

Multiclass instance segmentation optimization for fetal heart image object interpretation

Hadi Syaputra^{1,4}, Siti Nurmaini², Radiyati Umi Partan³, Muhammad Taufik Roseno^{1,4}

¹Doctoral Program in Engineering, Faculty of Engineering, Universitas Sriwijaya, Indralaya, Indonesia

²Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Palembang, Indonesia

³Department of Internal Medicine, Faculty of Medicine, Universitas Sriwijaya, Indralaya, Indonesia

⁴Computer Science Study Program, Faculty of Computer Science, Universitas Sumatera Selatan, Palembang, Indonesia

Article Info

Article history:

Received Aug 18, 2024

Revised Jun 17, 2025

Accepted Jul 10, 2025

Keywords:

Artificial intelligence

Fetal heart

Instance segmentation

Multiclass

ResNet

ABSTRACT

This research aims to develop a multi-class instance segmentation model for segmenting, detecting, and classifying objects in fetal heart ultrasound images derived from fetal heart ultrasound videos. Previous studies have performed object detection on fetal heart images, identifying nine anatomical classes. Further, these studies have conducted instance segmentation on fetal heart images for six anatomical classes. This research seeks to expand the scope by increasing the number of classes to ten, encompassing four main chambers left atrium (LA), right atrium (RA), left ventricle (LV), right ventricle (RV); four valves tricuspid valve (TV), pulmonary valve (PV), mitral valve (MV), and aortic valve (AV); one aorta (Ao), and the spine. By developing an instance segmentation method for segmenting ten anatomical structures of the fetal heart, this research aims to make a significant contribution to improving medical image analysis in healthcare. It also aims to pave the way for further research on fetal heart diseases using AI. The instance segmentation approach is expected to enhance the accuracy of segmenting fetal heart images and allow for more efficient identification and labeling of each anatomical structure in the fetal heart.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Siti Nurmaini

Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya

Palembang, Indonesia

Email: siti_nurmaini@unsri.ac.id

1. INTRODUCTION

The rapid development of AI technology has become an integral part of modern society. This is due to the capability of AI to rationalize and take actions or solutions that have the highest probability of achieving set goals [1]. In recent years, AI has been widely applied across various sectors, including government [2], infrastructure [3], agriculture [4], and healthcare [5]. By leveraging this technology, companies and organizations can integrate vast amounts of data to process information and make decisions. To support decision-making processes, an AI-based approach in developing models using machine learning (ML) algorithms is necessary. Various AI methodologies have been developed, one of which is ML. ML operates by utilizing neural networks to process data with the aim of generating knowledge that supports organizational or individual activities. In the process, ML extracts key features from data for model formation [6]–[8].

In the healthcare field, ML has been extensively used to aid medical professionals in decision-making. Research by Pullagura *et al.* [9] utilized ML to enhance the accuracy of fetal heart disease identification. Canadilla *et al.* [10] conducted research employing ML to improve the evaluation of fetal heart

function by optimizing image acquisition and measurements, thereby aiding in prenatal diagnosis of fetal heart remodeling and abnormalities. Hoodbhoy *et al.* [11] studied the accuracy of ML algorithm techniques in identifying high-risk fetuses through cardiotocography. Cömert and Kocamaz [12] used ML as a monitoring technique that provides crucial and vital information about fetal status during antepartum and intrapartum periods, as well as classifying fetal heart rate signals. However, previous studies have shown that ML methods have limitations in analyzing structured and limited data. Additionally, ML methods involve more complex stages, such as manual image augmentation, which can be time-consuming to produce actionable information for decision-making and actions [13].

To address the challenges of traditional ML methods, several studies have adopted a deep learning (DL) [14] approach for analyzing and predicting medical examination outcomes, especially in image classification and object detection to support fetal echocardiography examinations. By processing large amounts of data, DL has demonstrated potential in enhancing accuracy and efficiency in medical image analysis. DL methods are frequently employed in the medical field, such as in fetal cardiography image detection [15]. One of the primary advantages of DL techniques is their ability to extract significant insights, patterns, and information from images and videos. This is achieved through the development of algorithms and models that enable machines to analyze, process, and make decisions based on visual data [16]. Moreover, DL techniques can identify and depict individual objects in images while providing labels for each object, making them applicable in various fields such as object tracking [17] and medical imaging [18]. However, these studies mainly focus on the classification of medical images or videos by comparing one image object with another. Additionally, the classification technique in DL methods can only identify a single object within an image and categorize it based on that object. To overcome the limitations of DL classification techniques, a solution is required that can detect multiple objects within a single image or video [19]. In addition to classification and detection capabilities, DL methods also possess the ability to detect multiple objects in one image and video. For example, research conducted by Sapitri *et al.* [20] utilized DL for object detection in fetal ultrasound videos, identifying anatomical substructures of the fetal heart, including i) four main chambers: left atrium (LA), right atrium (RA), left ventricle (LV), right ventricle (RV); ii) four valves: tricuspid valve (TV), pulmonary valve (PV), mitral valve (MV), and aortic valve (AV); and iii) one aorta (Ao).

Subsequent developments in object detection [21], [22] have enabled the identification and categorization of every pixel in an image into meaningful object categories or areas, known as segmentation. Segmentation techniques include semantic segmentation and instance segmentation. Research by Rachmatullah *et al.* [23] used semantic segmentation methods to develop a semantic model that detects objects by assigning labels to each pixel in an image, ensuring that pixels with the same label have the same image. Simply put, semantic image segmentation is a technique used to identify specific object types within an image. However, semantic segmentation techniques have several drawbacks, including the inability to distinguish between individual objects in an image and difficulty identifying individual objects with similar textures [23], [24]. In contrast, instance segmentation can provide unique labels for each individual object [25], [26].

Efforts to recognize and separate each class of objects in an image rely heavily on instance segmentation, which in turn depends on the backbone architecture [27]. The backbone architecture plays a crucial role in instance segmentation by providing essential feature information of the areas to be segmented for the model [28]. Research conducted by Nurmaini *et al.* [29], has utilized the use of ResNet as the main structure to achieve optimal instance segmentation. The application of instance segmentation in the medical field includes automating the segmentation process and improving detection accuracy [30]. For instance, an instance segmentation approach for fetal echocardiography can simultaneously separate the four standard heart views and detect defects [29]. To accurately detect fetal heart abnormalities through fetal ultrasound, all heart substructures must be recognized in normal anatomy [20]. One of the most significant limitations associated with ultrasound involves interpersonal variability, meaning it depends on the examining doctor's skills and the patient's condition [28]. Referring to research by Sapitri *et al.* [20], which examined anatomical structure detection in fetal heart images, as well as research by Nurmaini *et al.* [28], which focused on instance segmentation for the four main chambers of the fetal heart and heart disease detection, this study expands its scope to include additional anatomical objects, namely the spine. The addition of the spine is crucial for medical practitioners in identifying the four-chamber view (A4C) of the fetal heart in images [31]. Therefore, the contributions of this study are the inclusion of ten anatomical objects of the fetal heart, namely LA, RA, LV, RV, TV, PV, MV, AV, Ao, and spine, and the development of a DL approach using instance segmentation methods for these ten anatomical structures. By developing a sample segmentation method for ten fetal heart anatomy objects and applying hyperparameter tuning to find the optimal settings [32], [33], this study aims to significantly improve medical image analysis in the healthcare field and pave the way for future research in detecting fetal heart disease. This approach promises accuracy in segmenting the fetal heart.

