Fetal organ detection using feature enhancement with attention and residual block

Nuswil Bernolian[1], Siti Nurmaini[2,3], Ade Iriani Sapitri[2,3], Annisa Darmawahyuni[2,3], Muhammad Naufal Rachmatullah[2,3], Bambang Tutuko[2,3], Firdaus Firdaus[2,3]
[1]Department of Fetomaternal, Dr. Mohammad Hoesin General Hospital, Palembang, Indonesia
[2]Artificial Intelligence-Medical Center of Excellence, Universitas Sriwijaya, Palembang, Indonesia
[3]Intelligent System Research Group, Universitas Sriwijaya, Palembang, Indonesia

Article Info

Article history:

Received Mar 14, 2025
Revised Jan 25, 2026
Accepted Feb 6, 2026

Keywords:

Computer vision
Decision
Deep learning
Instance segmentation
Medical imaging
Ultrasound
YOLO

ABSTRACT

The rapid advancements in fetal ultrasonography have significantly enhanced prenatal diagnosis in recent years. Deep learning (DL) architectures have further streamlined the process of organ detection, improved diagnostic accuracy, and reduced observer dependency. This study proposes a computer-aided DL approach for fetal organ segmentation using the you only look once (YOLO) algorithm, a state-of-the-art method for object detection and image segmentation. This study identified and classified 15 fetal organs, including the umbilical vein, stomach, abdomen, brain (trans-cerebellum, trans-thalamic, and trans-ventricular regions), femur, head, thorax (chest cavity), heart (circumference, left atrium, left ventricle, right atrium, right ventricle), and aorta. We compared the performance of YOLOv7, YOLOv8, YOLOv9, and YOLOv11 architectures. The results showed that YOLOv9 outperformed YOLOv7, YOLOv8, and YOLOv11 achieving mAP50 and mAP95 scores of 91.90% and 94.50%, respectively. This performance surpasses previous studies that focused on classifying only a limited number of fetal organs.

Corresponding Author:

Siti Nurmaini
Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya
Palembang, Indonesia
Email: siti_nurmaini@unsri.ac.id

1. INTRODUCTION

Fetal growth is the most rapid phase of growth in humans, and exponential growth occurs by six folds between 22 and 40 weeks [1]. A prenatal ultrasound is the primary diagnostic tool for examining fetal growth [2]. A propensity score (PS) cross-sectional analysis of each fetal organ is extensively required to evaluate fetal growth. In current clinical practice, this assessment is routinely performed by capturing images of the fetal organs in utero and measuring the size of various anatomical structures [3]. However, inherent challenges with PS imaging—such as acoustic shadows, speckle noise, motion blur, and unclear boundaries—can also contribute to the low detection rates of scanning planes of fetal organs [2], [4]. Accurate assessment and detection of the scanning plane of fetal organs is essential for effective pregnancy management [5]. Guiding the prenatal ultrasound probe through the intricate fetal anatomy to obtain the correct scanning plane of the fetal organ demands extensive training and deep knowledge [2]. Manual navigation can be prone to errors due to the structural similarities between planes of the same fetal organ, making the process both challenging and time-consuming [3]. An automated algorithm can be utilized to detect accurate fetal organs, streamlining subsequent retrospective analysis. Moreover, such automation ensures reproducibility and eliminates both inter- and intra-observer variability [3].

Ultrasound technology plays a vital role in gynecology, serving as a key tool for assessing fetal development and identifying potential complications during pregnancy. One of its essential applications in medicine is the detection and monitoring of maternal-fetal health plans. Computer-aided deep learning (DL) is regarded as the foremost artificial intelligence (AI) technique in the detection and monitoring of maternal-fetal health plans [6]–[10] With the rise of DL algorithms, many image detection and recognition tasks can now be automated, reducing the need for highly skilled operators [2]. You only look once (YOLO), as a representative of DL algorithm, is the latest method of object detection and medical image segmentation [11]. The structure of YOLO is simple and efficient, allowing it to directly output the position and class of the bounding box (BBox) through its neural network [12]. Kumar and Goel [13] adopted YOLOv4 architecture to detect and classify the healthy and unhealthy fetal brain with the accuracy of 97.27%. Sapitri *et al*. [14] proposed YOLOv7 architecture to detect fetal cardiac substructure objects, and the results achieved the highest mean average precision (mAP) of 82.10%. Ramla *et al*. [15] explored YOLO architecture for fetal head circumference at different gestational age. Dandil *et al*. [16] identified fetal head, arm, heart and body using YOLOv5 architecture with average F1-score of 90.25%.

The previous studies showed the results reported are noteworthy. Nevertheless, to best our knowledge, the previous studies only limited to classify fetal organ, which evaluate fetal image and its annotation it in categories. However, its approach does not provide information regarding where the fetal organ is located [2]. Additionally, the number of fetal organ objects studied was limited. Therefore, this study proposed two main tasks with the instance segmentation: combined image identification and segmentation in order to determine which fetal organ objects are present in fetal images and to delineate each fetal organ object by assigning a unique label to every pixel corresponding to a specific object. Driven by the limitations of current state-of-the-art methods in this field, the contribution of this study is automatic detecting of 15-fetal organ objects that consisted of the (umbilical vein, stomach, and abdomen), brain (trans-cerebellum, trans-thalamic, and trans-ventricular), femur, head, thorax (chest cavity, heart circumference), and heart (left atrial, left ventricular, right atrial, right ventricular, and aorta). The proposed algorithm is built upon the YOLO architecture, a neural network widely utilized in medical image processing and computer vision applications. This study aims to advance maternal healthcare by enhancing the accuracy and accessibility of maternal health services.

2. MATERIALS AND METHOD

This study aims to establish an efficient classifier to detect fetal ultrasound plans. This study employs instance segmentation (segmentation and classification) for fetal organ objects. The fetal organ objects consisted of six parts, i.e., abdomen, brain, femur, head, thorax, and heart. The research methodology of this study can be visualized in Figure 1, which lists: i) data preparation of fetal object images, ii) data annotation, iii) instance segmentation with YOLO algorithms, and iv) performance evaluation.

