dataset.

Segmentation models, in particular, require modification of this approach to improve the performance of models in relation to each other.

As part of this work, we developed a contribution (for segmenter) based on competition. It consists in jointly taking into account the accuracy and diversity of models and ranking them based on their own contribution. In this approach, machine learning models ”compete” for each data point and, accordingly, based on the measure of their quality, a resulting ranking can be generated for each data point (Eq. 5) and, subsequently, for the entire test dataset (Eq. 6).

Let I_i be a certain data point, in this case an image, $I_i \in \{I_1, \dots, I_K\}$ - the set of images of the test dataset, S_j - a certain segmenter, $S_j \in \{S_1, \dots, S_N\}$ - a set of segmenters; The task is to calculate the value of the result of the segmenter S_j at the data point I_i. The accuracy of the segmenter j at data point i we can see in Eq. 3, Y_i is the reference mask of data point i.

$$G_j^{(i)} = IOU\big(S_j(I_i), Y_i\big) \tag{3}$$

The following relations can be used to calculate the value of the result $G_j^{(i)}$ in comparison with other models. Enter the accuracy threshold (for example, 0.5), if the segmenter does not reach it, its score is reset to zero. Scale $G_j^{(i)}$ relative to the threshold (Eq. 4).

$$G_j^{(i)} = \big(G_j^{(i)} - 0.5\big) \ if \ G_j^{(i)} > 0.5 \ else \ 0 \tag{4}$$

Calculate the rating $R_j^{(i)}$ of the segmenter S_j at data point I_i relative to other segmenters:

$$R_j^{(i)} = \frac{G_j^{(i)}}{\sum_{j \in N} G_j^{(i)}} \tag{5}$$

Calculate the overall rating R_j of the segmenter S_j:

$$R_j = \sum_{i \in K} R_j^{(i)} \tag{6}$$

According to the results, we obtain the ranking of segmenters and proceed to the step of forming ensembles. This approach to evaluating models allows us to better take into account the results of models that performed well on data points where others performed poorly. In other words, the most accurate models with sufficient diversity are selected, i.e. those that cover the majority of the test dataset.

5.7. Approach to calculating the volume of vegetations

To calculate the volume, a necessary condition is the presence of echocardiography images in two perpendicular projections to create an approximate 3D vegetation map. As a result of segmentation of images from echocardiography in two perpendicular planes, we obtained a number of masks:

M_{it} is the mask for the ith projection at the moment of time t, $i \in \{1,2\}$; $M_{it} \in M_i$ where M_i is the set of masks in the i-th projection.

The task is to calculate the volume of the intersection of the masks on two planes located in projections perpendicular to each other. Note that the area of the mask corresponds to the number of pixels on it in square millimeters according to the scale of the image.

The proposed method for calculating the volume is as follows: each pair of binary masks in the perpendicular planes M_{1t} and M_{2t} is considered as projections of some three-dimensional figure. Having the forms of the projections, recreate the original figure using the method of constructive stereometry (Boolean intersection of two three-dimensional figures). The number of points in which corresponds to the desired estimate of the vegetation volume at time t.

Algorithm 1: Algorithm for calculating the optimal position of masks

Input: Two binary masks M_1 and M_2, both with shape (n, n)
Output: *best_rotation*, *best_shift*

```
Find the top, bottom, left and right boundaries for each mask;
Combine them as min(top1, top2), max(bottom1, bottom2), min(left1, left2) and max(right1, right2) to obtain the smallest rectangular region that covers both masks - (n1, n2);
Maximum intersection value max_overlap ← 0;
Shift value for the best result best_shift ← (0, 0);
Angle value for the best result best_angle ← 0;
Current shift value current_shift ← (0, 0);
Current angle value current_angle ← 0;
Shift directions directions ← [(−1, 0), (1, 0), (0, −1), (0, 1)];
Rotation angles angles ← [0, 90, 180, 270];
Number of iterations of the algorithm i ← 50;
Acceptable number of iterations without improvements staying_in_place ← 10;
while True do
    improved ← False;
    forall direction ∈ directions do
        new_shift ← (current_shift[0] + i × direction[0], current_shift[1] + i × direction[1]);
        forall angle ∈ angles do
            rotated_mask2 ← rotate(mask2, angle);
            shifted_mask2 ← shift(rotated_mask2, new_shift);
            overlap ← max((mask1 ∧ shifted_mask2).sum());
            if overlap > max_overlap then
                max_overlap ← overlap;
                best_shift ← new_shift;
                best_angle ← angle;
                improved ← True;
            end
        end
    end
    if improved then
        current_shift ← best_shift;
        current_angle ← best_angle;
    end
    else
        if i ≠ 1 then
            i ← i − 1;
        end
        else
            staying_in_place ← staying_in_place − 1;
            if staying_in_place == 0 then
                break;
            end
        end
    end
end
```

