



## Deep Learning Based Detection and Classification of Amniotic Fluid Echogenicity Type for Enhanced Prenatal Diagnosis

Putu Desiana Wulaning Ayu<sup>1\*</sup>    Gede Angga Pradipta<sup>1</sup>    I Made Darma Susila<sup>2</sup>  
 Dandy Pramana Hostiadi<sup>1</sup>    Made Liandana<sup>2</sup>

<sup>1</sup>*Department of Magister Information System, Institut Teknologi dan Bisnis STIKOM Bali, Bali, Indonesia,*

<sup>2</sup>*Department of Computer and Informatic, Institut Teknologi dan Bisnis STIKOM Bali, Bali, Indonesia*

\*Corresponding author's Email: wulaning\_ayu@stikom-bali.ac.id

**Abstract:** Meconium-stained and vernix caseosa are two significant components found in amniotic fluid, each serving distinct roles in fetal development. These components have potential to hinder fetal development when detected in excessive amounts. Due to the urgency of early detection, the development of a model for detecting and classifying echogenicity types is considered necessary. Detection model for echogenicity type based on digital images has not been developed previously. Therefore, In this study, we used a private dataset obtained from a gynecology clinic, consisting of 110 original images in JPG format. Our proposed model designed by applying a CNN based semantic segmentation approach to amniotic fluid images, with modifications made to the dense layer. Pretrained models incorporating various CNN architectures, are utilized to extract features from meconium-stain and vernix caseosa. Feature selection is carried out using three methods: Chi-Square, ANOVA, and Mutual Information. Xtreme Gradient Boosting algorithm for classification. Our proposed model achieve accuracy of 0.94 or 94% for classifying echogenicity type.

**Keywords:** Amniotic fluid, Vernix caseosa, Meconium stain, Semantic segmentation, Classification.

### 1. Introduction

All The examination of amniotic fluid focuses on two aspects: the volume of the fluid and its echogenicity. The echogenicity is caused by the presence of meconium-stain (feces) and vernix caseosa (a protective layer on the fetal skin) [1]. Meconium-stain and vernix caseosa are two significant components found in amniotic fluid, each serving distinct roles in fetal development [2]. Meconium, a sterile mixture produced in the fetal intestines, is composed of water, mucopolysaccharides, and solid materials such as vernix caseosa and lanugo [3]. Conversely, vernix caseosa, a protective substance covering the fetal, plays an important role in the skin's barrier function and may also serve as a site for dioxin excretion, highlighting its importance in the detoxification process [4]. Research indicates that vernix contains branched-chain fatty acids, which are absent in

meconium, suggesting that vernix may have a unique metabolic function. Meconium-stain (feces) is detected in the amniotic fluid, there is concern that the fetal may develop Meconium Aspiration Syndrome (MAS), a condition in which feces enter the lungs, potentially leading to fatal complications, fetal distress, and indicating underlying issues such as hypoxia or asphyxia, which can elevate the risk of perinatal morbidity and mortality [5]. The echogenicity instead of the fluid is caused by the presence of meconium (feces) and vernix caseosa. To determine the content of meconium and vernix in the amniotic fluid, clinicians perform clinical tests by examining the amniotic fluid obtained during delivery and comparing it with recorded examination results from pregnancy [6], as shown in Fig. 1. Fig. 1 (a) illustrates the presence of meconium in the amniotic fluid, while (b) shows the presence of vernix caseosa [6, 7].

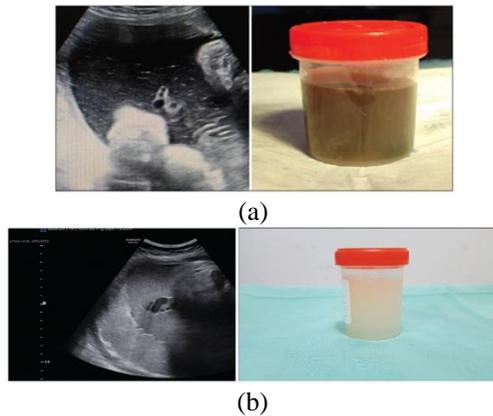


Figure. 1 Clinical test results of echogenicity: (a) meconium-stained in amniotic fluid and (b) vernix-caseosa in amniotic fluid

The early detection of meconium-stain distribution is crucial for managing fetal conditions. This enables a more prompt response from clinicians in mitigating the risk of the fetal developing Meconium Aspiration Syndrome (MAS).

Issues related to the detection and measurement of anatomical features in amniotic fluid images have been addressed with significantly satisfactory results through the use of artificial intelligence-based imaging analysis models, such as detection and segmentation. The classification of amniotic fluid volume, echogenic conditions, and the materials contributing to echogenicity is viewed as presenting promising opportunities for further research development.