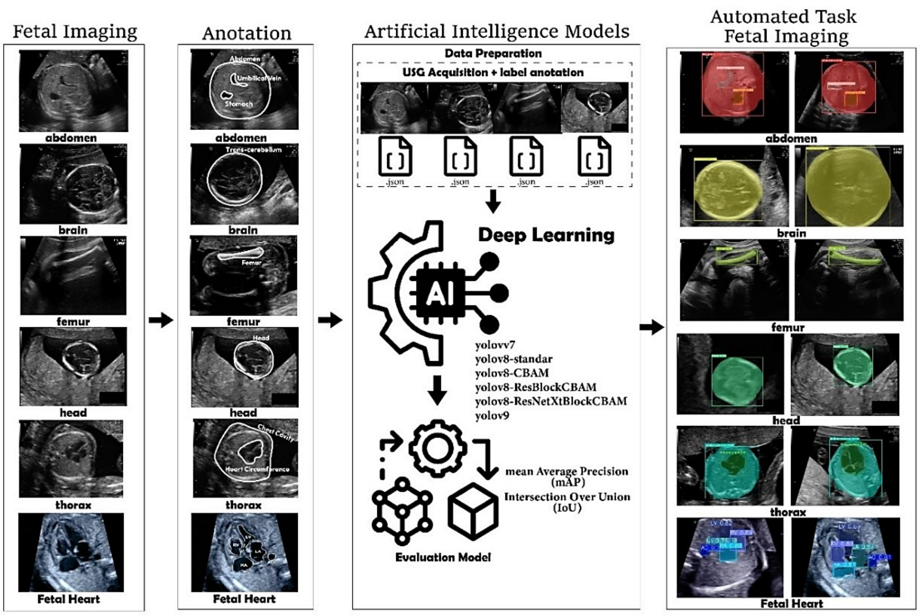


Figure 1. The stages of fetal organ segmentation

2.1. Data preparation

This study integrates two datasets, comprising both public and private sources. The combination of these datasets is intended to enhance data diversity and expand the size of the training set, thereby improving model robustness and accuracy. By learning from more diverse samples, the proposed models are better able to identify various fetal planes and contribute to advances in prenatal healthcare, particularly in fetal assessment and detection. The datasets used in this study consist of two sources.

i) The public dataset was obtained from multiple hospitals in Barcelona, Spain [17]. It comprises 12,400 two-dimensional ultrasound images collected from 896 pregnant women. These images are categorized into five fetal objects: thorax, femur, abdomen, brain, and head. All images were manually annotated by a gynecology specialist.
ii) The private dataset was collected at Dr. Mohammad Hoesin General Hospital in Indonesia and contains 800 two-dimensional ultrasound images focusing on the fetal heart.

Table 1 summarizes the distribution of six fetal objects across the two datasets and presents the number of samples allocated to the training, validation, and testing sets. To reduce bias and prevent overfitting during training, the DL models were trained using both the training and validation data. Model performance was then evaluated on the testing set, which consists of previously unseen samples. This evaluation strategy ensures that the model demonstrates good generalization capability and does not overfit the training data.

Table 1. The fetal object images distribution of training and test set

Dataset	Objects	Class	Number of data		
			Training set	validation set	Testing set
Public (Barcelona Spain Hospitals)	Abdomen	Umbilical vein (UV), stomach (STO), abdomen (ABD)	176	24	28
	Head	Head (HD)	709	150	361
	Brain	Trans-cerebellum (TC), trans-thalamic (TT), trans-ventricular (TV)	811	205	397
	Femur	Femur (FR)	260	40	44
	Thorax	Chest cavity (CC), heart circumference (HC)	223	77	87
Private (Dr. Mohammad Hoesin General Hospital, Indonesia)	Heart	Left atrial (LA), left ventricular (LV), right atrial (RA), right ventricular (RV), aorta (AO)	253	30	26

2.2. Data annotation

The annotation process generates ground truth data by converting ultrasound images into semantic segmentation masks using the Python LabelMe library (see Figure 2). In this study, Roboflow was employed to delineate polygonal regions corresponding to fetal organs, as illustrated in Figure 2, which shows fetal organ objects with standard planes of abdomen as in Figure 2(a), head as in Figure 2(b), brain as in Figure 2(c), femur as in Figure 2(d), thorax as in Figure 2(e), and heart as in Figure 2(f). All ground truth ultrasound images are provided in joint photographic experts group (JPEG/JPG) format.

2.3. Model evaluation

Instance segmentation involves predicting both the object instances and their corresponding binary segmentation masks, combining object detection, and semantic segmentation in one process [18]. This means that instance segmentation can be used simultaneously to complete counting tasks for different fetal objects. YOLO is a novel one-stage end-to-end DL model for instance segmentation [19]. YOLO is designed for fast and accurate object detection in both images and videos. It splits the image into a grid and predicts BBox and class probabilities for each grid cell, enabling the detection of multiple objects in a single pass. YOLO achieves high speed and efficiency by eliminating the need for region proposal networks [20]. This study compared YOLO architectures such as YOLOv7, YOLOv8, YOLOv9, and YOLOv11 architectures.

2.4. Attention mechanism

By integrating ResNet blocks, the model benefits from improved feature extraction capabilities due to deeper network layers [21], [22]. In addition, convolutional block attention module (CBAM) is an attention mechanism that enhances the representational power of convolutional neural networks (CNNs) by focusing on important features along both the channel and spatial dimensions. It sequentially applies channel and spatial attention to refine feature maps, enabling the network to concentrate on more informative parts of the input data [22], [23]. The overall architecture of the proposed model, incorporating residual blocks and CBAM, is illustrated in Figure 3.

