

Fetal heart chamber segmentation on fetal echocardiography image using deep learning

Sutarno¹, Naufal Rachmatullah², Abdurahman³, Rahmad Fadli Isnanto⁴

^{1,3} Computer Systems, Computer Science, University Sriwijaya, Palembang, Indonesia

² Informatics Engineering, Computer Science, University Sriwijaya, Palembang, Indonesia

⁴ Computer Engineering, Computer Science, University Sriwijaya, Palembang, Indonesia

ARTICLE INFO

Article history:

Received Jun 28, 2025

Revised Jul 18, 2025

Accepted Jul 21, 2025

Keywords:

Confusion Matrix;

Deep Learning;

Fetal Echocardiography;

Image Segmentation;

U-Net.

ABSTRACT

Advances in medical imaging and utilization have encouraged the development of more sophisticated image analysis technologies. In this context, image segmentation acts as a fundamental preprocessing step, but fetal echocardiography (FE) image segmentation still faces challenges in terms of accuracy and efficiency. The dataset for developing the FE image segmentation model was obtained from the examination results of patients at Muhammad Husein Hospital (RSMH) in Palembang who had normal conditions, atrial septal defect (ASD), ventricular septal defect (VSD), and atrioventricular septal defect (AVSD), totaling 650 FE images, which have been verified by experts. Compared to previous studies, this study focuses on creating a DL-based segmentation model for FE images using an open-source framework and the Python MIScnn library, which is specifically designed for medical imaging. This differs from previous DL frameworks that are more general, such as TensorFlow or PyTorch, which do not emphasize specialization for medical imaging. Furthermore, in an effort to improve model accuracy and efficiency, various configurations were tested, including variations in batch size and loss functions. The Model performance evaluation was conducted comprehensively using various important metrics in addition to pixel accuracy and IoU, such as F1 score, average accuracy, precision, recall, and False Positive Rate (FPR). This method is expected to provide a more in-depth picture of model performance compared to previous studies that may have only considered a few metrics. The best results were achieved using the U-Net architecture with a batch size of 32 and the binary cross-entropy loss function. This U-Net model demonstrated excellent overall performance, achieving a pixel accuracy of 0.996, an IoU of 0.995, a mean accuracy of 0.965, an FPR of 0.004, a precision of 0.929, a recall of 0.933, and an F1-score of 0.941. These findings highlight the significant potential of deep learning methods in improving the accuracy and efficiency of fetal echocardiography image analysis.

This is an open access article under the CC BY-NC license.



Corresponding Author:

Sutarno,
Computer Systems, Computer Science,
Sriwijaya University,
Jl. Raya Palembang - Prabumulih Km. 32, Indralaya, Sumatera Selatan, 30862, Indonesia
Email: sutarno@unsri.ac.id

1. INTRODUCTION

Inborn heart disorders stand as the predominant birth defect morphological disorder in infants that occurs in about 0.8 to 1 percent of live births and accounts for 20 to 40 percent of deaths. Digital technological innovations in diagnosis and monitoring with imaging techniques, biomarkers, advanced devices, and technical improvements over the last few decades can reduce morbidity and mortality rates (Bouma & Mulder, 2017; Dolc et al., 2011; Hoffman & Kaplan, 2002). There are various

types of cardiac recording techniques available, such as ultrasound-echocardiography, CT scan, MRI, NMR, and X-ray with various display modes, but ultrasound-echocardiography technique is the most frequently used investigative tool due to its non-invasive nature, lower cost, and less risk to the patient compared to other techniques (Rawat et al., 2018).

Fetal echocardiography (FE) serves as a commonly employed medical assessment for the prompt identification of congenital heart disorders in fetuses. This technique is particularly effective in high-risk pregnancies, where abnormal findings on routine ultrasound prompt further evaluation (Sun, 2021). The four-chamber view (4ChV) of the heart is an important view among early FE images for evaluating ventricular proportions, septal defects, and heart rhythm (Miller et al., 2018). Accurate segmentation of fetal heart anatomy in cardiac imaging is a useful and important step for early diagnosis and timely treatment of congenital heart disease (Nurmaini et al., 2021). For early diagnosis and prompt treatment of congenital heart disease, the critical and crucial step is to perform accurate segmentation of the main anatomical features in the fetal heart chamber FE image (Liang et al., 2024). With accurate segmentation information, doctors can develop more informed treatment plans. This may include prenatal intervention if possible or preparation for immediate postnatal care, such as surgery or catheterization procedures.

Due to the rapid development of DL learning models, our research focuses on applying a variant of the U-Net segmentation model for FE image segmentation in the four-chamber view of the fetal heart using an open-source framework with the Python MIScnn library. The feasibility of our proposed method has a complex architecture such as training, prediction, and automatic evaluation in the form of cross-validation. The research carried out aims to implement and compare the performance of the proposed method with several conventional segmentation methods and simulated segmentation in previous studies.