Several studies have investigated the segmentation and detection of amniotic fluid areas. For instance, ayu et al., segmented the amniotic fluid area using a pixel classification model based on local grey-level rectangle window sampling, achieving an average Dice Similarity Coefficient (DSC) of 84% and an Intersection over Union (IoU) of 72.7% [8]. Another study, using the Classification Pixel Using Local Window Information and Distance Angle Pixel method, classified amniotic fluid by separating it from objects based on a defined window size and incorporating several feature extractions, such as grey-level, grey-level local variance, and distance angle pixel, achieving a DSC of 87.6% [9]. A study employing a deep learning approach, specifically the dual path network method, reported a DSC of 85.9% [10]. Another study, using the AF-net model, a variation of the U-net architecture incorporating three complementary concepts atrous convolution, a multi-scale side-input layer, and a side-output layer achieved a DSC of 87.7% [11]. Additionally, the U-net method, when optimized through hyperparameter tuning, yielded a DSC of 88% [12]. Various

approaches have been applied in classifying amniotic fluid volume. For example, a deep learning model based on the ResNet150 network combined with Extreme Gradient Boosting (XG-Boost) demonstrated a significant improvement in accuracy, reaching 96.5% [13]. The Single Deepest Vertical Detection approach achieved an accuracy of 92.63% [14]. In contrast, the Support Vector Machine (SVM) method, analysed across three different kernels, attained an accuracy of 77%. Lastly, a hybrid model combining rule-based methods with a Random Forest classifier achieved an accuracy of 90.52% [15].

Based on previous studies, the primary focus has been on the segmentation and classification of amniotic fluid volume. The potential to enhance segmentation accuracy remains a critical focus, as it significantly impacts the diagnosis of amniotic fluids. Furthermore, this study explores the underlying causes of echogenicity in amniotic fluids, specifically investigating whether it results from meconium-stained or vernix caseosa. Therefore, this study focuses on classifying the types of echogenicity in amniotic fluid into two categories: meconium stained or vernix caseosa. The novelty and contributions of this study are stated as follows:

- In our previous research [13, 16, 17, 18], the detection and classification of amniotic fluid into two categories Normal and Echogenic were the primary focus. In this study, that work is extended by specifically identifying the underlying cause of echogenicity, distinguishing between meconium-stained (fetal feces) and vernix caseosa.
- Research related to the detection or classification of meconium and vernix caseosa from an image processing perspective has not yet been conducted. Considering the importance of early detection to assess the amount of meconium in amniotic fluid, which plays a crucial role in preventing MAS, the need for research on meconium detection is highly urgent as a development area for amniotic fluid studies in the field of image processing.
- Our contribution in terms of methodology involves developing a model to classify the types of echogenicity. We propose a segmentation model that is capable of achieving higher accuracy and is robust against noise and blurred areas. The development includes applying a CNN approach for a semantic segmentation model of amniotic fluid, with modifications made to the dense layer. Pretrained models with various CNN architectures are employed to extract features from meconium and vernix caseosa. Feature

selection is carried out using three methods: Chi-Square, ANOVA, and Mutual Information. For classification, the XG-Boost algorithm is implemented. This integrated model serves as a comprehensive solution for detecting and classifying the types of echogenicity in amniotic fluid.

The remainder of the research paper is structured as follows: Section 2 presents the literature review, and Section 3 is proposed methodology. Section 4 covers the experiment result, while Section 5 provides the conclusion along with future work.

## 2. Literature review

Several studies have focused on segmenting 2D images of amniotic fluid. The accuracy and precision of amniotic fluid area segmentation are crucial factors in identifying the type of amniotic fluid. The segmented area serves as a feature that influences the volume and turbidity of the fluid. The higher the accuracy of the segmentation, the more precise the categorization of the amniotic fluid type. Several studies have been conducted to obtain the amniotic fluid area. However, challenges such as noise, blurred edges, and the presence of other particles remain. Study by [11] introduced a deep learning model known as AF-Net, A hierarchical deep-learning-based method was developed, incorporating clinicians' anatomical knowledge-based approaches. The critical step involved the segmentation of the amniotic fluid pocket using the proposed deep learning network, AF-net. AF-net, which is a variation of U-net, integrates three complementary techniques: atrous convolution, a multi-scale side-input layer, and a side-output layer. The experimental results demonstrated that this method provides a measurement of the Amniotic Fluid Index (AFI) with robustness and precision comparable to that of clinicians. The method achieved a Dice similarity score of  $0.877 \pm 0.086$  for AF segmentation.