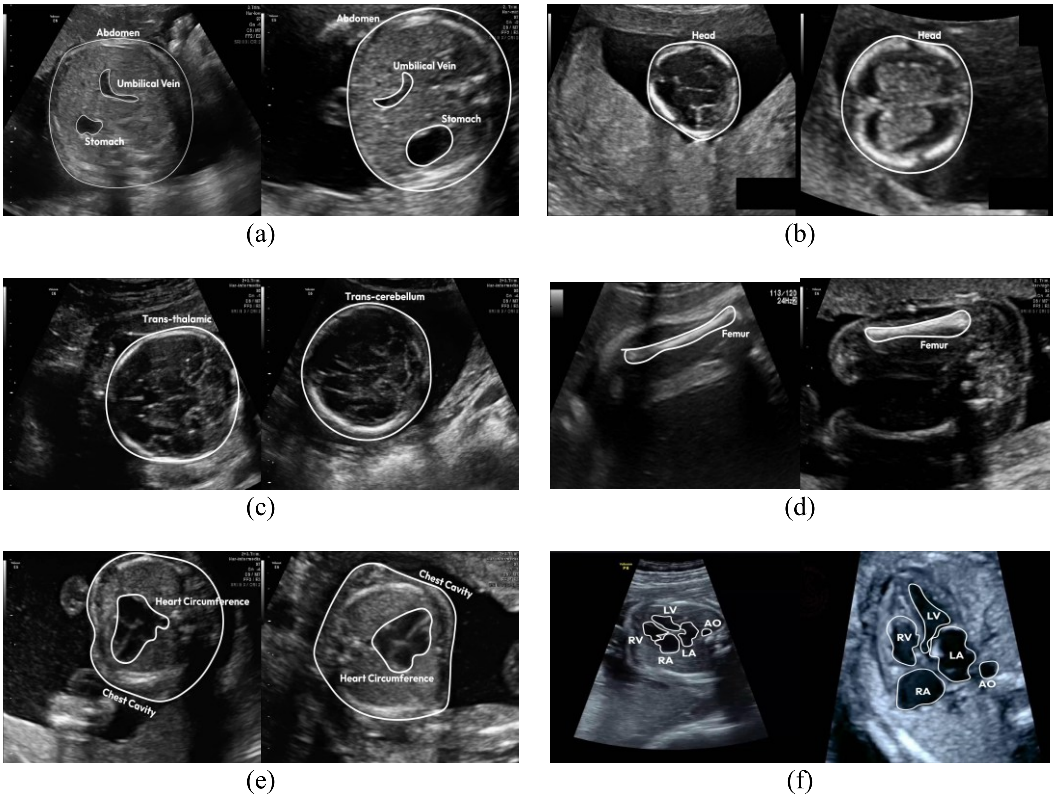


Figure 2. Fetal organ objects with standard planes of (a) abdomen, (b) head, (c) brain, (d) femur, (e) thorax, and (f) heart

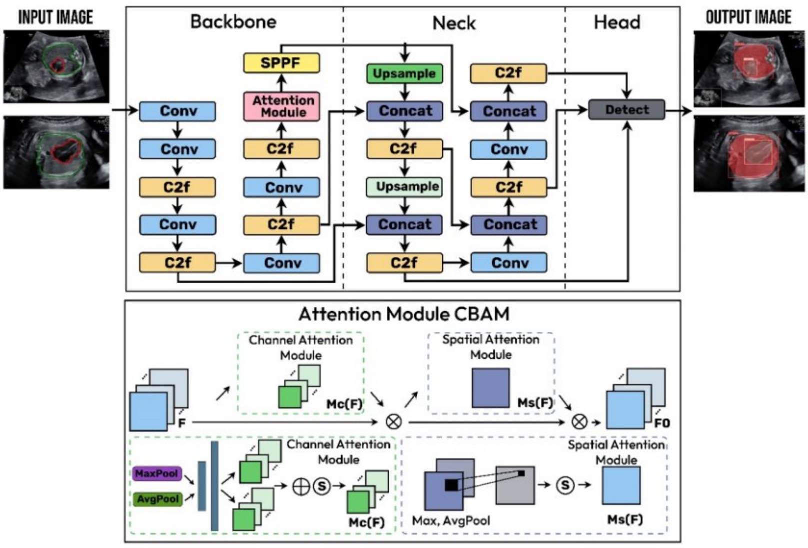


Figure 3. A new architecture of YOLO with residual block-CBAM

2.5. Metric evaluation

The evaluation metrics for an object detection model assess its ability to accurately detect and localize objects within an image. For evaluating the performance of fetal object detection, this study was used intersection over union (IoU) and mAP. IoU, also referred to as the Jaccard index, measures the accuracy by calculating the overlap between the predicted BBox and the ground truth box. IoU for comparing similarity between two arbitrary shapes $A, B \subseteq \mathbb{S} \in \mathbb{R}^n$ is attained by (1) [24].

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

Let A denote the ground truth and B represent the predicted output. By comparing the IoU value with a predefined threshold t, a detection can be classified as correct or incorrect. A detection is considered correct when IoU $\geq t$; otherwise, it is deemed incorrect when IoU $< t$.

mAP is a standard evaluation metric used to assess the performance of object detection models across all object classes. It considers both precision and recall at various IoU thresholds for each class. A higher mAP value indicates superior overall detection performance. The corresponding formulation is given as (2) [25].

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

With AP_i being the average precision in the i-th class and N is the total number of classes being evaluated.

2.6. Platform

The experiment works on Intel(R) Core(™) i9-14900K(32 CPUs), 3.2 GHz RAM, and NVIDIA GeForce RTX 2080 Ti. All experiments were run on Windows 11 Pro 64 Bit. Python code using VS Code, Pytorch, NumPy, Pandas, ScikitLearn, Matplotlib, Seaborn, and Roboflow was used.

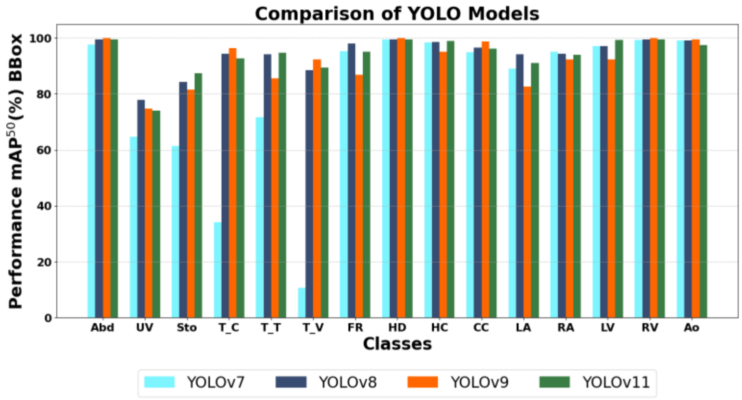
3. RESULTS AND DISCUSSION

This section presents a descriptive analysis and discusses the proposed model's results for fetal object detection. This study compared YOLOv7, YOLOv8, YOLOv9, and YOLOV11 to obtain the best model for 15-object fetal organ segmentation. Table 2 presents the results of YOLO models for fetal organ segmentation with mAP50 and mAP95. The mAP50 and mAP95 mean mAP were calculated at an IoU threshold of 0.50 and 0.95, respectively. As seen in Table 2, YOLOv9 outperformed YOLOv7, YOLOv8, and YOLOv11, with the mAP50 BBox achieved was 91.90% and mAP95 was 94.50%, while the mAP50 mask was 91.70%, and mAP90 was 94.40%. Table 2 also compares different YOLO models based on IoU metrics. The IoU of YOLOv9 is close to YOLOv8, with only o 0.03% difference. There is no significant change, and it has proven YOLOv9 still outperforms all YOLO models. Overall, YOLOv9 tends to show a stable performance with mAP50, mAP95, and IoU. YOLOv7 underperforms compared to YOLOv8 and YOLOv9, making it less optimal for fetal organ segmentation. Hence, this study proposed the YOLOv9 model for fetal organ segmentation.