This study aims to develop a highly accurate U-Net segmentation model for FE anatomy in a four-chamber view, identify the optimal configuration of its architecture, and demonstrate the effectiveness of DL implementation using open-source tools, as well as highlight the advantages of U-Net over conventional methods. The results will be highly beneficial for the early diagnosis of congenital heart defects and serve as a clinical decision-support tool that enhances technology accessibility. However, further research comparing U-Net with other modern DL architectures and conducting in-depth analyses of challenging cases and model limitations remains a critical gap to ensure better generalization.

Chamber segmentation is a challenging task due to several unfavorable factors: (a) engineering problems during recording and noise generated by ultrasound imaging, (b) complicated classification caused by similarity of anatomical structures and variation of scan angles, and (c) missing boundaries (Xu et al., 2020). Image segmentation is the process of grouping a digital image into segments pixel by pixel and has been applied to various medical images that aim to simplify and change the representation of an image into something more meaningful and easier to analyze (Hou, 2024; Kumar & Kumar, 2012).

Regarding the application of FE image segmentation methods that have been developed both conventionally, such as edge-based, threshold-based, region-based, clustering-based, watershed-based, and Partial Differential Equations (PDE), and semantic segmentation methods, such as Fully Convolution Network (FCN) (Long et al., 2015), SegNet (Badrinarayanan et al., 2017), PSPNet (Zhao et al., 2017), and DeepLab (L. C. Chen et al., 2018). This research does not attempt to find the problems and shortcomings of these methods.

Due to the rapid development of DL learning models, our research focuses on applying a variant of the U-Net segmentation model for FE image segmentation in the four-chamber view of the fetal heart using an open-source framework with the Python MIScnn library. The feasibility of our proposed method has a complex architecture such as training, prediction, and automatic evaluation in the form of cross-validation. The research carried out aims to implement and compare the performance of the proposed method with several conventional segmentation methods and simulated segmentation in previous studies.

Computer-aided diagnostics utilizing artificial intelligence (AI) offers a promising avenue for enhancing the efficiency and accessibility of the diagnostic process. Artificial intelligence is broadly defined as a field of computer science dedicated to developing data processing systems capable of reasoning, learning, and self-repairing (Annina Simon & Venkatesan, S, 2016). A key component of AI, Machine Learning (ML), specifically focuses on recognizing patterns from features within a dataset. ML incorporates computational learning theories from AI, studying and building algorithms that can

make predictions based on insights gleaned from dataset information (Toscano, 2024). Essentially, ML operates by constructing models from input training data to facilitate predictions or choices (Annina Simon & Venkatesan, S, 2016).

Deep Learning (DL), a sophisticated method within artificial intelligence, emulates the information processing of biological neural networks. DL models employ multiple layers; each composed of various linear and non-linear transformations. These DL methods prove highly effective for diverse medical diagnostic tasks, often surpassing human experts, particularly when compared to conventional machine learning algorithms in terms of achieving faster and more accurate results (Hesamian et al., 2019). Significant advancements in computer vision, driven by DL, have been developed and refined over time, notably through a specific algorithm known as the Convolutional Neural Network (ConvNet/CNN). CNN is a DL algorithm adept at processing input images, assigning importance (learnable weights and biases) to different aspects or objects within the image, and distinguishing between them. The preprocessing required for CNNs is considerably less than that for other classification algorithms, and with sufficient training, CNNs can learn these filters or features autonomously, unlike primitive methods where filters are manually engineered (Navab et al., 2015).

U-Net is a neural network architecture designed primarily for image segmentation (Navab et al., 2015). U-Net is an image segmentation technique designed for medical image analysis that can quickly and precisely segment images using limited training data. U-Net has a very high utility in the field of medical imaging and has resulted in the wide-reaching adoption of U-Net as the primary tool for medical imaging segmentation tasks. The success of U-Net is evident in its widespread use in all types of medical images, such as CT scans and MRIs, X-rays, and microscopes (Jasim et al., 2022; Siddique et al., 2020). In addition, U-Net architecture has the potential to be developed for other applications (Jasim et al., 2022; Siddique et al., 2020).

The fundamental design of the U-Net framework is divided into two segments. The initial segment acts as a contracting encoder or an analytical tool, resembling a conventional convolutional network while delivering classification insights; the subsequent segment is expanding as a decoder or synthesis, which consists of ascending convolution and merging of features from the first part (Patri et al., 2024). The development of the U-Net architecture enables the network to acquire specific classification insights. Furthermore, the expansion pathway enhances the output clarity, which can subsequently be relayed to the concluding convolutional layer to produce a completely segmented image. U-Net architecture is still being developed in many research fields, including Base U-Net, 3D U-Net, Attention U-Net, Inception U-Net, Residual U-Net, Recurrent Convolutional Network, Dense U-Net, U-Net++, Adversarial U-Net, Cascaded arrangement, and Parallel arrangement (Siddique et al., 2021).