In another study, [10] developed a deep learning network incorporating AF-Net and an auxiliary network, utilizing a dataset of 2,380 ultrasound images and achieving a DSC score of 0.8559. U-Net model by [12], segmented the amniotic fluid with U-net model, and achieved DSC 0.88. Additionally, [8] employed a pixel classification method for segmentation, using a dataset of 50 images, and comparing local window techniques, which resulted in a DSC of 0.814. Another approach was proposed by [17], combining pixel classification with feature extraction methods such as gray-level, local variance, and distance angle pixel to identify amniotic fluid regions. This technique yielded a DSC of 0.876.

However, the proposed model requires a relatively high computational time due to the involvement of multiple feature dimensions, resulting from the use of deep pixel information features.

Another study for Amniotic fluid classification by Ayu et al, aimed to attain higher identification outcomes than previous examinations by using a model comprising a convolutional neural network (CNN) (feature extractor), chi-square (feature selection), Safe Level Synthetic Minority Over-Sampling technique (SMOTE) (data oversampling), and XG-Boost (classifier), achieving accuracy 96.5% [13]. However, this study has not yet conducted an identification of the causes of turbidity in the amniotic fluid. In further research by Ayu et al., a classification of amniotic fluid conditions was performed using the Single Deep Pocket method and texture features (FOS and GLCM), classifying fluid conditions into six categories using an SVM with a Radial Basis Function (RBF) kernel, which resulted in an accuracy of 81.4% [14]. Another study, [19], classified amniotic fluid volume into two categories, use transfer learning models were utilized to classify the amniotic fluid (AF) levels as either normal or abnormal using ultrasound (US) images. The dataset, which consisted of 166 US images of pregnant women, was pre-processed before training the models. Five transfer learning models Xception, Densenet, Inception ResNet, MobileNet, and ResNet were applied. The results demonstrated that MobileNet achieved an overall accuracy of 0.94. Similar to the previous one, this study has not yet specifically categorized the causes of turbidity or the type of echogenicity in amniotic fluid. Research published by [20] employed fuzzy techniques to measure AFI, classifying conditions into oligohydramnios, borderline, normal, and polyhydramnios, achieving an accuracy of 0.925. However, this study is still focused on the categorization of amniotic fluid volume.

The study utilized a classification model that combined rule-based methods and Random Forest. The rule-based method classified amniotic fluid volume based on the Single Deepest Pocket (SDP) feature, while Random Forest was applied to classify the condition of the amniotic fluid based on First Order Statistical (FOS) and Gray Level Co-occurrence Matrix (GLCM) features, achieving an accuracy of 90.52% [15], where the classification of amniotic fluid conditions into six categories was performed. Ayu et al., [18] referred to as the modified Single Deepest Vertical Detection (SDVD) algorithm, was developed to automatically measure the longest vertical line according to medical guidelines and regulations. The SDVD algorithm was

designed to measure the depth of the amniotic fluid vertically by scanning the columns of pixels in the image sample, while excluding any intersections with the fetal body an average accuracy of 92.63% was achieved for the classification of amniotic fluid.

Based on several studies mentioned above, the accuracy of amniotic fluid area segmentation remains a major challenge that needs improvement, as it directly affects the outcome of feature extraction. Moreover, previous studies have not been able to identify the causes of echogenicity. As is known, in addition to categorizing amniotic fluid into echogenic and clear types, it is also important to identify the cause of the echogenic. Echogenic in amniotic fluid may originate from meconium-stain or vernix caseosa, both of which require different medical treatments. Therefore, this study proposes a more accurate amniotic fluid segmentation model and aims to develop a more in-depth detection model to identify the cause of echogenicity, whether it is meconium-stain or vernix caseosa.

### 3. Proposed methodology

The process of detecting and classifying meconium-stained and vernix caseosa in images of the amniotic fluid cavity is divided into four stages:

dataset acquisition, pre-processing (including data augmentation and segmentation of the amniotic fluid Region of Interest (ROI), feature extraction to obtain meconium-stain and vernix caseosa characteristics using the ResNet-18 model, feature selection using Chi-Square, ANOVA, and mutual information, and finally, classification using the XG-Boost method. The images are classified into two categories: meconium-stained and vernix caseosa. The stages of each process in this study are illustrated in Fig. 2.

#### 3.1 Data acquisition

The dataset used in this study is private dataset was obtained in collaboration with the Kasih Medika Bali Obstetrics Clinic in Bali, Indonesia. The utilize an Ultrasonography used Voluson S8 machine with a transducer frequency of 3.5 Hz and a lateral resolution of 3 mm. We have obtained ethical clearance to ensure the confidentiality of patient data. The images were saved in .jpg format with a resolution of 800 x 600 pixels.