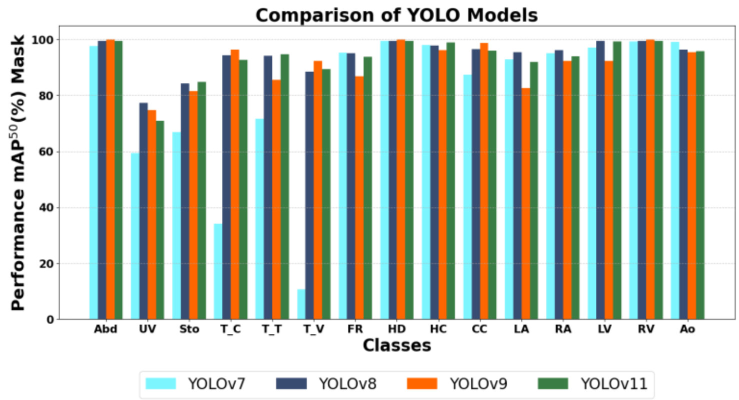
Table 2. The BBox and mask results with YOLOv7, YOLOv8, YOLOv9, and YOLOv11

Model	BBox (%)		Mask (%)		IoU
	mAP50	mAP95	mAP50	mAP95	
YOLOv7	80.51	52.26	80.27	50.00	70.03
YOLOv8	94.40	72.60	94.30	67.00	80.48
YOLOv9	91.90	94.50	91.70	94.40	80.45
YOLOv11	93.90	72.70	93.40	67.20	79.75

Figure 4 presents the comparison of the performance of three YOLO models (YOLOv7, YOLOv8, YOLOv9, and YOLOV11) across 15-object fetal organs. The y-axis represents mAP50, while the x-axis represents different object classes, labeled with abbreviations. In Figure 4(a), YOLOv9 (orange) and YOLOv11 (green) tend to have higher or similar mAP50 values compared to YOLOv7 (light blue). In some cases, YOLOv9 (orange) appears slightly ahead of YOLOv8 (dark blue) in performance. YOLOv8 and YOLOv9 both reach near 100% mAP50 in multiple classes (e.g., "Abd", "FR", "HD", "RV", "Ao"). YOLOv7 has noticeably lower performance in certain classes (e.g., "Sto", "T_C", and "T_T"). YOLOv8 and YOLOv9 perform nearly equally, with YOLOv9 having a slight edge in some cases. In Figure 4(b) with the mAP50 mask, YOLOv9 (orange) and YOLOv11 (green) perform better than YOLOv7 (light blue) in most classes. YOLOv7 shows noticeably lower performance in certain categories, especially "Sto" and "T_C". YOLOv9 (orange) and YOLOv8 (dark blue) have similar performance, but in some cases, YOLOv9 is slightly ahead. Many classes (e.g., "Abd", "FR", "HD", "RV", "Ao") reach near 100% mAP50 Mask. Based on overall Figure 4, YOLOv9 slightly leads in some cases.



(a)



(b)

Figure 4. Performance comparison of YOLO models (a) mAP50 comparison across 15 fetal organ classes and (b) mAP50 mask comparison across the same classes

Based on the comparison results above, this study has explored the YOLOV8 and YOLOv9 custom model with residual block and CBAM. Table 3 presents the highest IoU (81.01) is achieved by YOLOv8-ResNetBlockCBAM, indicating its superior ability in detecting and localizing objects accurately. YOLOv9 (91.90 mAP50) lags behind the YOLOv8 variations but excels in mAP95 (94.50), showing stronger generalization at higher precision. YOLOv9 (94.40 mAP95) is the best in high-precision segmentation. YOLOv9 performs well at high precision (mAP95), meaning it's better suited for tasks requiring highly accurate detections.

Figure 5 presents the learning curves for each model (YOLOv7, YLOv8, YOLOv8-CBAM, YOLOv8 ResNetBlockCBAM, YOLOv8 ResNetXBlockCBAM, YOLOv9, YOLOv9-CBAM, and YOLOv11). A learning curve refers to a plot that illustrates how a model's mAP changes over a range of conditions, such as the number of training epochs. This curve shows how the mAP performance of a model improves or plateaus as it learns from the data over time. As presented in Figure 5, with 100 epochs, the learning curve refers to the model performing well on both training and validation sets without underfitting or overfitting. Specifically, Figure 5(a) shows the BBox results, while Figure 5(b) presents the mask results. The mAP performance should converge and stabilize as the model learns the data well.

Table 3. The results of YOLO architectures for fetal organ segmentation

Model	BBox (%)		Mask (%)		IoU
	mAP50	mAP95	mAP50	mAP95	
YOLOv8	94.40	72.60	94.30	67.00	80.48
YOLOv8-CBAM	94.10	71.80	94.60	66.90	80.65
YOLOv8-ResNetBlockCBAM	93.70	88.88	93.80	89.00	81.01
YOLOv8-ResNetXBlockCBAM	94.20	71.90	94.20	67.10	80.85
YOLOv9	91.90	94.50	91.70	94.40	80.45
YOLOv9-CBAM	94.00	72.60	94.40	67.80	80.58