The segmentation of these ten anatomical structures was chosen based on clinical considerations as each has an important role in the diagnosis of congenital heart defects. The four main heart chambers (LA, RA, LV, RV) and the four valves (TV, PV, MV, AV) are the structures most frequently used in the functional assessment of the fetal heart via ultrasonography. The structure of the Ao is important in identifying blood outflow, while the spine helps to ensure correct anatomical orientation in the A4C. Accurate segmentation of these structures allows early identification of various abnormalities such as septal defects, valve stenosis, and abnormal positioning of the heart or other organs.

2. MATERIAL AND METHOD

Detecting the normal fetal heart anatomy from A4C video between 14 and 28 weeks of gestational age is a complex task. Segmentation aims to delineate cardiac structures using contour boundaries; however, this method is limited in capturing the spatial relationships among components. As illustrated in Figure 1, the workflow begins with the extraction and selection of video frames based on the A4C perspective. The selected frames are refined through cropping, filtering, and resizing, followed by manual annotation of fetal heart anatomy guided by expert knowledge. The dataset is then divided into training and testing sets. The model configuration includes hyperparameter tuning as well as refinement of anchor boxes and prediction layers within the region proposal network (RPN). The model is trained iteratively using various configurations. Its performance is evaluated using mean average precision (mAP), which reflects the accuracy of object detection across different recall levels in medical image analysis.

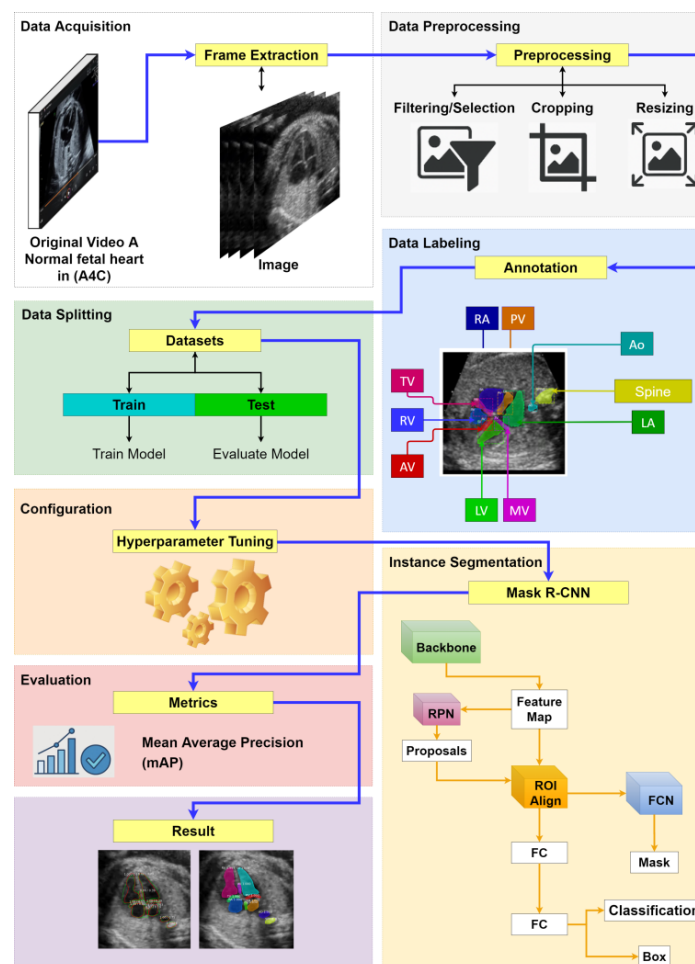


Figure 1. The flowchart of the AI-based models and experimental methods applied

2.1. Data acquisition

The initial phase of this study began with the acquisition of fetal echocardiography videos obtained from authorized online sources [34]. These videos display the fetal heart from the A4C perspective and are

provided in .mp4 format, with a file size of 13.7 MB, a duration of 178 seconds, and a frame rate of 30 fps. The entire video was converted into two-dimensional images with a resolution of 1280×720 pixels through a frame extraction process.

2.2. Data pre-processing

After the frame extraction process, a preprocessing stage was carried out to filter and select relevant images, ensuring the quality of the data used for model training. This stage consists of three main steps: filtering, cropping, and resizing. Filtering was performed to retain only the images that clearly depict fetal heart structures [35]. Cropping was applied to focus on the regions containing the fetal heart; in some cases, multiple crops were taken from a single image if it contained more than one fetal heart object. Finally, resizing was performed to standardize the image dimensions, with all images resized to 400×300 pixels.

2.3. Data labeling

Subsequently, the selected normal fetal heart images were manually annotated by fetal cardiology experts using a specialized graphical annotation tool, namely the makesense application [36]. The annotation process was conducted individually for each image, guided by expert knowledge of fetal cardiac anatomy. The annotated objects included: LA, RA, LV, RV, TV, PV, MV, AV, Ao, and spine. The annotation results were saved in JSON format and served as the ground truth for model training.

2.4. Data splitting

Following the annotation process, the dataset was divided into two primary subsets: training data and validation data, using an 80:20 split ratio. The splitting was performed randomly while ensuring that the class distribution remained balanced across both subsets. This approach allows the model to learn from the majority of the available data while reserving a portion for evaluating its generalization performance on unseen samples. Such a strategy is commonly employed in DL workflows to prevent overfitting and ensure an unbiased performance assessment.

2.5. Configuration

Prior to training, a hyperparameter tuning process was conducted, including the configuration of anchor boxes, learning rate, batch size, and number of epochs. The proposed model was developed and trained on a computer equipped with an Intel Core i3-4170 CPU @ 3.70 GHz (4 CPUs), 8 GB of RAM, and an Nvidia GeForce GTX 1050 Ti GPU featuring 768 CUDA cores, a GPU clock speed of 1392/1506 MHz, 4 GB of GDDR5 memory, and a memory bandwidth of 112.1 GB/s. The programming language used was Python 3.6.13, with TensorFlow 1.14.0, Keras 2.3.1, and Protobuf 3.19.6 libraries.

2.6. Instance segmentation

In the subsequent stage, the Mask region-based convolutional neural network (Mask R-CNN) [37] instance segmentation model is employed. This model consists of several key components: a backbone network (ResNet50) for feature extraction, a RPN for generating candidate object regions, region of interest (ROI) aligns for aligning proposed regions with the feature maps, fully connected layers for bounding box classification and regression; and a fully convolutional network (FCN) for generating binary masks of each detected object. The model is specifically designed to perform segmentation of the normal fetal heart anatomy based on the A4C view. Mask R-CNN was selected due to its ability to perform both object detection and instance-level segmentation with high accuracy. Mask R-CNN offers better performance on medical imaging datasets with limited data and complex object boundaries. No structural modifications were made to the original architecture, but model performance was optimized through hyperparameter tuning specific to fetal heart image characteristics.

2.7. Evaluation metrics

To evaluate the overall performance of the model, a specific metric called mAP was used. mAP is a widely employed metric for assessing the quality of object detectors. This metric measures the accuracy of the model in detecting objects by calculating the average precision (AP) for each class. mAP provides valuable insights into the performance of the DL model in the task of detecting fetal heart objects. To obtain the mAP value [17], the AP is first calculated by combining precision and recall at various threshold levels. The equations for AP and mAP are provided in (1) and (2), respectively:

$$AP = \sum_k (\text{Precision at recall point } k \times \Delta \text{recall}_k) \quad (1)$$

where precision at recall point k is the precision value at a specific recall; and $\Delta recall_k$ is the change in recall between two adjacent recall points.