A total of 110 images were collected, divided into two classes: 53 images in the ed-stained class and 57 images in the vernix caseosa class. An example of the original dataset is shown in Fig. 3

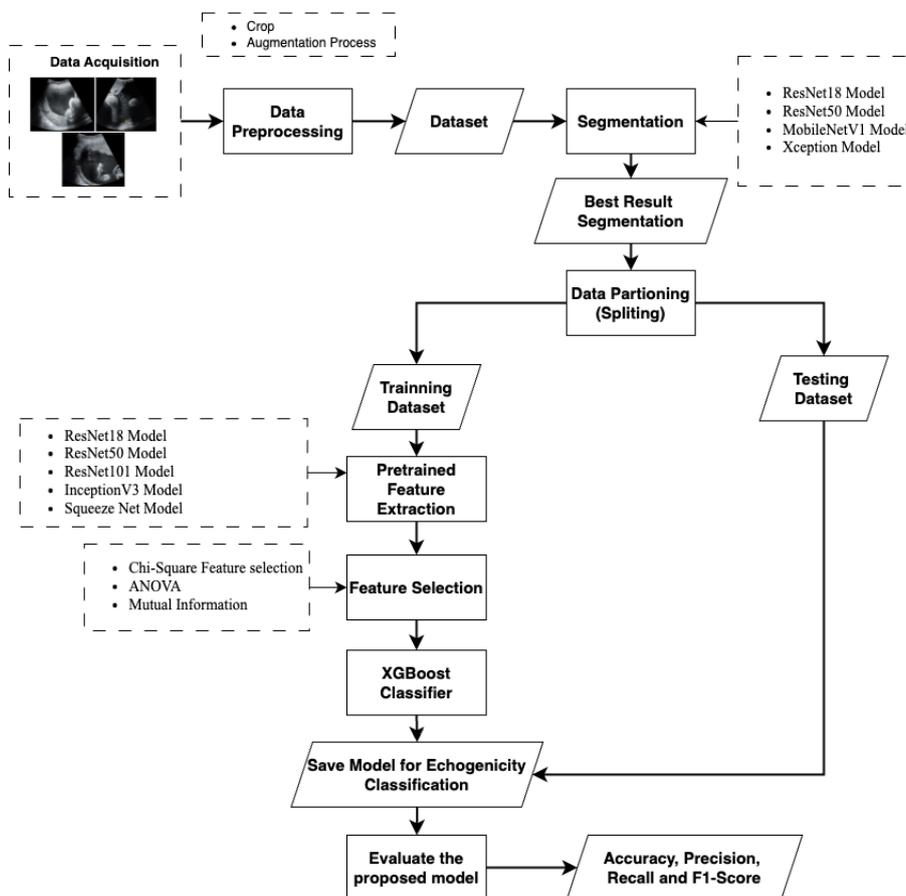


Figure. 2 The overall process of the proposed model for echogenicity classification

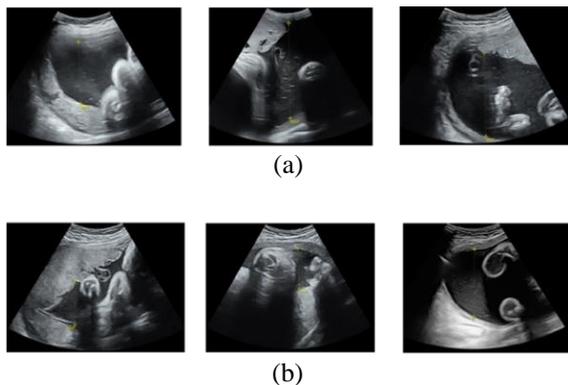


Figure. 3 Example of the dataset of amniotic fluid cavity images: (a) original dataset in the meconium-stained class and (b) original dataset in the vernix caseosa class

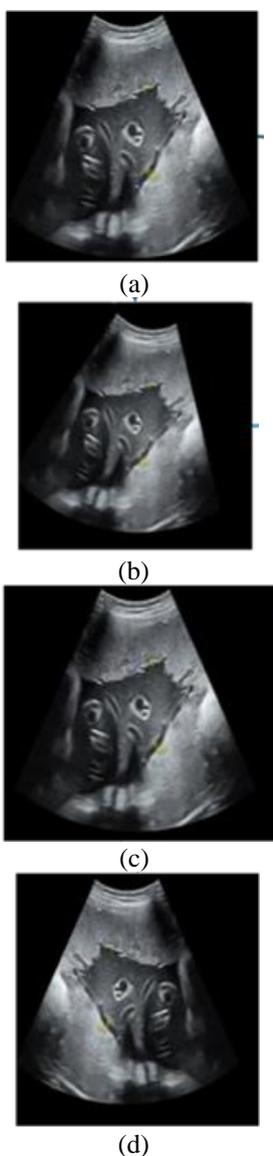


Figure. 4 Example of the augmented dataset results: (a) original image, (b) rotation, (c) translation, and (d) reflection

### 3.2 Data preprocessing

The initial step in preprocessing is automatic cropping. This cropping is performed by reducing the image size from the original 800×600 pixels to 300×400 pixels using the *imcrop* library in Matlab. The purpose of cropping is to remove unnecessary information, such as text or patient data annotations. The cropped images are then used as input for the augmentation process.