First, consider two important factors: (1) Frames from two different projections may not be synchronized under certain conditions. However, in our case, the echocardiograph recording was conducted starting from the same phase of the heart for two projections. (2) Two projections relative to each other may have some offset that needs to be corrected. For this, the following algorithm 1 was developed, which searches for the necessary transformations for M_{2t} relative to M_{1t}:

After obtaining the transformations, we apply them to the binary mask M_{2t} and calculate their intersection of both masks in the three-dimensional coordinate system [42]. A visualization of the result can be seen in Figure 6.

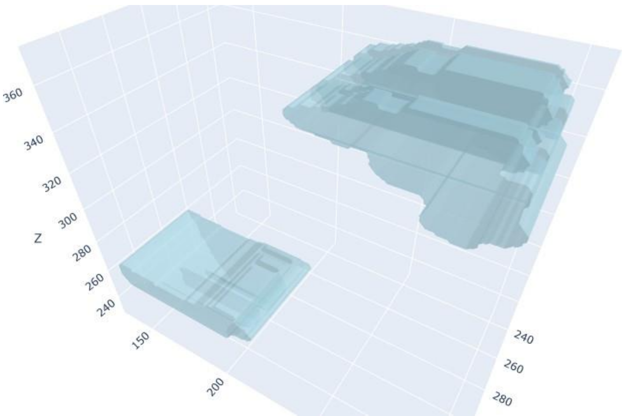

Fig. 6. 3D visualization of vegetation

Since one voxel of a figure corresponds to one square millimeter, the volume of the figure itself will correspond to the number of voxels of which it consists.

The general algorithm for calculating the volume will look like this:

a). We generate masks M_{it} for each of the two types of heart echos (i = 1 or 2) at a certain time t.
b). These masks show the areas of the image that interest us, and they are in two perpendicular planes.
c). The area of the mask is measured in square millimeters and depends on the number of pixels on it.
d). We apply the algorithm 1 to the mask M_{2t} to align it with the mask M_{1t}.
e). Each pair of masks M_{1t} and M_{2t} from two planes is considered a projection of some 3D figure in a three-dimensional coordinate system. We apply constructive solid geometry and obtain a set of voxexls that recreate this figure.
f). One voxel in a three-dimensional model corresponds to one square millimeter. The volume of the model is equal to the number of such cubes.

6. Results

6.1. Peculiarities of learning

The initial sample of 1300 images of 17 patients were evenly divided into three samples: training (800 images of 12 patients, 61.5%), validation (200 images of 3 patients, 15.5%), and test (300 images of 5 patients, 23%). All samples contain images from different patients to prevent data leakage and correctly estimate the generalization of the model.

The RTX 4090 graphics card was used to train the models. Training was performed in batches of 16 images with the Adam optimiser. Dice loss was used as a loss function. The learning rate was reduced when there was no change in the validation loss. The training was performed up to 200 epochs and stopped if the validation loss did not drop for more than 10 epochs.

6.2. Results of the proposed approach

Table 1 shows the results of the proposed architecture in compared to other approaches on the test sample.