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (2)$$

where N is the number of classes or objects; and AP_i is the AP for the i -th class.

To calculate precision and recall, use (3) and (4). Precision measures how many of the predicted positive cases are truly positive, and it decreases when there are many false positives. Recall indicates how many actual positive cases are correctly detected, and it decreases with high false negatives. Together, these values determine AP, which is then averaged to compute mAP, giving a robust overall measure of object detection performance.

$$P = \frac{TP}{TP + FP} \quad (3)$$

$$R = \frac{TP}{TP + FN} \quad (4)$$

where P is precision; R is recall, TP is true positive; and FP is false positive.

3. RESULTS AND DISCUSSION

3.1. Pre-processing of normal fetal heart image data

Following the preprocessing to obtain fetal heart images, the process involved converting ultrasound videos into still images, resulting in a total of 357 images. These images include those showing fetal heart objects, with some images containing one, two, or three fetal heart objects. Additionally, there are images that do not show any fetal heart objects and those where the fetal heart objects are out of focus or blurred, as illustrated in Figure 2. For images containing multiple fetal heart objects or where other text or objects are present in the image, cropping is performed to ensure that the data used for the instance segmentation model meets the specific requirements. This process aligns with the steps outlined in method section. The output of the video extraction process and the resulting images are summarized in Table 1.

Table 1. Video extraction

No	Image type	Number of extracted images
1.	Images showing fetal heart objects	114
2.	Image showing multiple fetal heart objects	50
3.	Images showing fetal heart objects but out of focus	105
4.	Images not showing any fetal heart objects	88
	Total	357

Table 1 presents the results of image extraction from fetal heart examination videos, categorized into four main groups based on the quality and presence of fetal heart structures. A total of 357 images with a resolution of 1280×720 pixels were obtained. Most of the images contain fetal heart objects with varying levels of clarity and object count, while others lack relevant features for further analysis. This classification supports the selection of suitable images for the annotation and model training stages.

Visually, Figure 2 illustrates four main categories resulting from the image extraction process. Figure 2(a) displays images that do not display any fetal heart object, Figure 2(b) presents images that contain a fetal heart object but are out of focus, Figure 2(c) shows images that clearly show a single fetal heart object, and Figure 2(d) presents images that display multiple fetal heart objects within a single frame. These categories are derived from the video-to-image conversion process and will subsequently undergo preprocessing as part of the dataset preparation for training the segmentation model.

Following the cropping and selection process for images displaying fetal heart objects, the total number of images was reduced to 176, which aligns with the requirements for the instance segmentation model, as shown in Figure 3. After obtaining the fetal heart images, the next step involved scaling the images to ensure uniform size across the dataset. The scaling process was conducted as described in the method section, with images resized to 400×300 pixels. Following this, all normal fetal heart images were annotated with ten labels corresponding to the anatomical features of the fetal heart. This annotation was performed using polygon points on the fetal heart object images. The annotation process is illustrated in Figure 4. The final annotated fetal heart images were exported in JSON file format.

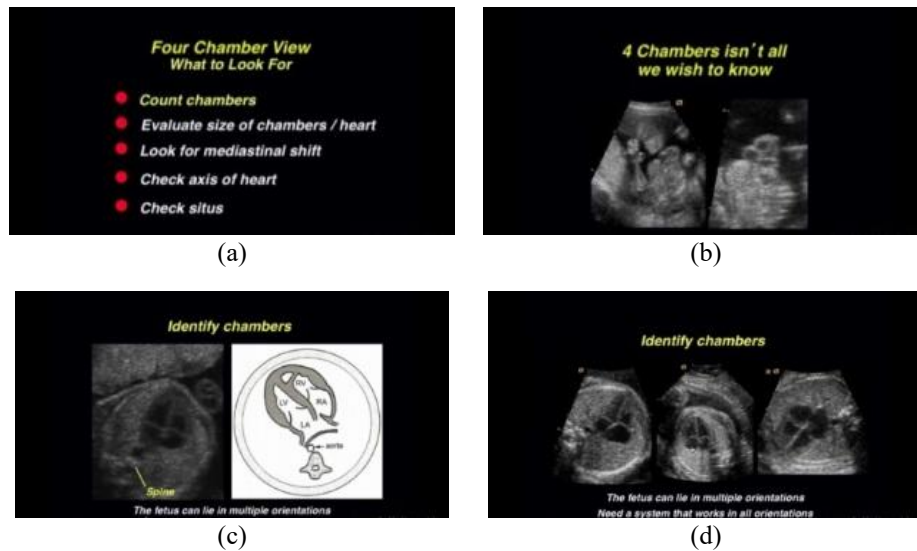


Figure 2. Four image categories from the extraction process of (a) not showing any fetal heart objects, (b) showing fetal heart objects but out of focus, (c) showing a fetal heart object, and (d) showing multiple fetal heart object



Figure 3. Fetal heart images from the A4C

Figure 4(a) represents the anatomical location of the fetal heart that has been determined based on expert designation, but has not gone through the AI-based modeling stage. This identification is done manually by the radiologist or specialist by considering the visual characteristics seen on the ultrasound image. The location of anatomical structures in this image serves as the ground truth, which becomes the reference in further annotation and modeling stages. Meanwhile, Figure 4(b) is the result of annotation performed using annotation tools, where each fetal heart structure has been labeled with a color mask and bounding box to distinguish specific areas. This annotation is an important part of preparing the dataset for training AI-based segmentation models.

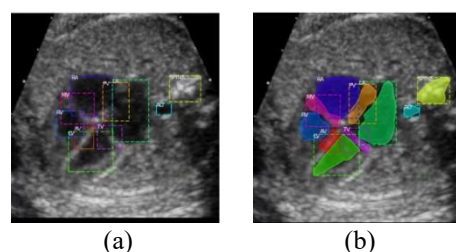


Figure 4. Annotation of fetal heart images (a) original image with manual identification and (b) annotated image with color masks and bounding boxes

After the annotation phase is complete, the JSON annotation files are paired with the annotated images. This combined dataset is then used to train the instance segmentation model for fetal heart image

objects. A sample of the annotation results is shown in Figure 5. Figure 5 shows the results of ground truth annotation for segmentation of anatomical structures in fetal heart ultrasound images. Figure 5(a) displays the original ultrasound image, while Figures 5(b) to 5(k) represent the manually annotated segmentation of various heart structures. The structures shown include Ao, AV, LA, LV, MV, PV, RA, RV, spine, and TV. The masking visualized in Figures 5(b) to 5(k) shows the areas identified as part of each anatomical structure based on the ground truth annotations. This image is generated from annotated data in JSON format imported into Python code and visualized using image processing libraries such as OpenCV or Matplotlib. The process involves mapping the JSON data into an array of binary images for each anatomical structure, then visualized against a blue background to clarify the segmented parts.