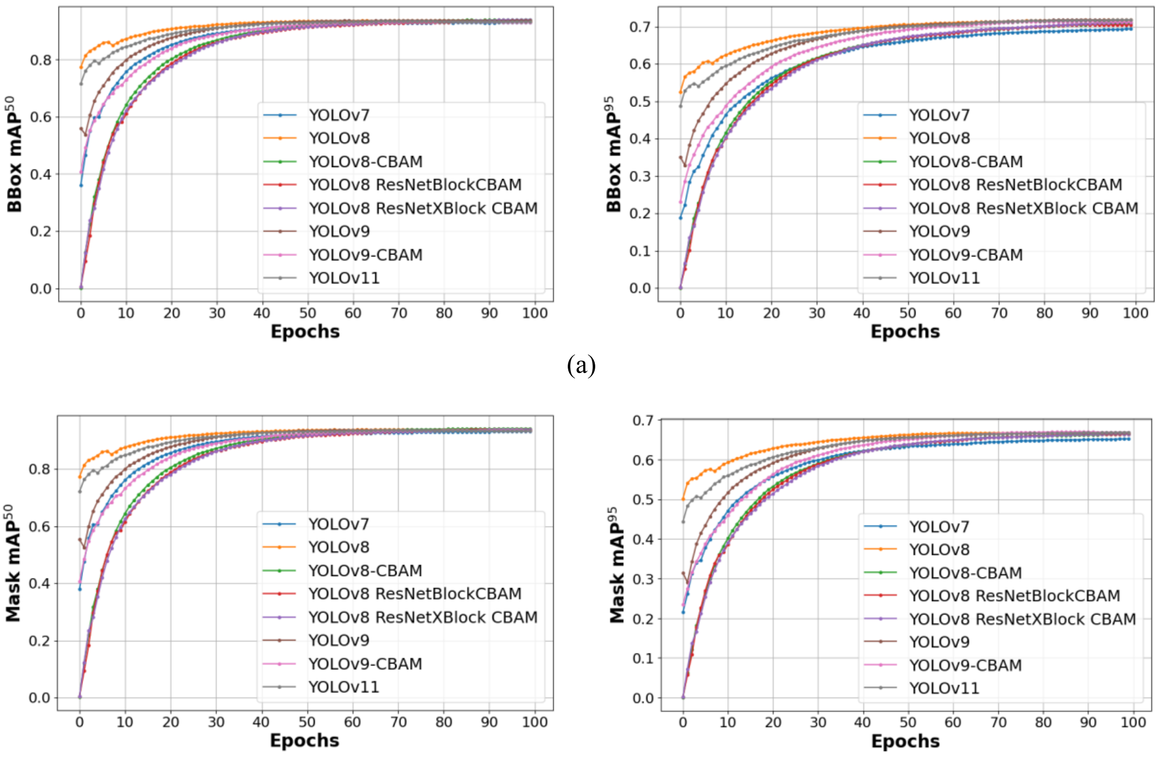


Figure 5. The learning curves of YOLO instance segmentation with mAP50 and mAP95 of (a) BBox and (b) mask

The performance of mAP50 and mAP95 for 15 fetal organ objects is illustrated in Figures 6 and 7, respectively, utilizing the proposed YOLOv9 architecture. The model demonstrates exceptional performance in detecting the abdomen, head, and right ventricular (heart) structures, achieving a perfect 100% mAP50 for these organs. This indicates that the YOLOv9 model accurately identified and localized these organs with at least 50% overlap, resulting in flawless precision and recall under this IoU threshold (see Figure 6). However, the performance in detecting the umbilical vein was significantly lower, with an mAP50 of only 74.8%. This discrepancy highlights the challenges associated with accurately detecting the umbilical vein, which may require further refinement of the model.

YOLOv9 incorporates advanced mechanisms, including programmable gradient information (PGI) and the generalized efficient layer aggregation network (GELAN). These innovations mitigate information loss in deep neural networks, thereby enhancing efficiency, accuracy, and adaptability [26], [27]. Specifically, PGI maintains critical information flow across network layers, whereas GELAN improves parameter efficiency and reduces computational complexity [27].

Figure 7 illustrates the performance of mAP95 for 15 fetal organ objects. Overall, each class achieved above 90% mAP95; however, the performance was notably affected by the detection of the umbilical vein. This is primarily because the umbilical vein is typically an isolated finding, yet it can be associated with other fetal malformations. Anomalies related to the umbilical vein may include the persistence of embryological structures, abnormal insertion points, atypical courses, and the presence of supernumerary vessels. These complexities can complicate detection and segmentation, underscoring the need for improved algorithms that can better handle such cases.

The predicted results of YOLOv9 with mAP95 in 15-fetal organ segmentation with unseen set can be presented in Figure 8. There are abdomen, head, brain, femur, thorax and heart that presented as different colors. Some masks in anatomical parts in segmentation (e.g., abdomen (Figure 8(a), head (Figure 8(b), and brain (Figure 8(c)) cover almost the entire structure, while others (e.g., femur (Figure 8(d) and thorax (Figure 8(e)) appear more refined. This suggests differences in segmentation quality across classes. The heart image in Figure 8(f) includes numerical values, likely confidence scores for different parts (RA, RV, and AO). The segmentation masks appear well-aligned with anatomical structures, indicating strong detection accuracy.

In recent years, several studies have implemented the DL for fetal organ segmentation (refer to Table 4). Dong *et al*. [28] proposed aggregated residual visual block network (ARVBNet) to detect fetal heart with 93.52% mAP and a test speed of 101 frame-per-second (FPs). Nurmaini *et al*. [29] detect the fetal heart object with 98% mAP using ResNet50. Also, they developed the instance segmentation based on a mask region-based convolutional neural network (Mask-RCNN) architecture. Komatsu *et al*. [30] proposed supervised object detection with normal data only (SONO), based on a CNN, to detect cardiac substructures and structural abnormalities in fetal. They obtained the average precision in umbilical vein and stomach, 94.4% and 96.9%, respectively. They also have detected the fetal heart, with 85.22% mAP.