#### 3.2.1. Data augmentation

Data augmentation is a process used to increase the size of a dataset by applying random transformations to the original data [21]. Given the limited number of original datasets, the augmentation technique is one approach to expanding the dataset to prevent overfitting, Table 1 showed the number of augmented image sued in this research. Several augmentation techniques used in this study include rotation, translation, and reflection. The parameters used in data augmentation can be seen in Table 2, while examples of each augmentation process are shown in Fig. 4.

Table 1. Number of augmented dataset

Image Classes	Original Dataset		Augmentation Dataset		
		Total	Training	Testing	Total
Meconium-Stained	53	53	252	108	360
Vernix-Caseosa	57	57	278	119	397

Table 2. The parameters for data augmentation that we utilized in this research.

Parameters name for data augmentation	Parameter Value	Action
Rotation range	10 degree	The input dataset create by 10 degree slightly counter clockwise
Translation	10 pixel	The input dataset shifted to the left by 10 pixels and down by 10 pixels
Reflection	'2' flips the image horizontally (along the columns).	horizontal reflection (left-right flip) on an image.

### 3.3 Segmentation

The next step is segmentation to obtain the ROI of the amniotic fluid, a fundamental process for detecting the amniotic fluid's ROI. In several of our previous studies, segmentation methods used include pixel-wise classification and U-Net. In this study, we employed a semantic segmentation approach, comparing transfer learning models with architectures such as ResNet18, ResNet50, MobileNetV1, and Xception. ResNet introduced the concept of residual learning, which enables the training of very deep networks without encountering performance degradation issues.

The ResNet architecture is constructed using residual blocks, each consisting of several convolutional layers followed by a shortcut connection. The purpose of each residual block is to learn the difference (residual) between the input and the desired output  $(n + x)th$  [22]. In ResNet18, each residual block consists of two convolutional layers (typically with  $3 \times 3$  filters). In ResNet50, each residual block consists of three convolutional layers with  $1 \times 1$ ,  $3 \times 3$ , and  $1 \times 1$  filters, arranged in a specific sequence. In this study, we used the ResNet18 and ResNet50 architectures from [22] and modified the input image size to  $300 \times 400$  pixels and the output of the Dense layer to two categories: amniotic fluid and background.

MobileNetV1 is a type of CNN specifically designed to operate on mobile devices with low-power processors. Its architecture is efficiently designed using depth wise-separable convolutions, allowing the development of lightweight neural networks with low latency, making it ideal for mobile and embedded devices. MobileNetV1 primarily uses pointwise separable convolutions instead of full convolutions. MobileNetV1 consists of standard  $3 \times 3$  convolution layers followed by 13 blocks of  $3 \times 3$  depthwise convolutions, batch normalization, and ReLU, as well as  $1 \times 1$  pointwise convolutions with batch normalization and ReLU [23]. There are no pooling layers between these depth wise-separable convolution blocks. Stride 2 is used to reduce the spatial dimensions of the input, and the number of output channels is doubled in the pointwise convolution layers. Each of these layers uses batch normalization. MobileNetV1 uses ReLU as its activation function. The modified architecture ends with a global average pooling layer, followed by three final layers: a flatten layer, a dense layer (128), and a dense layer (2).

Xception is a variant of 'Inception' that uses 'depth wise separable convolution,' where the spatial  $n \times n$  convolution on each channel separately is

called 'depth wise convolution. Xception is a CNN architecture with 71 layers. This architecture consists of 36 convolutional layers that form the feature extraction process, grouped into 14 modules. All modules, except the first and last, have linear residual connections. Xception is a linear stack of depth wise separable convolution layers with residual connections [24]. All modified architecture for segmentation model, showed in Fig. 5.

The performance of segmentation models was evaluated using the Dice Similarity Coefficient (DSC) and Intersection over Union (IoU), as defined in Eqs. (1) and (2). The DSC is defined in Eq. (1) and ranges from 0 to 1. Additionally, the Jaccard Coefficient/IoU is calculated by dividing the number of intersecting pixels between A and B by the union of A and B, as shown in Eq. (2) [14].

$$DSC = \frac{2|A \cap B|}{|A| + |B|} \quad (1)$$

$$Jaccard/IoU = \frac{|A \cap B|}{|A \cup B|} \quad (2)$$

### 3.4 Pretrain feature extraction

Feature extraction using pre-trained models with Convolutional Neural Network (CNN) architectures involves the utilization of CNNs that have been previously trained on large datasets, such as ImageNet, to obtain features from new data. This method is considered highly effective in deep learning, particularly when applied to relatively small or domain-specific datasets.