U-Net is a neural network architecture designed primarily for image segmentation (Navab et al., 2015). U-Net is a remarkable image segmentation method crafted specifically for analyzing medical images, capable of swiftly and accurately segmenting visuals with minimal training data. Its utility in the realm of medical imaging is unparalleled, leading to the extensive adoption of U-Net as the go-to tool for segmentation tasks in this field. The effectiveness of U-Net is showcased through its prevalent application across various medical images, including CT and MRI scans, X-rays, and microscopy. In addition, the U-Net architecture has the potential to be developed for other applications (Siddique, Sidike, Elkin, & Devabhaktuni, 2020). The following are some U-Net architectures that are still being developed in various fields of research, including:

The U-Net architecture is cleverly split into two segments: the initial segment compresses information through a CNN design, which features two sequential 3x3 convolutions succeeded by a ReLU activation function and a max-pooling layer, this sequence is replicated multiple times. What sets the U-Net apart is its expansion phase, where the feature map undergoes a 2x2 ascending convolution for sampling. The feature map from the contracting layer is then trimmed and merged with the upsampled version. This is followed by another round of two 3x3 convolutions and a ReLU activation. Finally, a singular 1x1 convolution is employed to refine the feature map to the desired number of channels, resulting in a segmented output image. (Yin et al., 2022).

2. METHODOLOGY AND EVALUATION

The research began with the collection of FE images, which were then optimized through a pre-processing stage. Following this, the design of the FE segmentation model using a variant of U-Net and FractalNet will be carried out. Subsequently, training and testing of the segmentation model will be conducted. The primary regions of interest in FE (arteries and ventricles) are the focus of

segmentation. The confusion matrix, also known as the confusion metric, will be used to evaluate the model's performance comprehensively. Finally, analysis of the test results and drawing of conclusions will be performed.

The overall process flow diagram of the input, process and output of the monitoring system can be seen in Figure 1 below.

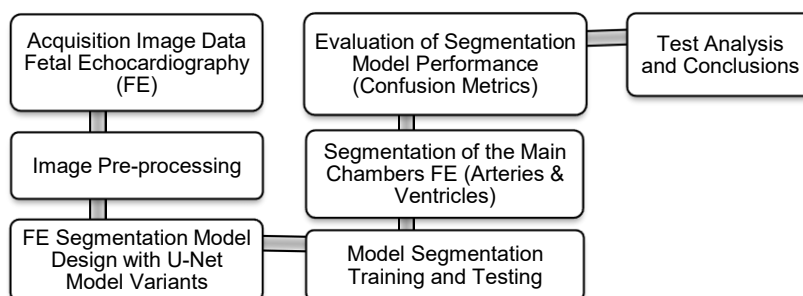


Figure 1. Flowchart of FE Image Processing Process

The following are some U-Net architectures that are still being developed in various fields of research, including:

a. BASE U-Net

The U-Net architecture is cleverly split into two segments: the initial segment compresses information through a CNN design, which features two sequential 3x3 convolutions succeeded by a ReLU activation function and a max-pooling layer, this sequence is replicated multiple times. What sets the U-Net apart is its expansion phase, where the feature map undergoes a 2x2 ascending convolution for sampling. The feature map from the contracting layer is then trimmed and merged with the upsampled version. This is followed by another round of two 3x3 convolutions and a ReLU activation. Finally, a singular 1x1 convolution is employed to refine the feature map to the desired number of channels, resulting in a segmented output image. (Yin et al., 2022). The foundational energy equation of the U-Net framework is derived using the subsequent formula:

$$E = \sum w(x) \log(p_{k(x)}(x)) \quad (1)$$

where p_k is the pixel-wise SoftMax function applied to the final feature map, and $a_k(x)$ denotes the activation in channel k .

$$p_k(x) = \frac{e^{a_k(x)}}{\sum_{k'=1}^K e^{a_{k'}(x)}} \quad (2)$$

b. 3D U-Net

The 3D U-Net architecture is an enhanced version of the foundational U-Net that facilitates segmentation in three-dimensional volumes (Çiçek et al., 2016). While the foundational design retains the contraction and expansion pathways, all two-dimensional processes are substituted with their three-dimensional equivalents, such as 3D convolution, 3D max concatenation, and 3D up convolution, culminating in a three-dimensional segmented output.