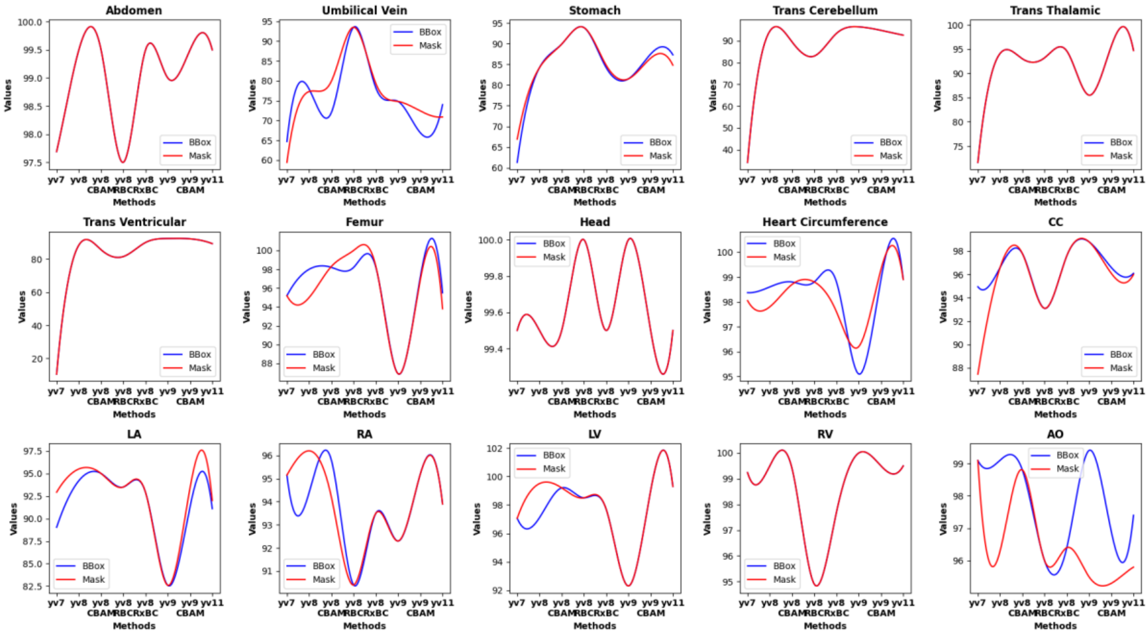


Figure 6. The performance of mAP50 in 15-fetal organ objects

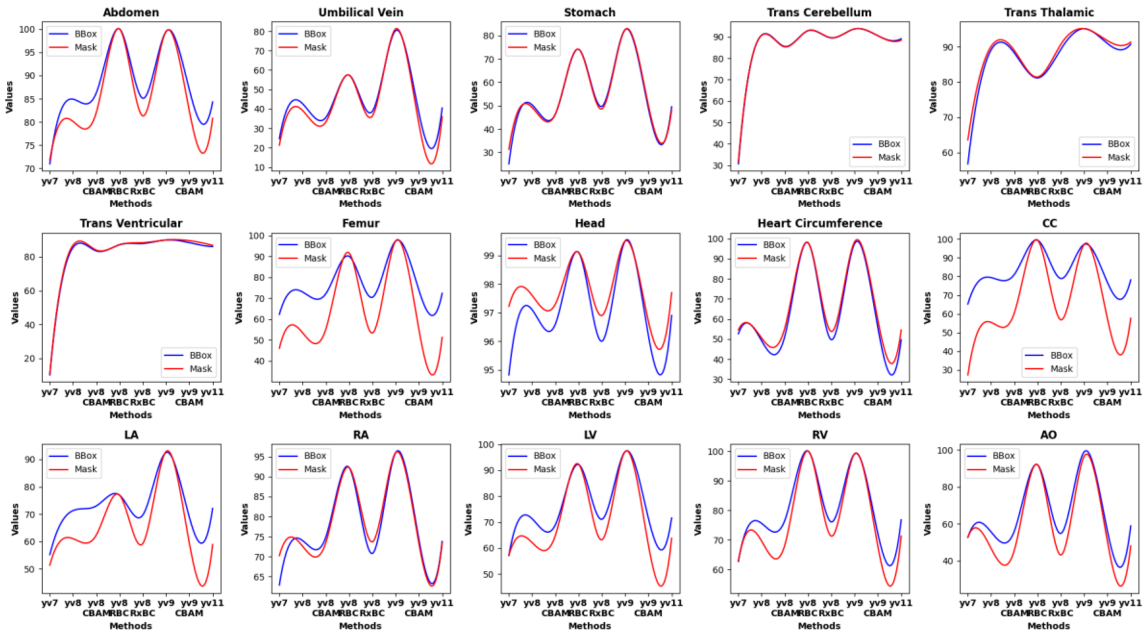


Figure 7. The performance of mAP95 in 15-fetal organ objects

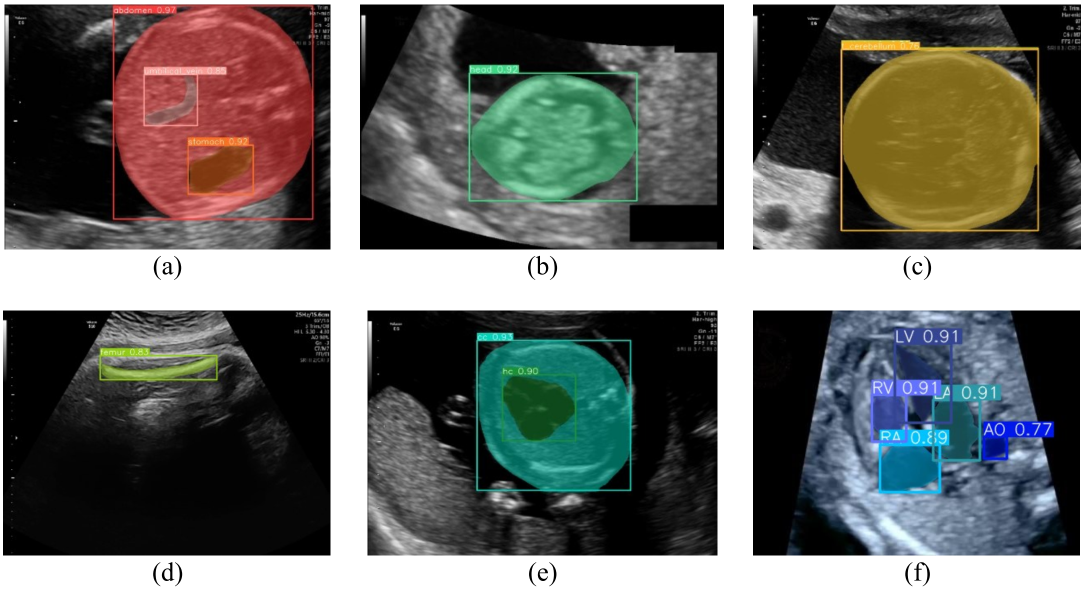


Figure 8. The predicted results of YOLOv9 with mAP95 in 15-fetal organ segmentation with unseen set of (a) abdomen, (b) head, (c) brain, (d) femur, (e) thorax, and (f) heart

Table 4 presented the proposed study outperformed the previous results [28]–[30], with the highest 94.50% mAP to detect multi fetal such as abdomen, brain, femur, head, thorax, and heart. With this study approach, the instance segmentation of YOLOv9 delivers pixel-level masks for each detected fetal organ object. It would allow for much more accurate localization of fetal organs, which is particularly important in medical imaging where exact boundaries are critical for diagnosis. Fetal organs can have irregular and complex shapes that are difficult to capture with just BBox, but the proposed YOLOv9 can better capture the fine details of each organ's structure, leading to more detailed and informative analyses. With outstanding performance, there are limitations in this study: i) limited to YOLOv7, YOLOv8, YOLOv9, and YOLOv11 architectures, with CBAM, ResNet and ResNetX backbones; and ii) this study did not handle the variations of fetal position and orientation; the fetal frequently change position or orientation during image acquisition, which poses a challenge for object detection models, potentially leading to reduced detection and segmentation accuracy.