In this approach, features are extracted from labelled images of amniotic fluid, categorized as meconium-stained and vernix caseosa, using a pre-trained model. General features such as edges and textures are captured by the early layers of the pre-trained CNN, while more complex and specific features, including shapes and objects, are captured by the subsequent layers. After feature extraction, the fully connected layers at the end of the CNN are typically discarded, and the extracted features are used as input for other machine learning models, such as XG-Boost. Five pretrained CNN models—ResNet18, ResNet50, ResNet101, Inceptionv3, and Squeeze Net—are employed in this study. Fig. 6 illustrates one of the pre-trained models utilized in this research using ResNet101. For all the pre-trained architectures utilized, feature selection processes using chi-square, ANOVA, and Mutual Information were applied to identify the best features. A fully connected layer was then used as the prediction layer.

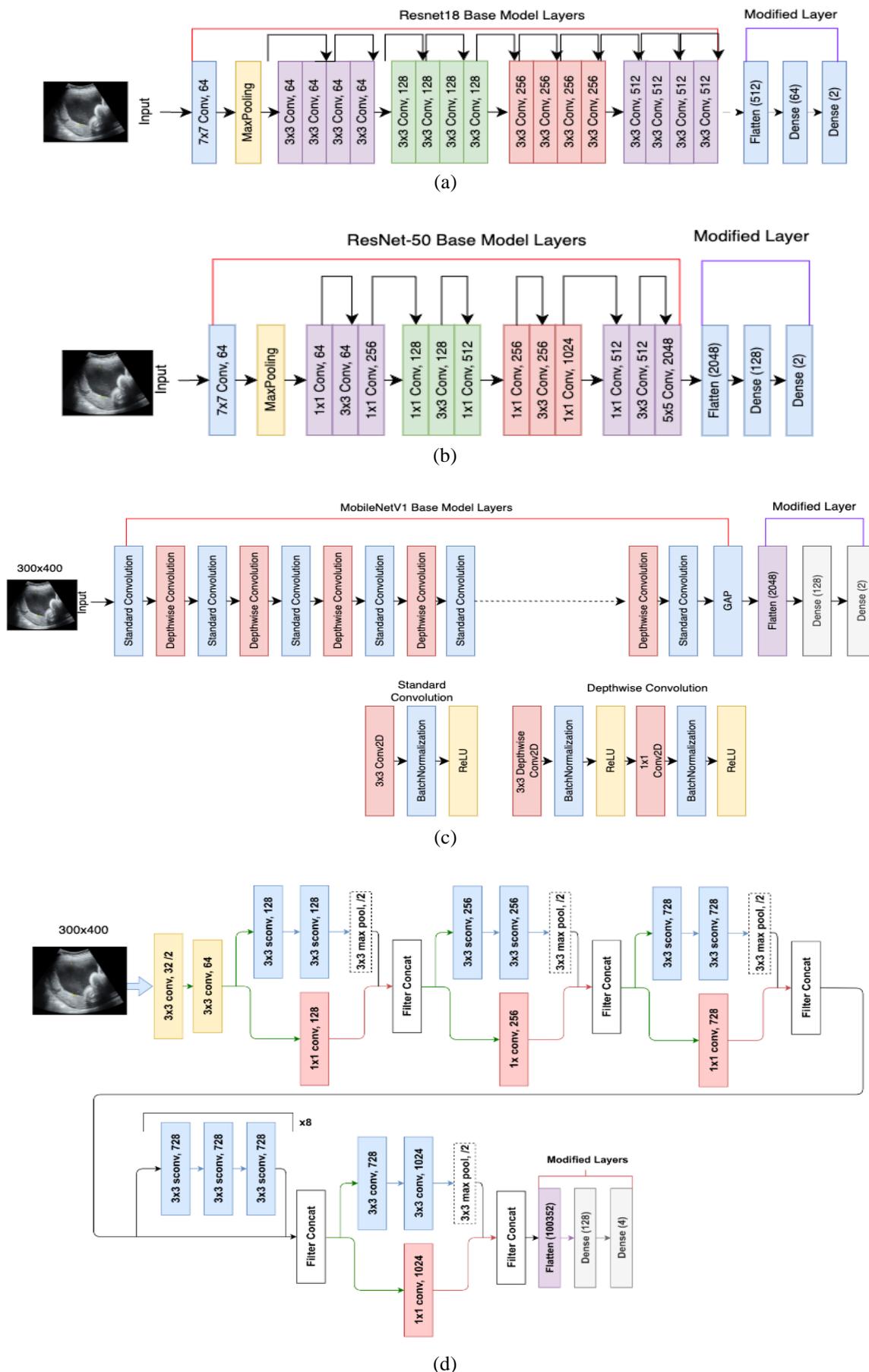


Figure. 5 The architecture for the segmentation model: (a) ResNet18, (b) ResNet50, (c) MobileNetV1, and (d) Xception  