Table 1. Results of the segmentation approaches on the test dataset

Approach	Encoder	IoU	F1 (Dice)	F2	Precision	Recall	Number of parameters
UNet++	ResNet34	0.5973	0,7428	0,8043	0,8512	0,6589	26.1 M
DeepLabV3+	ResNet50	**0.7091**	**0,8248**	**0,842**	**0,8539**	**0,7976**	26.7 M
SegTransformer	Mit_b2	0.5612	0,7193	0,762	0,7934	0,6578	27.3 M
UNETR	Mit_b2	0.6002	0,7481	0,7947	0,8291	0,6815	27.5 M
MANETR	Mit_b2	0.5872	0,7403	0,7845	0,8171	0,6767	35.0 M
PAN	ResNet50	0.6783	0,8113	0,8385	0,8576	0,7698	24.3 M

According to the results of calculating the contribution based on competition, a ranking of networks was formed. The summarised are shown in Table 2.

Table 2. Ranking of networks based on competition

Approach	Individual Contribution
SegTransformer	**64.0603**
DeepLabV3+	**61.8607**
PAN	**47.7629**
UNETR	43.9673
MANET	36.8912
UNet++	32.4574

Figure 7 shows a visualisation of the results of the competition-based contribution for each data point.

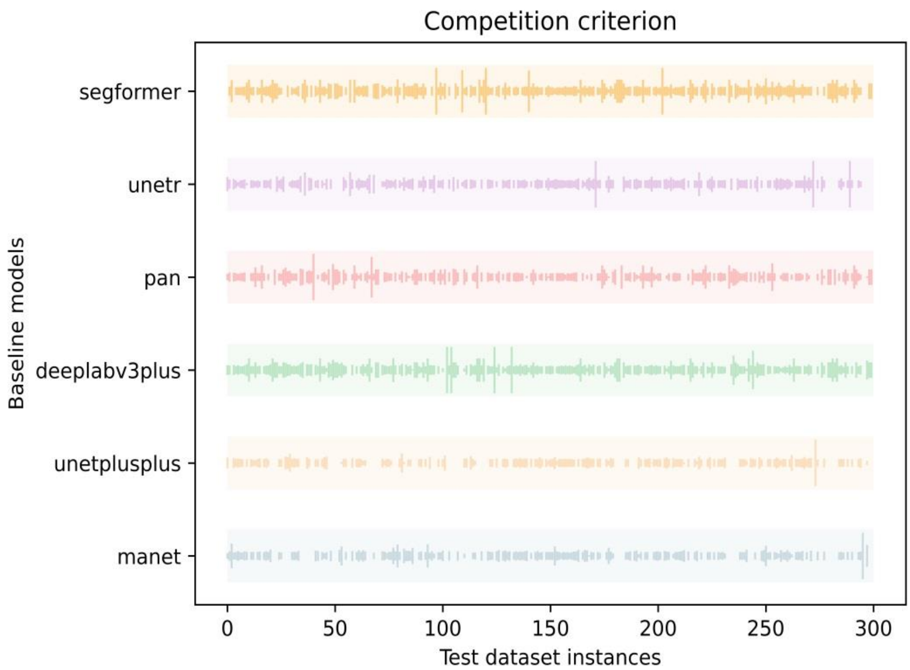

Fig. 7. Visualising the results of a competition-based contribution

Table 3 shows the results obtained by the ensembles of networks on the test sample using the DeepLabV3+, SegFormer and PAN models; these networks showed the best results in the individual contribution ranking.

Table 3. Results of model aggregation in an ensemble

Aggregation types	IoU	F1 (Dice)	F2	Precision	Recall
Average	0.6975	0.8218	0.7961	0.8688	0.7796
Maximum	0.6961	0.8207	0.8443	0.7844	0.8607
Meta-model PWCA	**0.7822**	**0.886**	**0.8891**	**0.8912**	**0.8810**

Having received the trained models and their results for images in two perpendicular planes, we can calculate the volume. At stage 1, we calculate vegetation masks of two echocardiography images in perpendicular planes. At stage 2, we construct projections of masks in 3D space and use algorithm 1 to find their best position. The projections of the masks in 3D space are shown in blue and yellow, and their intersection is shown in red. At stage 3 we extract the intersection of the 3D projections of the masks and calculate its volume based on knowledge of the size of single voxel in cubic millimeters. An example is shown in Figure 8.