This network excels at segmenting images with minimal annotated examples. This efficiency arises because 3D images encompass numerous recurring patterns and shapes, enabling faster training even with limited labeling. The 3D U-Net has found extensive application in the segmentation of volumetric CT and MR images, including the assessment of cardiac structures, skeletal frameworks, spinal columns, brain tumors, liver tumors, pulmonary nodules, nasopharyngeal cancers, multi-organ segmentation, and head and neck structures, among others. Furthermore, the 3D U-Net has demonstrated impressive outcomes (S. Chen et al., 2020).

c. Attention U-Net

An important thing that is desired in image processing is the ability to focus on specific objects and ignore unnecessary areas. Attention U-Net achieves this by utilizing an attention gate. Attention gates are units that prune features that are irrelevant to the ongoing task. Each layer in the vast corridors feature a focus portal through which the relevant attributes of the narrowing corridor must filter before they merge with those collected from the wider expanse. The repeated application of the focus portal after every layer significantly enhances segmentation efficacy while avoiding unnecessary computational burden on the model. (Wu et al., 2024).

d. Recurrent Convolutional Network U-net

Recurrent Neural Networks (RNNs) were initially developed to process sequential data like text or audio. Unlike traditional feedforward networks, RNNs incorporate feedback loops where a node's output is influenced by its own previous output. This generates an internal state or "memory" that grants the node temporal features, allowing its output to change in discrete time steps. When this idea is extended over an entire layer, it allows the network to analyze contextual information acquired from previous data in the sequence. (Ming Liang & Xiaolin Hu, 2015).

These recurrent feedback loops are immediately integrated into convolutional layers of RCNNs. Specifically, feedback is performed after both the convolution and activation functions, feeding the filter-generated feature map back into its related layers. This recurring attribute enables individual units to enhance their feature maps using context from surrounding units, resulting in higher accuracy and performance. An RCNN's output y is defined by the interaction of convolution, activation, and continuous feedback.

$$y_{ijk}^l(t) = (w_k^f)^T x_l^{f(i,j)}(t) + (w_k^r)^T x_l^{r(i,j)}(t-1) + b_k, \quad (3)$$

where this expression $x^f(t)$ is the feedforward input and $x_l(t-1)$ is the recurrent input for layer l , w^f is the feedforward weight, w^r is the recurrent weight, and b is the bias of the k^{th} feature map.

Confusion Metric

The confusion matrix is an important tool in evaluating the performance of FE segmentation models, as it provides insight into the accuracy of the model and the types of errors. This process involves several stages that improve the model's ability to effectively distinguish between classes (Manai et al., 2024). The following confusion matrix is used to improve the performance of the FE segmentation model:

a. Intersection Over the Union (IoU)

IoU is one of the methods used to measure segmentation performance, evaluating the overlapping areas between machine prediction and ground truth prediction. The IoU value can be calculated using the following equation:

$$IoU = \frac{Prediction \cap Ground Truth}{Prediction \cup Ground Truth} \quad (4)$$

b. Pixel Accuracy

Pixel accuracy is the ratio of correct predictions at each pixel of the overall data. Pixel accuracy is the simplest method to evaluate how well an image segmentation model performs. The parameters needed are true positive, true negative, false positive, and false negative.

$$Pixel Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

c. Mean Accuracy (MA)

MA is the average ratio of the accuracy that has been obtained. The MA value can be calculated using the following equation:

$$Mean Accuracy = \frac{\sum Pixel Accuracy}{n_{epoch}} \quad (6)$$

d. False Positive Rate (FPR)

FPR is a value that is negative but predicted to be positive. The FPR value can be calculated using the following equation:

$$False Positive Rate = \frac{FP}{FP + TN} \quad (7)$$

e. Precision

Precision is characterized as the proportion of correctly predicted positive instances to the total anticipated positive instances. Alternatively, accuracy refers to the fraction of true positive forecasts to all outcomes that are expected to be positive. The value of precision can be determined through the following formula.:

$$\text{False Positive Rate} = \frac{TP}{TP + FP} \quad (8)$$

f. Recall

Recall is the ratio of true positive predictions compared to all true positive data. Recall value can be calculated using the following equation:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (9)$$

g. F1-score

F1-score is a weighted average comparison of precision and recall. The F1-score value can be calculated using the following equation:

$$\text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

description:

- TP (True Positive) refers to the number of correctly categorized positive picture pixels in the groundtruth image, each worth one.
- TN (True Negative) refers to the number of zero-valued negative pixels in the groundtruth picture that the system properly classifies.
- FP (False Positive) refers to the number of positive pixel data in the groundtruth image that the algorithm incorrectly classifies.
- FN (False Negative) refers to the number of zero-valued negative pixels in the groundtruth picture that the system incorrectly classifies.
- The prediction is a pixel from the segmentation result image.

Image segmentation is the process of partitioning a digital image into different classes or image objects. In this research, an evaluation is needed to get the results of the tests that have been carried out comparing the segmentation model using the U-Net architecture, which will be tested with the FractalNet architecture, from the evaluation results that have been obtained in previous studies. The evaluation parameters of the segmentation model obtained are pixel accuracy, intersection over union, mean accuracy, FPR, precision, recall, and F1-score.