*International Journal of Intelligent Engineering and Systems, Vol.18, No.1, 2025* DOI: 10.22266/ijies2025.0229.18

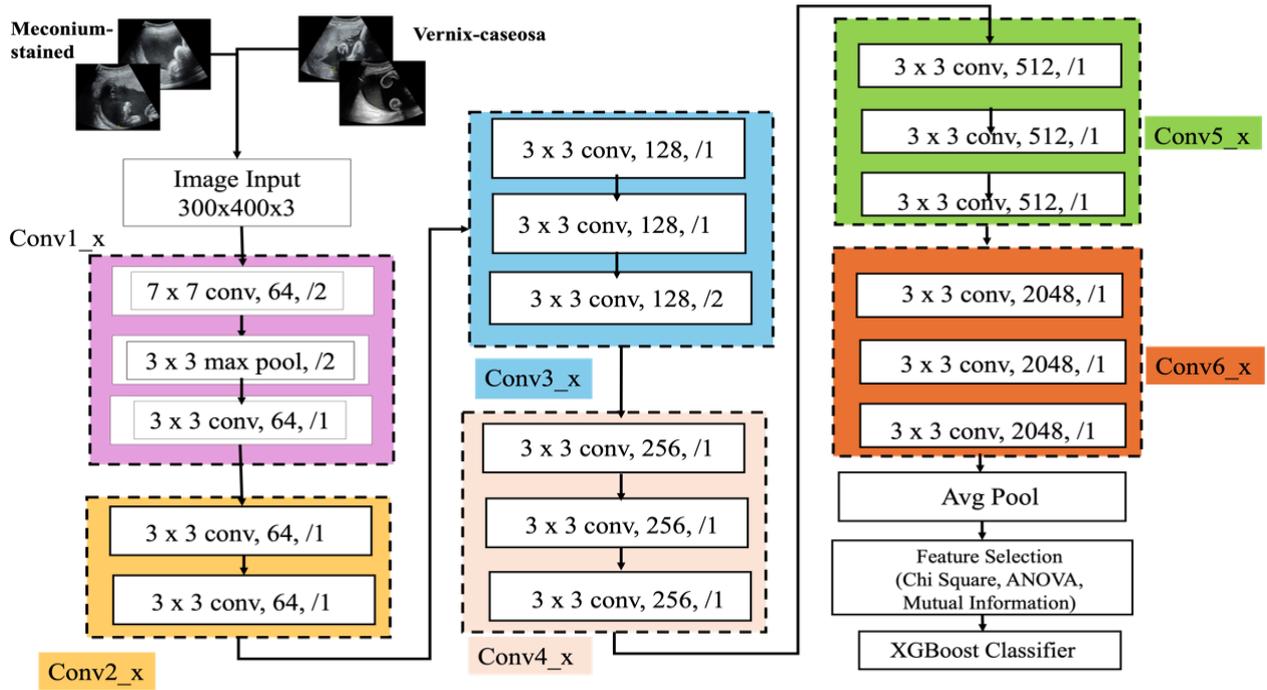


Figure. 6 The Resnet101 architecture for the pretrain

In this study, the fully connected neural network was replaced by a classifier using XG-Boost.

### 3.5 Feature selection

#### 3.5.1. Chi-Square

The Chi-square test was employed to assess the deviation from the expected distribution by evaluating independent features that were not dependent on class values. The chi-square value was determined using metrics such as true positives ( $tp$ ), dependent on class values. The chi-square value was determined using metrics such as true positives ( $tp$ ), false positives ( $fp$ ), true negatives ( $tn$ ), false negatives ( $fn$ ), the probability of positive cases ( $P_{pos}$ ), and the probability of negative cases ( $P_{neg}$ ) [13, 25] as detailed in Eq. (3)

$$\begin{aligned}
 \text{Chi-square\_metric} = & \\
 & t(t_p, (t_p + f_p)P_{pos}) + \\
 & t(f_n, (f_n + t_n)P_{pos}) + \\
 & t(f_p, (t_p + f_p)P_{neg}) + \\
 & t(f_n, (f_n + t_n)P_{neg}
 \end{aligned} \tag{3}$$

Where  $t(\text{count, expect}) = (\text{count} - \text{expect})^2 / \text{expect}$ .