Table 4. The benchmarking results of fetal organ object detection

Authors	Method	Fetal organ objects	mAP (%)
Dong *et al.* [28]	ARVBNet	Heart	93.52
Komatsu *et al.* [30]	SONO-CNN	Umbilical vein and stomach, heart	88.7
The proposed study	YOLOv9	Abdomen, brain, femur, head, thorax, and heart	94.50

4. CONCLUSION

The analysis of anatomical structures is a crucial aspect of fetal organ screening, as it offers clear evidence of potential fetal malformations. Fetal organs are fully developed between 18 and 22 weeks of pregnancy, such as brain, face, neck, chest/heart, abdomen, skeletal system, placenta, umbilical cord, genitalia, and cervix. This study proposed an instance segmentation based on computer-aided DL for 15-fetal organ objects, include abdomen, brain, femur, head, thorax, and heart. This study compared various YOLO architectures, including YOLOv7, YOLOv8, YOLOv9, and YOLOv11. The results showed that YOLOv9 outperformed YOLOv7, YOLOv8, and YOLOv11, achieving mAP50 and mAP95 scores of 91.90% and 94.50%, respectively. These findings indicate that YOLOv9 surpasses previous related work, which was limited to classifying fetal organs. Beyond performance improvements, the proposed approach has the potential to support clinical fetal organ screening by assisting clinicians in accurately localizing multiple anatomical structures, thereby reducing operator dependency and improving screening efficiency. Such a system may serve as a complementary tool in routine ultrasound examinations, particularly in high-workload clinical environments where rapid and consistent assessments are required. Nevertheless, this study has several limitations. The proposed model was trained and evaluated on a limited dataset, which may affect its generalizability across different ultrasound devices, imaging protocols, and patient populations. Future work will focus on expanding the dataset, improving robustness across diverse clinical scenarios, and integrating the proposed framework into real-time ultrasound systems for comprehensive clinical validation.

ACKNOWLEDGMENTS

This work was supported by the Intelligent System Research Group (ISysRG), Universitas Sriwijaya. We gratefully acknowledge Dr. Mohammad Hoesin Central General Hospital for their invaluable contributions and support throughout this project.

FUNDING INFORMATION

This research was funded by Universitas Sriwijaya under Professional Research Grant No. 0016/UN9/SK.LP2M.PT/2024, which provided the necessary resources and tools to develop and validate our DL model for the assessment of fetal heart enlargement.

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
Nuswil Bernolian	✓	✓		✓					✓					✓
Siti Nurmaini		✓							✓			✓		
Ade Iriani Sapitri			✓			✓		✓		✓				
Annisa Darmawahyuni					✓					✓				
Muhammad Naufal Rachmatullah			✓		✓					✓				
Bambang Tutuko						✓				✓				
Firdaus Firdaus						✓	✓			✓				

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [SN], upon reasonable request.

REFERENCES

[1] S. Shrestha, A. Thakur, S. Goyal, P. Garg, and N. Kler, "Growth charts in neonates," *Current Medicine Research and Practice*, vol. 6, no. 2, pp. 79–84, Mar. 2016, doi: 10.1016/j.cmrp.2016.03.009.

[2] R. R. Zegarra and T. Ghi, "Use of artificial intelligence and deep learning in fetal ultrasound imaging," *Ultrasound in Obstetrics & Gynecology*, vol. 62, no. 2, pp. 185–194, Aug. 2023, doi: 10.1002/uog.26130.

[3] P. Sridar *et al.*, "Automatic identification of multiple planes of a fetal organ from 2D ultrasound images," in *Proceedings of the MICCAI Workshop Perinatal, Preterm Paediatric Image Anal (PIPPI)*, 2016, pp. 1–10.

[4] B. Rahmatullah and J. A. Noble, "Anatomical object detection in fetal ultrasound: computer-expert agreements," in *Biomedical Informatics and Technology (ACBIT 2013)*, 2014, pp. 207–218, doi: 10.1007/978-3-642-54121-6_18.

[5] M. C.-Castrejon, T. Jansson, and T. L. Powell, "Fetal maternal unit," in *Encyclopedia of Reproduction*, 2018, pp. 477–483, doi: 10.1016/B978-0-12-801238-3.64674-3.

[6] E. Medjedovic *et al.*, "Artificial intelligence as a new answer to old challenges in maternal-fetal medicine and obstetrics," *Technology and Health Care*, vol. 32, no. 3, pp. 1273–1287, May 2024, doi: 10.3233/THC-231482.