3. RESULTS AND DISCUSSIONS

Data Acquisition

The training data required for the research is in the form of fetal echocardiography (FE) video data in format (*.mp4) which is the result of ultrasound (USG) recording of the fetal heart in the four-chamber view consisting of six video recordings obtained from the Central General Hospital DR. Mohammad Housein Palembang. This video data will then be pre-processed to get good data and in accordance with the input format of the training and testing process. Recapitulation of research data acquisition can be seen in Table 1 below.

Table 1. Fetal Heart Research Data

No	Name	Data Types	Duration (seconds)	Frame Rate (fps)	Number of Frames	Selection Frame
1	Video_1	Normal	04:51	25	120	80
2	Video_2	Normal	06:27	25	155	100
3	Video_3	Normal	12:06	30	362	200
4	Video_4	ASD	03:50	25	90	60
5	Video_5	VSD	05:46	25	135	100
6	Video_6	AVSD	06:05	30	180	110
Total data used					1042	650

Pre Image Processing

The pre-processing process is needed to confirm that the data format can be processed according to the needs of deep learning that has been designed, here in Figure 2 is a flowchart of the pre-processing stages carried out.

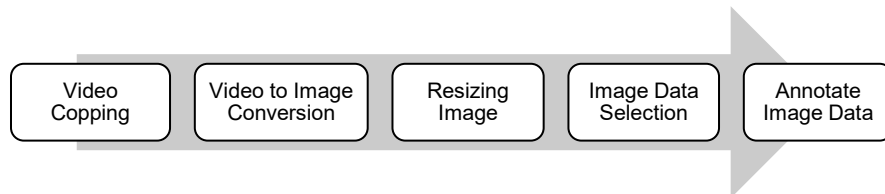


Figure 2. Flowchart of Research Data Pre-Processing Process

a. Video Cropping

This stage is done to get the fetal heart object in the video to be more focused, and other objects that are not needed can be reduced. Cutting is done by first determining 4 coordinate points, namely the upper left, upper right, lower left, and lower right boundaries of the fetal heart.

b. Convert Video to Image (Converting)

Video data is a collection of frames that are displayed sequentially. The next step is to convert the video data into images, because further processing will use digital image processing techniques. Converting video to image will also facilitate the process of labeling (ground truth) training data. The video cropping and conversion process is done using the Python library.

c. Resizing the Image

After the fetal heart image is obtained from the previous process, it turns out that the size or image resolution of each video is different; this will cause the dimensions of the training input data to vary and cannot be tolerated by CNN, so the image must be resized to be the same and in accordance with the architecture used. Here the image size is changed from 854x480 and 480x360, to a size of 256x256 pixels.

d. Data Selection

The converted image data does not entirely display the fetal heart object, because in the ultrasound recording process the doctor does not immediately find the correct position of the heart. For this reason, it is necessary to select data by selecting images with a complete view of the heart chamber and images with incomplete chamber views so that detection does not cause problems and confuse the system.

f. Data Annotation

The process of annotating image data uses the Labellmg tools application on the repository <https://github.com/tzutalin/labelImg>, determining the location of the heart room based on the results of consultation with experts (pediatric cardiologist) at Muhammad Hussein Hospital Palembang. The labeled image will then be stored as ground truth or training data.

Training and Testing

The image segmentation model is created through a neural network that receives the ground truth as a target. The ground truth is a correctly labeled image that provides information to the neural network regarding the expected output.

a. Division of test data and training data

At this stage, the image will be divided into training data and test data with a size of 70% for training data and 30% for test data. Data division will be done randomly so that 455 frames are obtained as training data and 195 frames as test data. Details of the division of training and testing data can be seen in Table 2.

Table 2. Division of Training Data and Testing Data

No	Name	Data Types	Number of Frames	Selection Frame	Ground Truth	Test Data
----	------	------------	------------------	-----------------	--------------	-----------

1	Video_1	Normal	120	80	56	24
2	Video_2	Normal	155	100	70	30
3	Video_3	Normal	362	200	140	60
4	Video_4	ASD	90	60	42	18
5	Video_5	VSD	135	100	70	30
6	Video_6	AVSD	180	110	77	33
Total data used				650	455	195

b. Training

The training process is carried out using the U-Net architecture using the following parameters: epoch, batch size, input layer, hidden layer, output layer, optimization, neurons, learning rate, and loss. Table 3 shows the contents of the training parameters.

Table 3. Division of Training Data and Testing Data

Model Parameters	Fill in the parameters
Hidden Layer Activation Function	Relu
Output Layer Activation Function	Sigmoid
Optimisation	Adam
Learning Rate	1e-5
Epoch	1000
Input Layer	256,256,1

The training process is carried out in 2 models using tuning parameters focusing on hyperparameters, viz., batch size and loss function. Batch size controls the number of training samples that must be performed before the internal parameters of the model are updated, and the loss function describes how well the model performs in the training process. The contents of the tuning parameters can be seen in Table 4.