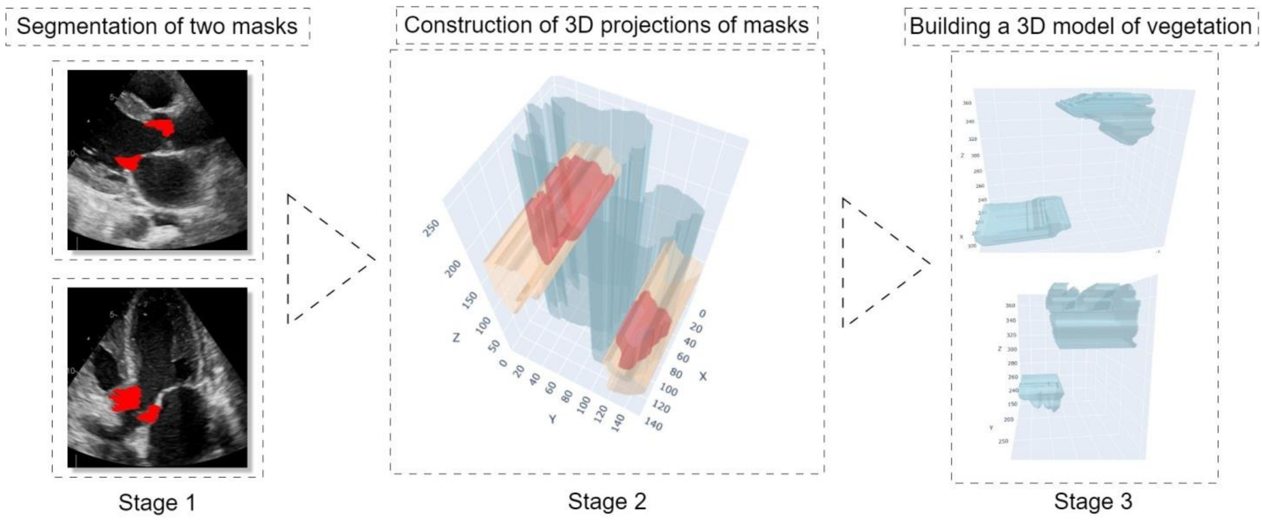

Fig. 8. Example of volume calculation

7. Discussion

The results in Table 1 show a significant dominance of classical segmentation methods based on convolutional neural networks over hybrid networks that include a transformers architecture. However, according to the results in Table 2, it can be generalised that transformers networks can better cover the target sample and are able to ensure the stability of the results.

It can be concluded that transformers approaches need to be further refined and developed to be applied to the health- care sector. Most medical images require the network to pay close attention to the texture features of the input data, and based on research, transformers model configurations are poorly adapted to this.

An analysis of existing work has shown that transformers require a substantial sample size for effective training, so we can assume that the worst results are precisely related to this.

Ensemble aggregations in this work did not significantly affect the result and showed controversial efficiency. How- ever, the meta-model we proposed showed a large boost in results and its relevance.

8. Conclusion

In this paper, for the first time, an approach to solving the problem of determining the size of vegetations using the example of infective endocarditis based on the results of segmentation of cardiac echocardiography using neural networks was developed.

We investigated CNN and hybrid neural network models based on modern visual transformers for segmenting vegeta- tions on echocardiographic images for their further analysis. For the tested methods, the following segmentation accuracy was achieved: iou 0.7822, dice score 0.886.

The analysis of the results showed a significant advantage in accuracy of the DeepLabV3+ model compared to other CNN and VIT models. According to the results of this study, other model configurations, although slightly behind, also have significant potential and require careful research. The pointwise weighted channel aggregation method we proposed showed the best results in terms of ensemble.

As part of this work, we developed our own criterion for selecting models for the ensemble - a contribution based on competition.

For the first time, the volume of vegetation was calculated using the theory of constructive solid geometry. The obtained results of segmentation models indicate a great potential for research into the problem of segmentation of vege- tations in echocardiography images. The result of the work is a full-fledged approach to diagnostics, the implementation of which, using available clinical resources, allows for increasing the efficiency and accuracy of detection of IE and other diseases accompanied by the occurrence of vegetations at different stages of the disease, contributing to faster and more accurate treatment of patients.