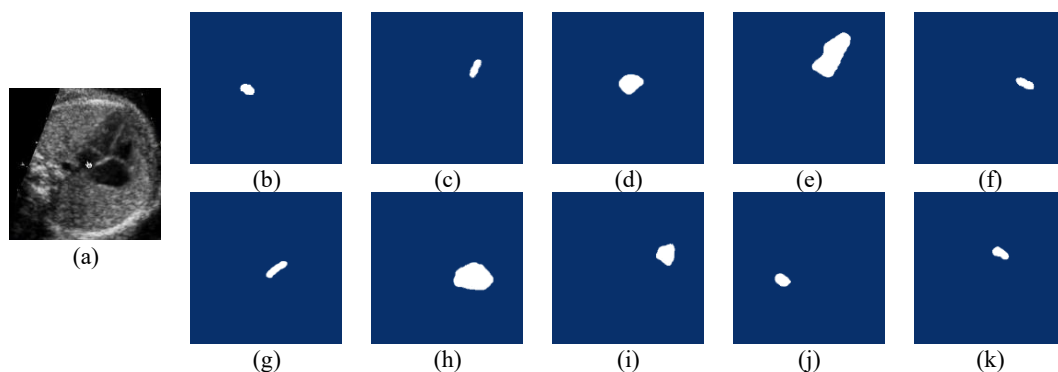


Figure 5. Ground truth of annotation results (a) original image, (b) Ao, (c) AV, (d) LA, (e) LV, (f) MV, (g) PV, (h) RA, (i) RV, (j) spine, and (k) TV

3.2. Splitting data

This study utilizes a dataset that is divided into two parts: the training set and the validation set. The training set is used to train the model, while the validation set is used to evaluate the model's performance on data that was not seen during the training process. Out of the total 176 images, the dataset is split into 140 images for the training set and 36 images for the validation set, as shown in Table 2.

Table 2. Dataset split for training and validation sets

No	Data	Number of images
1.	Training data	140
2.	Validation data	36

3.3. Model segmentation design

This study employs the Mask R-CNN method, optimizing the model by fine-tuning the hyperparameters specific to Mask R-CNN. The hyperparameters used are listed in Table 3. These hyperparameters are critical for improving accuracy with the image data. The selected hyperparameters result in 24 model combinations. According to Table 4, the parameters include image size, learning rate, learning momentum, with 8 epochs and 500 steps per epoch, a ResNet-50 backbone architecture, stochastic gradient descent (SGD) optimizer, and a batch size of 1.

Table 3. Hyperparameters used for Mask R-CNN model training

No	Configurations	Hyperparameter
1.	Image size	[64,128,256,512]
2.	Learning rate	[0.01,0.001,0.0001]
3.	Learning momentum	[0.7,0.9]

Table 4 presents 24 combinations of convolutional neural network (CNN) architecture models experimented to evaluate the impact of various hyperparameters on classification performance. The varied parameters include input image size, learning momentum, and learning rate. All models use the ResNet-50 backbone, optimized with the SGD algorithm, and trained for 8 epochs with 500 steps per epoch, and with a batch size of 1.

Table 4. Model experimentations with ResNet-50 backbone and SGD optimizer

Model	Size	Learning momentum	Learning rate
R50_sgd_1	(64,64)	0.7	0.01
R50_sgd_2	(64,64)	0.9	0.01
R50_sgd_3	(64,64)	0.7	0.001
R50_sgd_4	(64,64)	0.9	0.001
R50_sgd_5	(64,64)	0.7	0.0001
R50_sgd_6	(64,64)	0.9	0.0001
R50_sgd_7	(128,128)	0.7	0.01
R50_sgd_8	(128,128)	0.9	0.01
R50_sgd_9	(128,128)	0.7	0.001
R50_sgd_10	(128,128)	0.9	0.001
R50_sgd_11	(128,128)	0.7	0.0001
R50_sgd_12	(128,128)	0.9	0.0001
R50_sgd_13	(256,256)	0.7	0.01
R50_sgd_14	(256,256)	0.9	0.01
R50_sgd_15	(256,256)	0.7	0.001
R50_sgd_16	(256,256)	0.9	0.001
R50_sgd_17	(256,256)	0.7	0.0001
R50_sgd_18	(256,256)	0.9	0.0001
R50_sgd_19	(512,512)	0.7	0.01
R50_sgd_20	(512,512)	0.9	0.01
R50_sgd_21	(512,512)	0.7	0.001
R50_sgd_22	(512,512)	0.9	0.001
R50_sgd_23	(512,512)	0.7	0.0001
R50_sgd_24	(512,512)	0.9	0.0001

3.4. Results of Mask R-CNN model optimization

The evaluation of the Mask R-CNN model was conducted by calculating the mAP, which reflects the overall accuracy of the model. The results of the model evaluation from various experiments are detailed in Table 5. This table presents the mAP results for each category within the training dataset, indicating that the model has successfully learned to recognize all classes.

Table 5. AP and mAP for training dataset original

Model	AP (IoU)=50										mAP
	Ao	LA	LV	RV	RA	Spine	TV	MV	PV	AV	
R50_sgd_1	0.000	0.278	0.244	0.278	0.278	0.000	0.000	0.000	0.000	0.000	0.1078
R50_sgd_2	0.000	0.000	0.222	0.247	0.222	0.000	0.000	0.000	0.000	0.000	0.0691
R50_sgd_3	0.000	0.056	0.188	0.222	0.222	0.000	0.000	0.000	0.000	0.000	0.0688
R50_sgd_4	0.000	0.139	0.278	0.191	0.250	0.000	0.000	0.000	0.000	0.000	0.0858
R50_sgd_5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0
R50_sgd_6	0.000	0.000	0.139	0.000	0.028	0.000	0.000	0.000	0.000	0.000	0.0167
R50_sgd_7	0.139	0.278	0.278	0.278	0.278	0.222	0.000	0.000	0.000	0.000	0.1473
R50_sgd_8	0.244	0.278	0.278	0.250	0.278	0.250	0.000	0.000	0.000	0.000	0.1578
R50_sgd_9	0.000	0.194	0.278	0.250	0.278	0.000	0.000	0.000	0.000	0.000	0.1
R50_sgd_10	0.028	0.278	0.278	0.250	0.278	0.163	0.000	0.000	0.000	0.000	0.1275
R50_sgd_11	0.000	0.000	0.194	0.056	0.111	0.000	0.000	0.000	0.000	0.000	0.0361
R50_sgd_12	0.000	0.139	0.222	0.250	0.250	0.000	0.000	0.000	0.000	0.000	0.0861
R50_sgd_13	0.028	0.083	0.056	0.083	0.278	0.028	0.000	0.000	0.000	0.000	0.0556
R50_sgd_14	0.000	0.250	0.250	0.222	0.194	0.167	0.037	0.000	0.000	0.000	0.112
R50_sgd_15	0.250	0.222	0.250	0.219	0.278	0.219	0.054	0.000	0.000	0.000	0.1492
R50_sgd_16	0.278	0.250	0.278	0.278	0.278	0.250	0.151	0.000	0.000	0.000	0.1763
R50_sgd_17	0.000	0.000	0.139	0.111	0.056	0.000	0.000	0.000	0.000	0.000	0.0306
R50_sgd_18	0.056	0.222	0.216	0.222	0.222	0.056	0.000	0.000	0.000	0.000	0.0994
R50_sgd_19	0.842	0.250	0.278	0.278	0.278	0.222	0.231	0.188	0.182	0.000	0.2749
R50_sgd_20	0.565	0.278	0.278	0.278	0.278	0.250	0.267	0.25	0.155	0.042	0.2641
R50_sgd_21	0.333	0.278	0.250	0.239	0.278	0.111	0.007	0.000	0.000	0.007	0.1503
R50_sgd_22	0.278	0.278	0.278	0.278	0.278	0.222	0.165	0.042	0.157	0.125	0.2101
R50_sgd_23	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0
R50_sgd_24	0.222	0.159	0.099	0.028	0.222	0.134	0.000	0.000	0.000	0.000	0.0864