[7] T. Vaidya, A. N. Patil, K. Mankad, A. A. Aghade, and V. R. Tambade, "The role of artificial intelligence in improving maternal fetal health outcomes: a mini review," *Archives of Medical Reports*, vol. 1, no. 3, pp. 19–22, 2024.
[8] I. Yaseen and R. Rather, "A theoretical exploration of artificial intelligence's impact on feto-maternal health from conception to delivery," *International Journal of Women's Health*, vol. 16, pp. 903–915, May 2024, doi: 10.2147/IJWH.S454127.
[9] H. N. Xie *et al.*, "Using deep-learning algorithms to classify fetal brain ultrasound images as normal or abnormal," *Ultrasound in Obstetrics & Gynecology*, vol. 56, no. 4, pp. 579–587, Oct. 2020, doi: 10.1002/uog.21967.
[10] B. Al-Bander, T. Alzahrani, S. Alzahrani, B. M. Williams, and Y. Zheng, "Improving fetal head contour detection by object localisation with deep learning," in *Medical Image Understanding and Analysis (MIUA 2019)*, 2020, pp. 142–150, doi: 10.1007/978-3-030-39343-4_12.
[11] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A review of YOLO algorithm developments," *Procedia Computer Science*, vol. 199, pp. 1066–1073, 2022, doi: 10.1016/j.procs.2022.01.135.
[12] M. Hussain, "YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection," *Machines*, vol. 11, no. 7, Jun. 2023, doi: 10.3390/machines11070677.
[13] N. S. Kumar and A. K. Goel, "Detection, localization and classification of fetal brain abnormalities using YOLO v4 architecture," *International Journal of Performability Engineering*, vol. 18, no. 10, pp. 720–729, 2022, doi: 10.23940/ijpe.22.10.p5.720-729.
[14] 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.
[15] M. Ramla, S. Sangeetha, and S. Nickolas, "Towards building an efficient deep neural network based on YOLO detector for fetal head localization from ultrasound images," in *Edge-of-Things in Personalized Healthcare Support Systems*, 2022, pp. 137–156, doi: 10.1016/B978-0-323-90585-5.00005-9.
[16] E. Dandil, M. Turkan, F. E. Urfali, İ. Biyik, and M. Korkmaz, "Fetal movement detection and anatomical plane recognition using YOLOv5 network in ultrasound scans," *European Journal of Science and Technology*, no. 26, pp. 208–216, Jun. 2021, doi: 10.31590/ejosat.951786.
[17] X. P. B.-Artizzu *et al.*, "Evaluation of deep convolutional neural networks for automatic classification of common maternal fetal ultrasound planes," *Scientific Reports*, vol. 10, no. 1, Jun. 2020, doi: 10.1038/s41598-020-67076-5.
[18] E. Mohamed, A. Shaker, A. El-Sallab, and M. Hadhoud, "INSTA-YOLO: real-time instance segmentation," Sep. 2024, arXiv: 2102.06777.
[19] D. Kolesnikov, "SAHI-like tool for instance segmentation and detection with support of YOLOv8, YOLOv9, FastSAM, and RTDETR," *GitHub*. Accessed: Sep. 13, 2023. [Online]. Available: https://github.com/orgs/ultralytics/discussions/9381
[20] P. Hurtik, V. Molek, J. Hula, M. Vajgl, P. Vlasanek, and T. Nejezchleba, "Poly-YOLO: higher speed, more precise detection and instance segmentation for YOLOv3," *Neural Computing and Applications*, vol. 34, no. 10, pp. 8275–8290, May 2022, doi: 10.1007/s00521-021-05978-9.
[21] Y. Hu, H. Tang, and G. Pan, "Spiking deep residual networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 8, pp. 5200–5205, Aug. 2023, doi: 10.1109/TNNLS.2021.3119238.
[22] Y. Luo and Z. Wang, "An improved ResNet algorithm based on CBAM," in *2021 International Conference on Computer Network, Electronic and Automation (ICCNEA)*, Sep. 2021, pp. 121–125, doi: 10.1109/ICCNEA53019.2021.00036.
[23] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: convolutional block attention module," in *Computer Vision – ECCV 2018*, 2018, pp. 3–19. doi: 10.1007/978-3-030-01234-2_1.
[24] F. van Beers, A. Lindström, E. Okafor, and M. Wiering, "Deep neural networks with intersection over union loss for binary image segmentation," in *Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods*, SCITEPRESS - Science and Technology Publications, 2019, pp. 438–445, doi: 10.5220/0007347504380445.
[25] R. Padilla, S. L. Netto, and E. A. B. da Silva, "A survey on performance metrics for object-detection algorithms," in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, Jul. 2020, pp. 237–242, doi: 10.1109/IWSSIP48289.2020.9145130.
[26] M. L. Ali and Z. Zhang, "The YOLO framework: a comprehensive review of evolution, applications, and benchmarks in object detection," *Computers*, vol. 13, no. 12, Dec. 2024, doi: 10.3390/computers13120336.
[27] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "YOLOv9: learning what you want to learn using programmable gradient information," in *Computer Vision – ECCV 2024: 18th European Conference*, 2025, pp. 1–21. doi: 10.1007/978-3-031-72751-1_1.
[28] J. Dong, S. Liu, and T. Wang, "ARVBNet: real-time detection of anatomical structures in fetal ultrasound cardiac four-chamber planes," in *Machine Learning and Medical Engineering for Cardiovascular Health and Intravascular Imaging and Computer Assisted Stenting*, Springer, Cham, 2019, pp. 130–137, doi: 10.1007/978-3-030-33327-0_16.
[29] 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, Nov. 2021, doi: 10.3390/s21238007.
[30] M. Komatsu *et al.*, "Detection of cardiac structural abnormalities in fetal ultrasound videos using deep learning," *Applied Sciences*, vol. 11, no. 1, Jan. 2021, doi: 10.3390/app11010371.

BIOGRAPHIES OF AUTHORS

Nuswil Bernolian is currently a specializing in obstetrics and gynecology. His expertise includes preconceptional care and counseling, color doppler examination for obstetric and gynecologic cases (systolic/diastolic ratio, resistance index, pulsatile index of the umbilical artery, uterine artery, middle cerebral artery, ductus venosus, and aorta), as well as the diagnosis and management of pregnancy in patients with heart disease (decompensation cordis, atrial septal defect, ventricular septal defect, myocardial infarction, mitral stenosis, and tetralogy of Fallot). He can be contacted at email: nuswilbernoli@gmail.com.

Siti Nurmaini received the master's degree in Control System from the Institut Teknologi Bandung, Indonesia, in 1998, and the Ph.D. degree in Computer Science from the Universiti Teknologi Malaysia, in 2011. She is currently a professor with the Faculty of Computer Science, Universitas Sriwijaya. She is member of IEEE. 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.

Ade Iriani Sapitri is currently a doctoral student at the Faculty of Engineering, Universitas Sriwijaya, Indonesia. In addition to her studies, she is also a research assistant affiliated with the Intelligent System Research Group. Her research interests include medical imaging, deep learning, and machine learning. She can be contacted at email: adeirianisapitri13@gmail.com.

Annisa Darmawahyuni is currently a lecturer and a researcher with the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. Her research interests include biomedical signal processing, deep learning, and machine learning. She can be contacted at email: riset.annisadarmawahyuni@gmail.com.

Muhammad Naufal Rachmatullah is currently a lecturer and a researcher with the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. His research interests include medical imaging, biomedical signal processing, deep learning, and machine learning. He can be contacted at email: naufalrachmatullah@gmail.com.

Bambang Tutuko received the master's degree in Control System from the Institut Teknologi Bandung, Indonesia, in 1998, and the doctor degree in Computer Science from the Sriwijaya University, in 2019. He is currently a professor with the Faculty of Computer Science, Universitas Sriwijaya. His research interests include biomedical engineering, deep learning, machine learning, image processing, control systems, and robotic. He can be contacted at email: bambangtutuko60@gmail.com.

Firdaus Firdaus is currently a lecturer and a researcher with the Intelligent System Research Group, Faculty of Computer Science, Universitas Sriwijaya, Indonesia. His research interests include pattern classification, text analysis, bibliographic system, cardiovascular system, classification, computer aided instruction, control engineering computing, data mining, diseases, and natural language processing. He can be contacted at email: virdauz@gmail.com.