The limitation of this approach is the low specialization of the finished system for different clinical cases, due to the difficult process of producing a training dataset. That is, such a system will pay attention to any anomalies of the endocardium, including endocarditis.

This work opens up new perspectives in the application of artificial intelligence and computer vision for the diagnosis of infective endocarditis, making this area open for further research and development.

References

[1] E.A. Ashley and J. Niebauer, *Cardiology Explained*. Explained Series. Remedica, 2004. ISBN: 9781901346220.

[2] Alberto Cresti et al., "Epidemiological and mortality trends in infective endocarditis, a 17-year population-based prospective study". In: *Cardiovascular Diagnosis and Therapy* 7.1 (Feb. 2017), pp. 27–35. DOI: 10.21037/ cdt.2016.08.09.

[3] Richard Rezar et al., "Infective endocarditis – A review of current therapy and future challenges". In: *Hellenic Journal of Cardiology* 62.3 (2021), pp. 190–200. ISSN: 1109-9666. DOI: 10.1016/j.hjc.2020.10.007.

[4] David Adams and Emily Forsberg, "Conducting a Cardiac Ultrasound Examination". In: *Echocardiography*. Lon- don: Springer London, 2009, pp. 31–46. ISBN: 978-1-84882-293-1. DOI: 10.1007/978-1-84882-293-1_2.

[5] A Evangelista and M T Gonzalez-Alujas, "Echocardiography in infective endocarditis". In: *Heart* 90.6 (2004), pp. 614–617. ISSN: 1355-6037. DOI: 10.1136/hrt.2003.029868.

[6] Zuning Ren et al., "Preoperative false-negative transthoracic echocardiographic results in native valve infective endocarditis patients: a retrospective study from 2001 to 2018". In: *Cardiovascular Ultrasound* 19 (Jan. 2021). DOI: 10.1186/s12947-020-00229-8.

[7] Guillermo Cuervo et al., "Current Challenges in the Management of Infective Endocarditis". In: *Frontiers in Medicine* 8 (2021). ISSN: 2296-858X. DOI: 10.3389/fmed.2021.641243.

[8] R Netzer et al., "Infective endocarditis: Determinants of long term outcome". In: *Heart (British Cardiac Society)* 88 (July 2002), pp. 61–6.

[9] Y.P. Ostrovsky, F.J. Yanovsky, and H. Rohling, "Turbulence and Precipitation Classification based on Doppler- Polarimetric Radar Data". In: *2006 International Radar Symposium*. 2006, pp. 1–4. DOI: 10.1109/IRS.2006. 4338026.

[10] Oleksandr Solomentsev et al., "Substantiation of Probability Characteristics for Efficiency Analysis in the Process of Radio Equipment Diagnostics". In: *IEEE 3rd Ukraine Conference on Electrical and Computer Engineering* (Aug. 2021), pp. 535–540. DOI: 10.1109/UKRCON53503.2021.9575603.

[11] Olga Shcherbyna et al., "Diagnostic Process Efficiency Analysis for Block Diagram of Electric Field Parameters Meter". In: *2021 IEEE 12th International Conference on Electronics and Information Technologies (ELIT)*. 2021, pp. 5–9. DOI: 10.1109/ELIT53502.2021.9501136.

[12] Timothy Barry et al., "The Role of Artificial Intelligence in Echocardiography". In: *Journal of Imaging* 9.2 (2023). ISSN: 2313-433X. DOI: 10.3390/jimaging9020050.

[13] Fei Liu et al., "Deep pyramid local attention neural network for cardiac structure segmentation in two-dimensional echocardiography". In: *Medical Image Analysis* 67 (2021), p. 101873. ISSN: 1361-8415. DOI: 10.1016/j.media.2020.101873.

[14] Amirata Ghorbani el al. "Deep learning interpretation of echocardiograms". In: *NPJ Digital Medicine* 3 (Dec. 2020), p. 10. DOI: 10.1038/s41746-019-0216-8.

[15] Moomal Farhad et al., "Cardiac phase detection in echocardiography using convolutional neural networks". In: *Scientific Reports* 13.1 (June 2023), p. 8908. DOI: 10.1038/s41598-023-36047-x.