Table 5 presents the Mask R-CNN model evaluation results based on AP at intersection over union (IoU)=50 for each fetal heart anatomy category in the training dataset as well as the mAP as a measure of overall model performance. Based on the results obtained, models R50_sgd_19 and R50_sgd_20 showed the best performance with mAP of 0.2749 and 0.2641, indicating the ability to recognize various anatomical structures more accurately than other models. Cardiac structures such as right RV, LV, RA, LA, and AV tended to have higher AP values, indicating that the models were able to recognize these parts better than

other structures, such as TV or PV, which had lower or even zero AP values. The evaluation results also show that there are some models with AP value=0.000 in certain categories, indicating that the model failed to detect objects of that class in the training dataset. This could be due to various factors, such as a limited amount of annotation data or the complexity of anatomical structures that are difficult for the model to recognize. In addition, models such as R50_sgd_5 and R50_sgd_23 have mAP=0, indicating that they did not successfully segment any objects in the dataset. Models with higher mAP show better performance in detecting and labeling fetal heart structures, while models with many values of 0.000 or mAP=0 show weaknesses in the learning process from the available data.

Figure 6 displays the mAP for various ResNet-50 models trained using the SGD optimizer with different hyperparameter combinations. mAP is a commonly used metric to evaluate the performance of object detection models, with higher values indicating better performance. From the graph, it is evident that models R50_sgd_19 and R50_sgd_20 achieved the best results, with mAP values of approximately 0.27 and 0.26, respectively. This suggests that models with an input image size of 512×512 and a learning rate of 0.01 perform better in detecting objects within the dataset used. Other models, such as R50_sgd_1, R50_sgd_7, and R50_sgd_15, also showed fairly good performance with mAP values ranging from 0.1 to 0.15. However, their performance was still below that of models R50_sgd_19 and R50_sgd_20. Some models exhibited very low or even zero performance, such as R50_sgd_5 and R50_sgd_23. This may be attributed to suboptimal hyperparameter combinations for the dataset. Overall, these results highlight the importance of selecting the appropriate image input size and learning rate to achieve optimal performance in object detection models using the ResNet-50 architecture with the SGD optimizer.

Although Mask R-CNN is a well-established method, this study presents a novel application by integrating instance segmentation with targeted hyperparameter optimization tailored for A4C fetal heart ultrasound images. The combination of input resolution tuning, learning rate, and momentum on a dataset with ten anatomical classes represents a unique contribution, as previous studies typically limited segmentation to fewer structures or did not perform systematic model optimization. This approach addresses the complexity of fetal cardiac imaging and demonstrates improved class-wise recognition in a clinically relevant context.

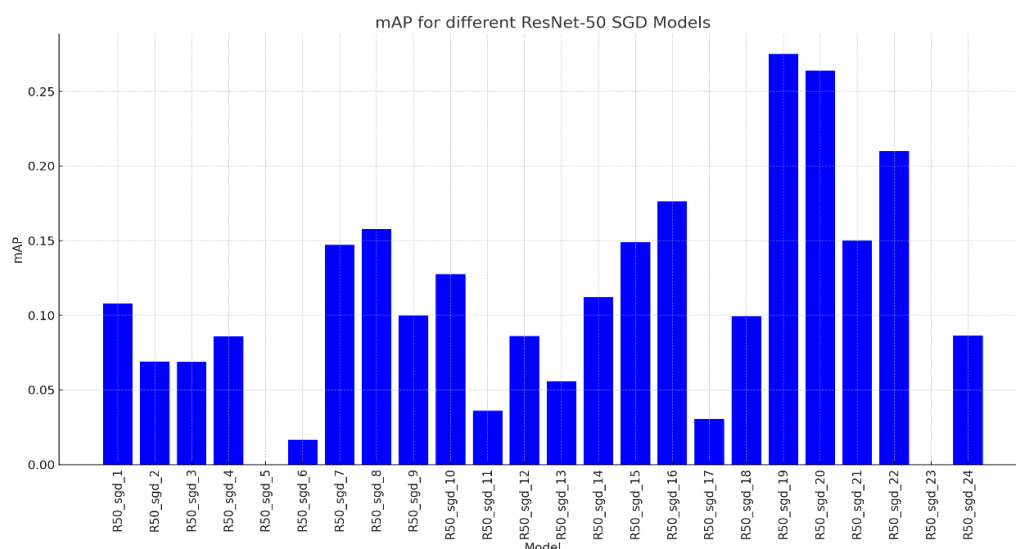


Figure 6. Results and analysis of model performance

The Figure 7 illustrates the AP at an IoU threshold of 50 for each class across various ResNet-50 models trained with the SGD optimizer. Each line in the graph represents a class, with AP values for each model plotted as points along that line. The analysis reveals that the class Ao demonstrates significant performance variation across models, with some models such as R50_sgd_19 and R50_sgd_20 achieving high AP values. Other classes, including LA, LV, and RV, also show noticeable variation in performance among the tested models. Models R50_sgd_19 and R50_sgd_20 exhibit more consistent performance across many classes compared to others. Certain classes like spine, TV, MV, PV, and AV frequently show low or even zero AP values in many models, indicating that detection for these classes is more challenging. Overall, models with larger input image sizes and lower learning rates appear to deliver better and more consistent results across various classes. The best-performing model in this evaluation is R50_sgd_19, which demonstrates the highest performance across most classes. Out of the 24 identified models, named from

R50_sgd_1 to R50_sgd_24, the research selected two models with optimal detection performance for classes such as Ao, LA, LV, RV, RA, TV, MV, PV, AV, and spine. The evaluation, based on mAP values, identified the first optimal model as R50_sgd_19, which achieved the highest mAP of 0.2749, although it failed to detect the AV class. The second model, R50_sgd_20, successfully detected all classes with a mAP of 0.2641. Both models demonstrated strong overall performance. Table 6 presents these two optimal models based on the results from Table 5. These models were selected due to their high mAP values and consistent performance across most fetal heart anatomical classes.

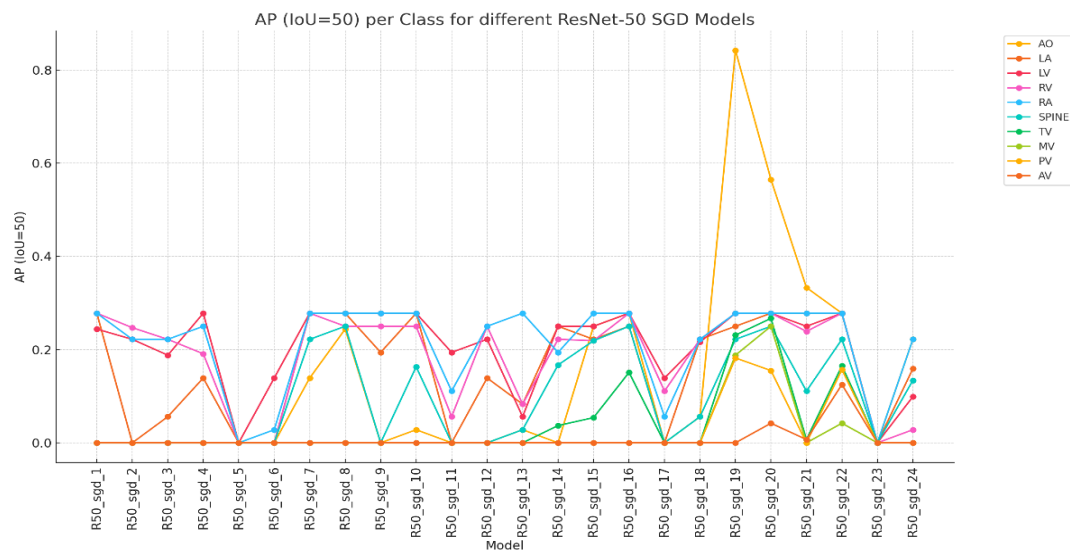


Figure 7. Comparison of AP values between object classes of fetal heart images

Table 6. Optimal parameters for Mask R-CNN models

Method	Model	Parameters
Mask R-CNN	R50_sgd_19	Image size: 512×512, learning momentum: 0.7, learning rate: 0.01
	R50_sgd_20	Image Size: 512×512, learning momentum: 0.9, learning rate: 0.01