Table 4. Tuning Parameter

Model	Batch Size	Loss Function
Model_U-Net	32	Binary Crossentropy
Model_FractalNet	32	Binary Crossentropy

c. Testing

In the testing phase related to segmentation results using U-Net and FractalNet architectures using the model in Table 4. The expected result is an accurate prediction of the segmentation of the four-chamber view of the child's heart. In the evaluation process, the image provides the results obtained with the two architectures. Image data that has gone through the ground truth stage is used as a standard in processing the evaluation. The performance of each model is assessed from the evaluation parameters of the segmentation model obtained in the form of pixel accuracy, intersection over union, mean accuracy, FPR, precision, recall, and F1-score. The outcomes of testing the U-Net model and the FraktalNet model are displayed in Table 5 and Table 6, and the accuracy and loss function graphs for each model are shown in Figure 5 and Figure 6.

(1) U-Net Model

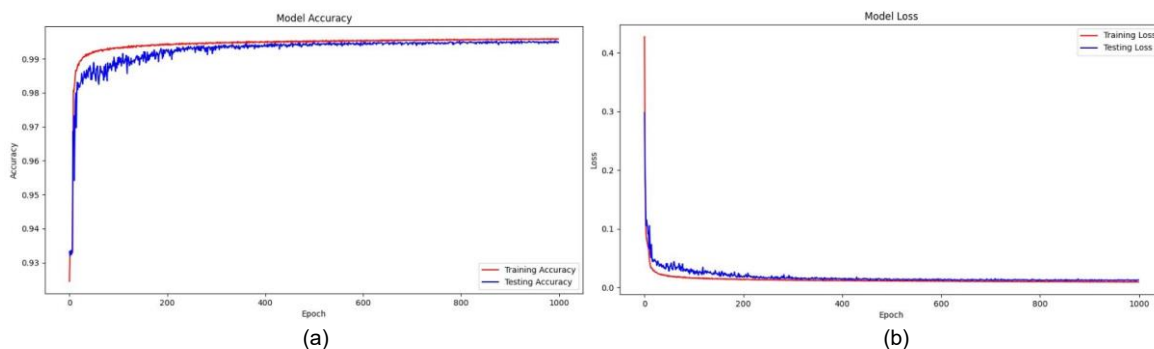


Figure 5. U-Net Model (a) Accuracy Graph (b) Loss Function Graph

Table 5. U-Net Model Evaluation Results

Assessment Parameters	Score (%)				
	Normal	ASD	VSD	AVSD	Average
IoU	99,47	99,38	99,43	99,63	99,48
Mean Accuracy	96,79	96,14	96,55	95,48	96,24
Pixel Accuracy	99,55	99,36	99,48	99,64	99,51
FPR	0,50	0,42	0,38	0,29	0,40
Precision	92,97	93,04	93,50	91,94	92,86
Recall	92,09	93,45	94,01	93,65	93,30
F1 Score	93,65	94,02	95,72	93,14	94,13

(2) FractalNet Model

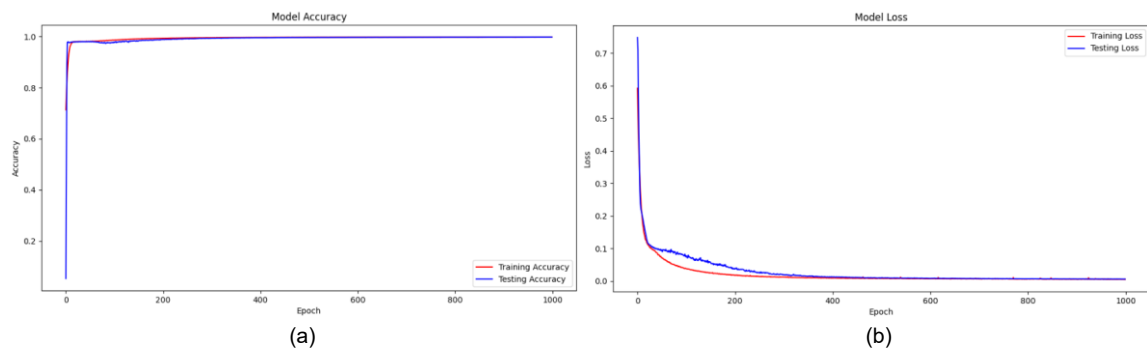


Figure 6. FractalNet Model (a) Accuracy Graph, (b) Loss Function Graph

Table 6. FractalNet Model Evaluation Results

Assessment Parameters	Score (%)				
	Normal	ASD	VSD	AVSD	Average
IoU	91,32	92,25	93,68	92,52	92,44
Mean Accuracy	95,73	95,18	96,80	96,12	95,96
Pixel Accuracy	99,20	99,10	99,22	99,14	99,17
FPR	0,45	0,32	0,30	0,22	0,32
Precision	90,23	91,30	91,10	89,98	90,65
Recall	91,91	92,34	93,30	92,88	92,61
F1-score	90,90	91,71	90,41	91,82	91,21

Analysis of Test Results

At this stage, the performance of the two models that have been built will be evaluated with the aim of measuring and comparing the performance of each architecture and model used. The parameters used to obtain the evaluation results include pixel accuracy, intersection over union (IoU), average accuracy, false positive rate (FPR), precision, recall, and F1-score. The comparison of the U-Net and FractalNet model results is shown in Table 7.