[16] James Howard et al., "Improving ultrasound video classification: an evaluation of novel deep learning methods in echocardiography". In: *Journal of Medical Artificial Intelligence* 3 (Mar. 2020), pp. 4–4. DOI: 10.21037/jmai. 2019.10.03.

[17] T. Shirakawa et al., "Automated heart segmentation using a convolutional neural network accelerates 3D model creation for cardiac surgery". In: *European Heart Journal - Cardiovascular Imaging* 22 (Feb. 2021). DOI: 10. 1093/ehjci/jeaa356.353.

[18] Chayakrit Krittanawong et al., "Deep Learning for Echocardiography: Introduction for Clinicians and Future Vision: State-of-the-Art Review". In: *Life* 13 (Apr. 2023), p. 1029. DOI: 10.3390/life13041029.

[19] L Vogel et al., "Neuronal nets and logistic regression analysis provide improved prediction of infective endocarditis compared to the modified Duke Score: a post-hoc analysis of the prospective PRO-ENDOCARDITIS study". In: *European Heart Journal - Digital Health* 3 (Dec. 2022). DOI: 10.1093/ehjdh/ztac076.2776.

[20] Li Luo et al., "Machine Learning–Based Risk Model for Predicting Early Mortality After Surgery for Infective Endocarditis". In: *Journal of the American Heart Association* 11 (June 2022). DOI: 10.1161/JAHA.122. 025433.

[21] V Sineglazov et al., "Intelligent tuberculosis activity assessment system based on an ensemble of neural networks". In: *Computers in biology and medicine* 147 (2022), p. 105800. DOI: 10.1016/j.compbiomed.2022.105800.

[22] Felix Yanovsky et al., "Classification algorithms for weather radar". In: *2008 12th International Conference on Mathematical Methods in Electromagnetic Theory*. 2008, pp. 366–368. DOI: 10.1109/MMET.2008.4580996.

[23] Van Tho Dang and F.J. Yanovsky, "Fuzzy Detection and Classification of Dangerous Weather Phenomena Using Dual-Polarimetric Radar Measurements". In: *Proceedings of SPIE - The International Society for Optical Engi- neering* 7502 (June 2009). DOI: 10.1117/12.838058.

[24] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *LNCS*. Vol. 9351. Oct. 2015, pp. 234–241. ISBN: 978-3-319-24573-7. DOI: 10.1007/978- 3-319-24574-4_28.

[25] Liang-Chieh Chen et al., "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmenta- tion". In: *Computer Vision – ECCV 2018: 15th European Conference, Munich, Germany, September 8–14, 2018, Proceedings, Part VII*. Munich, Germany: Springer-Verlag, 2018, pp. 833–851. ISBN: 978-3-030-01233-5. DOI: 10.1007/978-3-030-01234-2_49.

[26] Alexey Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale". In: *International Conference on Learning Representations*. 2021. DOI: 10.48550/arXiv.2010.11929.

[27] Bichen Wu et al., "Visual Transformers: Token-based Image Representation and Processing for Computer Vision". In: *CoRR* abs/2006.03677 (2020). DOI: 10.48550/arXiv.2006.03677.

[28] Sixiao Zheng et al., "Rethinking Semantic Segmentation from a Sequence-to-Sequence Perspective with Transform- ers". In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2021, pp. 6877–6886. DOI: 10.1109/CVPR46437.2021.00681.

[29] Imran Ahmed et al., "Comparison of Deep-Learning-Based Segmentation Models: Using Top View Person Images". In: *IEEE Access* 8 (2020), pp. 136361–136373. DOI: 10.1109/ACCESS.2020.3011406.

[30] Ashish Vaswani et al., "Attention is All you Need". In: *Advances in Neural Information Processing Systems*. Ed. By I. Guyon et al. Vol. 30. Curran Associates, Inc., 2017. DOI: 10.48550/arXiv.1706.03762.

[31] Vajira Lasantha Thambawita et al., "DivergentNets:Medical Image Segmentation by Network Ensemble". In: *En- doCV@ISBI*. 2021. DOI: 10.48550/arXiv.2107.00283.