The results of instance segmentation from the two optimal models are displayed in Figure 8. This figure shows the segmentation results for ten classes of fetal heart objects. These segmentation outputs are essential to evaluate the model’s ability to differentiate each anatomical structure accurately. Figure 8(a) shows the segmentation results for several anatomical structures in a medical image, likely an echocardiogram of the heart. The segmentation successfully identifies and labels several key parts of the image with high confidence levels, including: RV with a confidence of 0.995 LV with a confidence of 0.999 AV with a confidence of 0.985 MV with a confidence of 0.975 RA with a confidence of 1.000 PV with a confidence of 0.972 LA with a confidence of 0.999 TV with a confidence of 0.970 Ao with a confidence of 0.995 spine with a confidence of 0.997. This segmentation demonstrates that the model has very high accuracy in identifying and labeling various anatomical structures within the medical image. Each segment is clearly delineated, and the high confidence values suggest that this model is reliable for diagnostic purposes and further medical analysis. These results are highly favorable for medical applications, particularly in assisting physicians with the identification and analysis of critical parts of echocardiographic images. Figure 8(b) presents the segmentation results of several key anatomical structures in an echocardiographic image, with extremely high confidence levels. Detailed explanations for each identified structure are as follows. RV: this structure is identified with a confidence of 0.998, indicating that the model is highly confident in its identification. LV: similarly, the LV is identified with a very high confidence of 0.998. AV: this valve is identified with a confidence of 0.954. Although slightly lower than other structures, this value remains very high. MV: marked with a confidence of 0.993, indicating nearly perfect confidence in identifying this valve. RA: with a confidence of 0.999, the RA is identified with nearly perfect confidence. PV: this valve is identified with a confidence of 0.945, which remains within a high confidence range. TV: with a confidence of 0.988, the TV is segmented with very good accuracy. Ao: this structure is marked with a confidence of 0.998, indicating highly accurate identification. Spine: the spine is segmented with a confidence of 0.996, showing high confidence in the identification of this structure. Overall, this

segmentation demonstrates that the model excels in identifying various important anatomical structures in echocardiographic images. With nearly perfect confidence values for most structures, the model is highly reliable for diagnostic and further medical analysis.

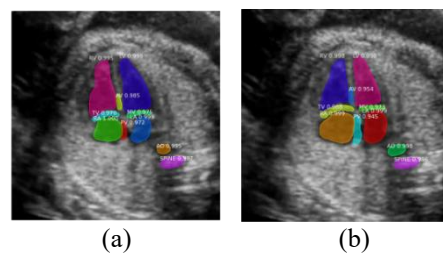


Figure 8. Instance segmentation results instance segmentation of fetal heart objects using model:
(a) R50_sgd_19 and (b) R50_sgd_20

3.5. Comparative review with previous studies

Unlike previous studies that focused solely on segmenting the four main heart chambers or performing object detection without segmentation, this study incorporates all anatomically and clinically relevant structures for comprehensive diagnosis. The inclusion of the spine, along with eight additional cardiac structures, enables a more complete interpretation of fetal cardiac conditions and supports early screening for abnormalities such as tetralogy of fallot, ventricular septal defects (VSD), and cardiac malposition. Therefore, the proposed model not only expands the number of anatomical classes identified but also enhances the clinical relevance of the segmentation results.

This research develops an instance segmentation approach using the DL method with ResNet architecture, which is optimized through hyperparameter settings to produce more accurate fetal heart anatomy segmentation. In this study, a total of ten anatomical objects in the fetal heart were successfully segmented, namely: LA, RA, LV, RV, TV, PV, MV, AV, Ao, and spine. The addition of the spine in this segmentation provides additional useful information for medical practitioners in comprehensively identifying the four-chambered view of the heart, which is important in the diagnosis of fetal heart health. The results of this study show that instance segmentation with the ResNet architecture backbone is able to provide more comprehensive results that are capable of detection and segmentation with many objects in fetal heart anatomy. This approach has the potential to improve medical image analysis in the health sector, especially in detecting fetal heart disease more accurately through ultrasound images. This research is a continuation of previous studies, as can be seen from Table 7 comparison with previous research results.

Table 7. Comparison of previous research with current research

Researcher	Method	Segmented objects	Anatomy coverage	Disease coverage	Segmentation type
Sapitri <i>et al.</i> [20]	Anatomical structure detection	Detection object without segmentation	Nine objects in the fetal heart image	None	Detection (without segmentation)
Nurmaini <i>et al.</i> [29]	Instance segmentation (ResNet)	Fetal heart image object	Four major chambers of the fetal heart image and heart disease	Heart disease	Segmentation
This research	Instance segmentation (ResNet + tuning hyperparameter)	Ten anatomical objects (LV, RV, TV, PV, MV, AV, Ao, Spine)	Nine objects of fetal heart and spine image	Focus on normal structures to detect abnormalities	Complete object segmentation

This study shows significant improvements in fetal cardiac anatomy segmentation using the DL method based on the ResNet architecture, which has been optimized through hyperparameter settings. By implementing the segmentation instance, this study successfully identified ten major anatomical objects in the fetal heart, including the four main chambers, important valves, aorta, and spine. The addition of the spine in this segmentation provides more comprehensive information, which is beneficial in the identification of the complete four-chambered view of the heart, aiding in the diagnosis of fetal heart health. Although this study is superior in anatomical object coverage compared to previous studies (which only focused on four chambers or no segmentation at all), the results show that the mAP value achieved is still relatively low. This is due to the challenge of identifying more diverse classes of objects in the fetal heart image, as well as the high level of noise in the video. These conditions degrade the accuracy of the model in detecting and classifying objects accurately, which impacts the overall performance of the segmentation. In addition, the

noise factor in ultrasound images can complicate the segmentation process as it depends on the varying video quality. Compared to previous studies that have fewer object classes, this study shows limitations in accurately analyzing more objects under non-ideal video conditions. Nonetheless, this study still shows the potential to improve fetal heart image detection and segmentation in the future through improved preprocessing techniques and improved quality of the video data used.