Table 7. Comparison of U-Net Model and FractalNet Model Results

Assessment Parameters	U-Net Model				FractalNet Model			
	Normal	ASD	VSD	AVSD	Normal	ASD	VSD	AVSD
IoU	99,47	99,38	99,43	99,63	91,32	92,25	93,68	92,52
Mean Accuracy	96,79	96,14	96,55	95,48	95,73	95,18	96,80	96,12
Pixel Accuracy	99,55	99,36	99,48	99,64	99,20	99,10	99,22	99,14
FPR	0,50	0,42	0,38	0,29	0,45	0,32	0,30	0,22
Precision	92,97	93,04	93,50	91,94	90,23	91,30	91,10	89,98
Recall	92,09	93,45	94,01	93,65	91,91	92,34	93,30	92,88
F1-score	93,65	94,02	95,72	93,14	90,90	91,71	90,41	91,82
Rata-rata	95,75	95,90	96,45	95,58	93,22	93,65	94,09	93,74

The model that produces the best evaluation value in the comparison between U-Net and FractalNet architecture is U-Net architecture with batch size 32, and the loss function used is binary crossentropy. The evaluation values on the U-Net architecture model are IoU 99.63%, pixel accuracy 99.64%, mean accuracy 96.79%, FPR 0.50%, precision 93.50%, recall 94.01%, and F1-score 95.72%.

4. CONCLUSION

Based on the results of the comparative analysis of the performance of paediatric heart segmentation using U-Net and FractalNet architectures, it can be concluded that U-Net shows the best performance in this task. With a batch size configuration of 32 and the use of binary crossentropy as the loss function, the U-Net model achieves very good average evaluation values: pixel accuracy 99.51%, IoU 99.48%, average accuracy 96.524%, FPR 0.40%, precision 92.86%, recall 93.30%, and F1-score 94.13%. These results clearly demonstrate the superiority of the U-Net architecture in paediatric heart segmentation compared to FractalNet, as measured by various evaluation parameters in the confusion matrix. These results are expected to aid in the early diagnosis of congenital heart defects and assist in clinical decision-making by improving the availability of technology. However, for better generalization, model limitations are still needed, and further research comparing U-Net with other contemporary DL architectures, as well as analysing difficult cases, is required.

ACKNOWLEDGEMENTS

The author would like to thank Sriwijaya University for its support through the 2024 Research Grant. Appreciation is also extended to the Digital Image Processing Laboratory for the facilities and resources provided.

REFERENCES

- Annina Simon, M. S. D., & Venkatesan, S. D. R. R. B. (2016). Machine Learning and its Applications: An Overview. *University of Glasgow, Department of Computing, January*.
- Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12), 2481–2495. <https://doi.org/10.1109/TPAMI.2016.2644615>
- Bouma, B. J., & Mulder, B. J. M. (2017). Changing Landscape of Congenital Heart Disease. *Circulation Research*, 120(6), 908–922. <https://doi.org/10.1161/CIRCRESAHA.116.309302>
- Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4), 834–848. <https://doi.org/10.1109/TPAMI.2017.2699184>
- Chen, S., Hu, G., & Sun, J. (2020). Medical Image Segmentation Based on 3D U-net. *2020 19th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)*, 130–133. <https://doi.org/10.1109/DCABES50732.2020.00042>
- Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). *3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation*. <http://arxiv.org/abs/1606.06650>
- Dolk, H., Loane, M., & Garne, E. (2011). Congenital heart defects in Europe: Prevalence and perinatal mortality, 2000 to 2005. *Circulation*, 123(8), 841–849. <https://doi.org/10.1161/CIRCULATIONAHA.110.958405>
- Hesamian, M. H., Jia, W., He, X., & Kennedy, P. (2019). Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. *Journal of Digital Imaging*, 32(4), 582–596. <https://doi.org/10.1007/s10278-019-00227-x>
- Hoffman, J. I. E., & Kaplan, S. (2002). The incidence of congenital heart disease. *Journal of the American College of Cardiology*, 39(12), 1890–1900. [https://doi.org/10.1016/S0735-1097\(02\)01886-7](https://doi.org/10.1016/S0735-1097(02)01886-7)
- Hou, Y. (2024). Applications of Image Segmentation Techniques in Medical Images. *EAI Endorsed Transactions on E-Learning*, 10. <https://doi.org/10.4108/eetel.4449>
- Jasim, S. H., Haleot, R. A., & Thajeel, S. A. (2022). Brain Stroke Segmentation Based on U-NET Algorithm. *2022 International Conference on Data Science and Intelligent Computing (ICDSIC)*, 208–211. <https://doi.org/10.1109/ICDSIC56987.2022.10076085>
- Kumar, P., & Kumar, S. (2012). *Analyzing the Medical Image by using Clustering Algorithms Through Segmentation Process* (Z. Zeng & Y. Li, Eds.; pp. 83490T-83490T – 5). <https://doi.org/10.1117/12.923776>
- Liang, B., Peng, F., Luo, D., Zeng, Q., Wen, H., Zheng, B., Zou, Z., An, L., Wen, H., Wen, X., Liao, Y., Yuan, Y., & Li, S. (2024). Automatic segmentation of 15 critical anatomical labels and measurements of cardiac axis and cardiothoracic ratio in fetal four chambers using nnU-NetV2. *BMC Medical Informatics and Decision Making*, 24(1). <https://doi.org/10.1186/s12911-024-02527-x>
- Long, J., Shelhamer, E., & Darrell, T. (2015). *Fully Convolutional Networks for Semantic Segmentation*.
- Manai, E., Mejri, M., & Fattahi, J. (2024). Confusion Matrix Explainability to Improve Model Performance: Application to Network Intrusion Detection. *2024 10th International Conference on Control, Decision and Information Technologies (CoDIT)*, 1–5. <https://doi.org/10.1109/CoDIT62066.2024.10708595>
- Miller, O. I., Simpson, J., & Zidere, V. (2018). Abnormalities of the Four Chamber View. In *Fetal Cardiology* (pp. 71–99). Springer International Publishing. https://doi.org/10.1007/978-3-319-77461-9_7