[32] Dominik Muller, Inaki Soto-Rey, and Frank Kramer, "An Analysis on Ensemble Learning Optimized Medical Image Classification with Deep Convolutional Neural Networks". In: *IEEE Access* 10 (2022), pp. 66467–66480. DOI: 10.48550/arXiv.2201.11440.

[33] Viktor Sineglazov and Anatoly Kot, "Design of hybrid neural networks of the ensemble structure". In: *Eastern- European Journal of Enterprise Technologies* 1 (Feb. 2021), pp. 31–45. DOI: 10.15587/1729-4061.2021. 225301.

[34] Zongwei Zhou et al., "UNet++: A Nested U-Net Architecture for Medical Image Segmentation". In: *Deep Learn- ing in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer International Publishing, 2018, pp. 3–11. ISBN: 978-3-030-00889-5. DOI: 10.1007/978-3-030-00889-5_1.

[35] Enze Xie et al. "SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers". In: *CoRR* (2021). DOI: 10.48550/arXiv.2105.15203.

[36] Dhruvin Gandhi, Vaishnavi Shah, and Pramila M. Chawan, "A Vision Transformer Approach for Classification an A Small-Sized Medical Image Dataset". In: *2022 5th International Conference on Advances in Science and Technology (ICAST)*. 2022, pp. 519–524. DOI: 10.1109/ICAST55766.2022.10039593.

[37] Ali Hatamizadeh et al., "UNETR: Transformers for 3D Medical Image Segmentation". In: *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 2022, pp. 1748–1758. DOI: 10.1109/WACV51458. 2022.00181.

[38] Rui Li et al., "Multiattention Network for Semantic Segmentation of Fine-Resolution Remote Sensing Images". In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), pp. 1–13. DOI: 10.1109/TGRS.2021. 3093977.

[39] Garas Gendy et al., "A Deep Pyramid Attention Network for Single Image Super-resolution". In: *2021 9th Interna- tional Japan-Africa Conference on Electronics, Communications, and Computations (JAC-ECC)*. 2021, pp. 14–19. DOI: 10.1109/JAC-ECC54461.2021.9691443.

[40] M. Zgurovsky, V. Sineglazov, and E. Chumachenko, "Classification and Analysis of Multicriteria Optimization Methods". In: *Artificial Intelligence Systems Based on Hybrid Neural Networks*. Vol. 904. Studies in Computational Intelligence. Springer, 2021. ISBN: 978-3-030-48452-1. DOI: 10.1007/978-3-030-48453-8_2.

[41] Viktor Sineglazov, Kirill Ryazanovskiy, and Olena Chumachenko, "Multicriteria conditional optimization based on genetic algorithms". In: *System research and information technologies* (Dec. 2020), pp. 89–104. DOI: 10.20535/ SRIT.2308-8893.2020.3.07.

[42] Gopal Sharma et al., *CSGNet: Neural Shape Parser for Constructive Solid Geometry*. 2018. DOI: https://doi.org/10.48550/arXiv.1712.08290

Authors' Profiles

Victor Sineglazov is a Doctor of Engineering Science and a Professor, serving as the Head of the Aviation Computer-Integrated Complexes Department at the Faculty of Air Navigation Electronics and Telecommunications, National Aviation University in Kyiv, Ukraine. He graduated from Kyiv Polytechnic Institute in 1973. His research areas include air navigation, air traffic control, identification of complex systems, wind and solar power plants, and artificial intelligence. He has authored more than 700 publications.

Kirill Ryazanovskiy, PhD student at the Department of Artificial Intelligence of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute". Research interests include computer vision, semi-supervised learning, medical image processing. Author of more than 5 publications.

Andrew Sheruda obtained a master's degree from the Department of Artificial Intelligence at the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" in Kyiv, Ukraine. Research interests include computer vision, semi-supervised learning, medical image processing.

How to cite this paper: Victor Sineglazov, Kirill Ryazanovskiy, Andrew Sheruda, "Intelligent Infective Endocarditis Diagnostic System Based on Echocardiography", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.16, No.5, pp. 105-120, 2024. DOI:10.5815/ijigsp.2024.05.08