In addition to prior works focusing on image segmentation and detection, recent studies have explored alternative approaches to fetal cardiac analysis, such as digital twin modeling and entropy-based analysis of fetal heart rate variability (HRV). For instance, Lwin *et al.* [38] proposed a digital twin framework combined with entropy measures to enhance fetal monitoring systems, demonstrating how physiological signal analysis can complement image-based techniques. Similarly, Zin and Tin [39] applied Markov chain models to analyze HRV, highlighting the integration of AI with physiological data for diagnostic support. While these approaches differ in modality, they align with the broader goal of improving fetal cardiac assessment using AI, and this study complements them by advancing structural image-based segmentation.

4. CONCLUSION

This study successfully applied Mask R-CNN for sample segmentation on fetal heart ultrasound images, able to identify and label anatomical structures. The R50_sgd_19 and R50_sgd_20 models showed good performance, with mAP values of 0.2749 and 0.2641, respectively. These models accurately detected and labeled major cardiac structures including RV, LV, AV, MV, RA, PV, LA, TV, Ao, and spine with confidence values ranging from 0.970 to 1.000, demonstrating the robustness of the models. Systematic data preprocessing, annotation, hyperparameter optimization, and model training were critical to this success. The results provide a valuable tool for medical practitioners, enabling more precise diagnosis and contributing significantly to the assessment of fetal heart health. Furthermore, this research can serve as a foundation for the integration of AI-based diagnostic support in fetal cardiology. Future research can explore advanced architectures, dataset expansion, integration with other imaging modalities, real-time clinical applications, and user-friendly interfaces to further improve the utility and accuracy of the model.

ACKNOWLEDGMENTS

We would like to express very great appreciation to Universitas Sumatera Selatan and Lembaga Penelitian dan Pengabdian kepada Masyarakat Universitas Sumatera Selatan (LPPM) for support during the development of this research work.

FUNDING INFORMATION

This research was supported by financial assistance from Universitas Sumatera Selatan for publication funding.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Hadi Syaputra	✓	✓	✓		✓			✓	✓	✓				
Siti Nurmaini	✓	✓		✓						✓		✓		
Radiyah Umi Partan	✓	✓				✓				✓		✓		
Muhammad Taufik			✓			✓	✓	✓		✓				
Roseno														

C : **C**onceptualization

M : **M**ethodology

So : **S**oftware

Va : **V**alidation

Fo : **F**ormal analysis

I : **I**nvestigation

R : **R**esources

D : **D**ata Curation

O : **O**riginal Draft

E : **E**diting

Vi : **V**isualization

Su : **S**upervision

P : **P**roject administration

Fu : **F**unding acquisition

CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

DATA AVAILABILITY

The authors confirm that the data supporting the findings of this study are available within the article.




REFERENCES

- [1] Z. Akkus *et al.*, "A survey of deep-learning applications in ultrasound: artificial intelligence-powered ultrasound for improving clinical workflow," *Journal of the American College of Radiology*, vol. 16, no. 9, pp. 1318–1328, 2019, doi: 10.1016/j.jacr.2019.06.004.
- [2] R. Medaglia, J. R. G.-Garcia, and T. A. Pardo, "Artificial intelligence in government: taking stock and moving forward," *Social Science Computer Review*, vol. 41, no. 1, pp. 123–140, 2023, doi: 10.1177/08944393211034087.
- [3] M. Y. A.-Kader, A. M. Ebid, K. C. Onyelowe, I. M. Mahdi, and I. A.-Rasheed, "(AI) in infrastructure projects—gap study," *Infrastructures*, vol. 7, no. 10, 2022, doi: 10.3390/infrastructures7100137.
- [4] X. Liu, K. H. Ghazali, and A. A. Shah, "Sustainable oil palm resource assessment based on an enhanced deep learning method," *Energies*, vol. 15, no. 12, 2022, doi: 10.3390/en15124479.
- [5] T. Davenport and R. Kalakota, "The potential for artificial intelligence in healthcare," *Future Healthcare Journal*, vol. 6, no. 2, pp. 94–98, 2019, doi: 10.7861/futurehosp.6-2-94.
- [6] A. M. Hassan, J. A. Nelson, J. H. Coert, B. J. Mehrara, and J. C. Selber, "Exploring the potential of artificial intelligence in surgery: insights from a conversation with ChatGPT," *Annals of Surgical Oncology*, vol. 30, no. 7, 2023, doi: 10.1245/s10434-023-13347-0.
- [7] F. Qin and J. Gu, "Artificial intelligence in plastic surgery: current developments and future perspectives," *Plastic and Aesthetic Research*, vol. 10, no. 1, 2023, doi: 10.20517/2347-9264.2022.72.
- [8] S. Oh, J. H. Kim, S.-W. Choi, H. J. Lee, J. Hong, and S. H. Kwon, "Physician confidence in artificial intelligence: an online mobile survey," *Journal of Medical Internet Research*, vol. 21, no. 3, 2019, doi: 10.2196/12422.
- [9] L. Pullagura, M. R. Dontha, and S. Kakumanu, "Recognition of fetal heart diseases through machine learning techniques," *Annals of the Romanian Society for Cell Biology*, vol. 25, no. 6, pp. 2601–2615, 2021.
- [10] P. G.-Canadilla, S. S.-Martinez, F. Crispi, and B. Bijmens, "Machine learning in fetal cardiology: what to expect," *Fetal Diagnosis and Therapy*, vol. 47, no. 5, pp. 363–372, 2020, doi: 10.1159/000505021.
- [11] Z. Hoodbhoy, M. Noman, A. Shafique, A. Nasim, D. Chowdhury, and B. Hasan, "Use of machine learning algorithms for prediction of fetal risk using cardiotocographic data," *International Journal of Applied and Basic Medical Research*, vol. 9, no. 4, pp. 226–230, 2019, doi: 10.4103/ijabmr.IJABMR_370_18.
- [12] Z. Cömert and A. Kocamaz, "A study of artificial neural network training algorithms for classification of cardiotocography signals," *Bitlis Eren University Journal of Science and Technology*, vol. 7, no. 2, pp. 93–103, 2017, doi: 10.17678/beuscitech.338085.
- [13] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017, doi: 10.1016/j.media.2017.07.005.
- [14] H. Zhou *et al.*, "A deep learning approach for medical waste classification," *Scientific Reports*, vol. 12, no. 1, 2022, doi: 10.1038/s41598-022-06146-2.
- [15] M. C. Fiorentino, F. P. Villani, M. D. Cosmo, E. Frontoni, and S. Moccia, "A review on deep-learning algorithms for fetal ultrasound-image analysis," *Medical Image Analysis*, vol. 83, 2023, doi: 10.1016/j.media.2022.102629.
- [16] Y. Matsuzaka and R. Yashiro, "AI-based computer vision techniques and expert systems," *AI*, vol. 4, no. 1, pp. 289–302, 2023, doi: 10.3390/ai4010013.
- [17] Z. Soleimanitaleb and M. A. Keyvanrad, "Single object tracking: a survey of methods, datasets, and evaluation metrics," *arXiv-Computer Science*, pp. 1–15, 2022.
- [18] A. A. Adegun, S. Viriri, and R. O. Ogundokun, "Deep learning approach for medical image analysis," *Computational Intelligence and Neuroscience*, vol. 2021, no. 1, 2021, doi: 10.1155/2021/6215281.
- [19] Y. Dai, Y. Gao, and F. Liu, "TransMed: transformers advance multi-modal medical image classification," *Diagnostics*, vol. 11, no. 8, 2021, doi: 10.3390/diagnostics11081384.
- [20] A. I. Sapitri *et al.*, "Deep learning-based real time detection for cardiac objects with fetal ultrasound video," *Informatics in Medicine Unlocked*, vol. 36, 2023, doi: 10.1016/j.imu.2022.101150.
- [21] D. V. C. Gowda and R. Kanagavalli, "Video semantic segmentation with low latency," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 22, no. 5, pp. 1147–1156, 2024, doi: 10.12928/TELKOMNIKA.v22i5.25157.
- [22] S. An *et al.*, "A category attention instance segmentation network for four cardiac chambers segmentation in fetal echocardiography," *Computerized Medical Imaging and Graphics*, vol. 93, 2021, doi: 10.1016/j.compmedimag.2021.101983.
- [23] M. N. Rachmatullah, S. Nurmaini, A. I. Sapitri, A. Darmawahyuni, B. Tutuko, and Firdaus, "Convolutional neural network for semantic segmentation of fetal echocardiography based on four-chamber view," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 4, pp. 1987–1996, 2021, doi: 10.11591/EEI.V10I4.3060.
- [24] H. Cheng *et al.*, "Semantic segmentation method for myocardial contrast echocardiogram based on DeepLabV3+ deep learning architecture," *Mathematical Biosciences and Engineering*, vol. 20, no. 2, pp. 2081–2093, 2022, doi: 10.3934/mbe.2023096.
- [25] A. M. Hafiz and G. M. Bhat, "A survey on instance segmentation: state of the art," *International Journal of Multimedia Information Retrieval*, vol. 9, no. 3, pp. 171–189, 2020, doi: 10.1007/s13735-020-00195-x.
- [26] V. Iglovikov, S. Seferbekov, A. Buslaev, and A. Shvets, "TernausNetV2: fully convolutional network for instance segmentation," *arXiv-Computer Science*, pp. 4321–4325, 2018, doi: 10.1109/CVPRW.2018.00042.
- [27] J. Yan, T. Yan, W. Ye, X. Lv, P. Gao, and W. Xu, "Cotton leaf segmentation with composite backbone architecture combining convolution and attention," *Frontiers in Plant Science*, vol. 14, 2023, doi: 10.3389/fpls.2023.1111175.
- [28] S. Nurmaini *et al.*, "Deep learning-based computer-aided fetal echocardiography: application to heart standard view segmentation for congenital heart defects detection," *Sensors*, vol. 21, no. 23, 2021, doi: 10.3390/s21238007.
- [29] S. Nurmaini *et al.*, "Accurate detection of septal defects with fetal ultrasonography images using deep learning-based multiclass instance segmentation," *IEEE Access*, vol. 8, pp. 196160–196174, 2020, doi: 10.1109/ACCESS.2020.3034367.
- [30] A. Boccatonda, "Emergency ultrasound: is it time for artificial intelligence?," *Journal of Clinical Medicine*, vol. 11, no. 13, 2022, doi: 10.3390/jcm11133823.
- [31] J. P. McGahan, "Sonography of the fetal heart: findings on the four-chamber view," *American Journal of Roentgenology*, vol. 156, no. 3, pp. 547–553, 1991, doi: 10.2214/ajr.156.3.1899755.