- Ming Liang, & Xiaolin Hu. (2015). Recurrent convolutional neural network for object recognition. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3367–3375. <https://doi.org/10.1109/CVPR.2015.7298958>
- Navab, N., Hornegger, J., Wells, W. M., & Frangi, A. F. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9351(Cvd), 12–20. <https://doi.org/10.1007/978-3-319-24574-4>
- Nurmaini, S., Rachmatullah, M. N., Sapitri, A. I., Darmawahyuni, A., Tutuko, B., Firdaus, F., Partan, R. U., & Bernolian, N. (2021). Deep learning-based computer-aided fetal echocardiography: Application to heart standard view segmentation for congenital heart defects detection. *Sensors*, 21(23). <https://doi.org/10.3390/s21238007>
- Patri, H. V., Priya, M. B., Mothukuri, M. B., Kumar, D. M., Ratna Prabha, K. V., & Gandham, S. R. K. (2024). U-Net Advancements in Semantic Segmentation for Autonomous Vehicles. *2024 10th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2292–2296. <https://doi.org/10.1109/ICACCS60874.2024.10717217>
- Rawat, V., Jain, A., & Shrimali, V. (2018). Automated techniques for the interpretation of fetal abnormalities: A review. *Applied Bionics and Biomechanics*, 2018. <https://doi.org/10.1155/2018/6452050>
- Siddique, N., Paheding, S., Elkin, C. P., & Devabhaktuni, V. (2021). U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications. *IEEE Access*, 9, 82031–82057. <https://doi.org/10.1109/ACCESS.2021.3086020>
- Siddique, N., Sidike, P., Elkin, C., & Devabhaktuni, V. (2020). *U-Net and its variants for medical image segmentation: theory and applications*. <http://arxiv.org/abs/2011.01118>
- Sun, H. Y. (2021). Prenatal diagnosis of congenital heart defects: echocardiography. *Translational Pediatrics*, 10(8), 2210–2224. <https://doi.org/10.21037/tp-20-164>
- Toscano, R. (2024). *Machine Learning* (pp. 203–238). https://doi.org/10.1007/978-3-031-52459-2_7
- Wu, Z., Liu, W., & Chang, J. (2024). Improving U-Net Performance for Tumor Segmentation Using Attention Mechanisms. *Journal of Computer Science Research*, 6(4), 66–72. <https://doi.org/10.30564/jcsr.v6i4.7271>
- Xu, L., Liu, M., Shen, Z., Wang, H., Liu, X., Wang, X., Wang, S., Li, T., Yu, S., Hou, M., Guo, J., Zhang, J., & He, Y. (2020). DW-Net: A cascaded convolutional neural network for apical four-chamber view segmentation in fetal echocardiography. *Computerized Medical Imaging and Graphics*, 80. <https://doi.org/10.1016/j.compmedimag.2019.101690>
- Yin, X.-X., Sun, L., Fu, Y., Lu, R., & Zhang, Y. (2022). U-Net-Based Medical Image Segmentation. *Journal of Healthcare Engineering*, 2022, 1–16. <https://doi.org/10.1155/2022/4189781>
- Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid Scene Parsing Network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6230–6239. <https://doi.org/10.1109/CVPR.2017.660>