- [32] S. Iqbal, A. N. Qureshi, A. Ullah, J. Li, and T. Mahmood, "Improving the robustness and quality of biomedical CNN models through adaptive hyperparameter tuning," *Applied Sciences*, vol. 12, no. 22, 2022, doi: 10.3390/app122211870.
- [33] H. Y. Oh, M. S. Khan, S. B. Jeon, and M.-H. Jeong, "Automated detection of greenhouse structures using cascade Mask R-CNN," *Applied Sciences*, vol. 12, no. 11, 2022, doi: 10.3390/app12115553.
- [34] myminifellowship, "Mastering the fetal heart: step 1," *YouTube*. United States, 2015. Accessed: Mar. 31, 2023. [Online Video]. Available: <https://www.youtube.com/watch?v=xpxJFpORFmo>
- [35] B. Zhang, L. Niu, X. Zhao, and L. Zhang, "Human-centric image cropping with partition-aware and content-preserving features," *arXiv-Computer Science*, pp. 1–27, 2022.
- [36] P. Skalski, "Makesense: free to use online tool for labelling photos," *Kaggle*. 2023. Accessed: Mar. 25, 2023. [Online]. Available: <https://github.com/SkalskiP/make-sense>
- [37] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *2017 IEEE International Conference on Computer Vision*, 2017, pp. 2980–2988, doi: 10.1109/ICCV.2017.322.
- [38] T. C. Lwin, T. T. Zin, P. P. Kyaw, P. Tin, E. Kino, and T. Ikenoue, "Enhancing fetal monitoring through digital twin technology and entropy-based fetal heart rate variability analysis," *International Journal of Innovative Computing, Information and Control*, vol. 21, no. 1, pp. 185–196, 2025, doi: 10.24507/ijicic.21.01.185.
- [39] T. T. Zin and P. Tin, "Markov chain modelling for heart rate variability analysis: bridging artificial intelligence and physiological data," in *2023 IEEE 13th International Conference on Consumer Electronics - Berlin*, 2023, pp. 163–166, doi: 10.1109/ICCE-Berlin58801.2023.10375625.

BIOGRAPHIES OF AUTHORS






Hadi Syaputra    received a bachelor's degree in Computer Engineering from Universitas Bina Darma in 2006. Obtained a master's degree in Informatics Engineering at Universitas Bina Darma, Indonesia in 2011. Currently a researcher of USS Artificial Intelligence Research Group at the Universitas Sumatera Selatan, South Sumatra, Palembang, Indonesia. His research interests include deep learning, machine learning, image processing, and software engineering. He can be contacted at email: hadiisyaputra@uss.ac.id.






Siti Nurmaini    received the master's degree in Control System from the Institut Teknologi Bandung (ITB), Indonesia, in 1998, and the Ph.D. degree in Computer Science from the Universiti Teknologi Malaysia (UTM), in 2011. She is currently a Professor with the Faculty of Computer Science, Universitas Sriwijaya, Indonesia. Her research interests include biomedical engineering, deep learning, machine learning, image processing, control systems, and robotic. She can be contacted at email: siti_nurmaini@unsri.ac.id.



Radiyati Umi Partan    is a rheumatologist, staff at the Department of Internal Medicine, Faculty of Medicine, Universitas Sriwijaya in Palembang, South Sumatra Indonesia. She is member of the Indonesian Rheumatology Association and the Indonesian Osteoporosis Association. Graduated from the Faculty of Medicine, Universitas Sriwijaya in 1997. She completed specialist in Internal Medicine in 2008. Received master's degree in Pharmacology, a post-graduate program at Universitas Sriwijaya in 2010. Obtained a Doctoral degree and subspecialty in Rheumatology at Brawijaya University and completed it in 2016. Currently, she is in clinical practice and conducting several researches, particularly in rheumatology. Her research interests include osteoporosis, osteoarthritis, and systemic lupus erythematosus. She can be contacted at email: radiyati.u.p@fk.unsri.ac.id.



Muhammad Taufik Roseno    obtained a bachelor's degree in Industrial Engineering from Telkom Institute of Technology in 2001. Obtained a master's degree in Informatics Engineering from Bina Darma University, Indonesia in 2012. Currently a researcher of USS Artificial Intelligence Research Group at the Universitas Sumatera Selatan, South Sumatra, Palembang, Indonesia. His research interests include deep learning, machine learning, image processing, and software engineering. He can be contacted at email: mtroseno@gmail.com.