#### 3.5.2. ANOVA

ANOVA compares the variability between groups (inter-group variability) with the variability

within groups (intra-group variability). This ratio is summarized in the F-statistic[26]. F-statistic: Measures how much the variability between groups compares to the variability within groups. The larger the F-statistic, the greater the likelihood that the difference in means between groups is not due to chance. The equation for the F-statistic showed in Eq. (4).

$$F = \frac{MSB}{MSW} \tag{4}$$

Where, MSB is the Mean Square Between-groups (measures the variance between the group means), and MSW is the Mean Square Within-groups (measures the variance within the groups).

#### 3.5.3. Mutual information

Mutual information is utilized as a statistical measure in feature selection to assess how much information a feature provides in predicting the target variable. Mutual Information (MI) is a measure from information theory that quantifies the amount of information one random variable provides about another. In feature selection, MI is used to evaluate the dependency between a feature ( $X$ ) and the target variable ( $Y$ ). Features that share more information with the target are more relevant. The mutual information ( $I(X; Y)$ ) between two random variables ( $X$ ) and ( $Y$ ) is defined as in Eq. (5).

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} P(x, y) \log \left( \frac{P(x, y)}{P(x)P(y)} \right) \quad (5)$$

Where:  $(P(x, y))$  is the joint probability distribution of  $(X)$  and  $(Y)$ ,  $(P(x))$  and  $(P(y))$  are the marginal probability distributions of  $(X)$  and  $(Y)$ ,  $(\log)$  is the natural logarithm. The mutual information  $(I(X; Y))$  is 0 if  $(X)$  and  $(Y)$  are independent, and it increases as their dependence grows. Mutual Information is closely related to the concept of entropy, which measures the uncertainty of a random variable. The entropy of a random variable  $(X)$  is given by Eqs. (6) and (8)[27].

$$H(X) = - \sum_{x \in X} P(x) \log P(x) \quad (6)$$

Joint Entropy of  $(X)$  and  $(Y)$  defined as :

$$H(X, Y) = - \sum_{x \in X} \sum_{y \in Y} P(x, y) \log P(x, y) \quad (7)$$

The conditional entropy of  $(Y)$  given  $(X)$  is:

$$H(Y|X) = - \sum_{x \in X} \sum_{y \in Y} P(x, y) \log P(y|x) \quad (8)$$

Mutual Information can also be expressed in terms of entropy by Eq. (9).

$$I(X; Y) = H(X) - H(X | Y) \quad (9)$$

This shows that mutual information measures the reduction in uncertainty about  $(X)$  given knowledge of  $(Y)$ . To apply mutual information for feature selection, the mutual information between each feature  $(X_i)$  and the target variable  $(Y)$  is computed in Eq. (10).

$$(I(X_i; Y)) \quad (10)$$

Steps for feature selection for general can be first compute the mutual information  $(I(X_i; Y))$  for each feature  $(X_i)$  in the dataset. Second steps are rank the features based on their mutual information scores and third step is select the top-k features with the highest mutual information.

### 3.6 Xtreme gradient boosting classifier

Boosting is known as a form of ensemble learning, where several different decision trees are typically used in the model. Errors that arise in the existing model can be corrected by adding another model. New models are continuously added until no further improvements in performance can be achieved. Gradient boosting utilizes a method called gradient descent to minimize the loss function in order to

adjust the new models to the existing ones. Additionally, XG-Boost enhances the performance of weak learners in both classification and regression problems. When the method takes previously predicted values as input, it can build a new tree to refine the predictions further. In the proposed model, the XG-Boost classifier was employed to replace the Fully Connected Layer (FCL) in five CNN pretrain architecture. Where  $\hat{y}_i$  represents the value of prediction,  $n$  is the total cases in the training sample,  $K$  is the total number of trees that need to be constructed, and  $f_k$  is a member of the group of trees known as the ensemble trees.

The final prediction is the sum of the prediction scores for each tree showed in Eqs. (11) and (19) [13, 28].

$$\hat{y}_1 = \varphi(X_i) = \sum_{k=1}^K f_k(X_i), f_k \in F, \quad (11)$$

Where the  $X_i$  are members of the training set and  $y_i$  are the corresponding class labels,  $f_k$  is the leaf score for the  $k^{th}$  tree and  $F$  is the set of all  $K$  scores for all CARTs. Regularization is applied to improve the final result:

$$L(\varphi) \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (12)$$

The first term,  $l$ , represents the differentiable loss function, which measures the difference between target  $y_i$  and the prediction  $\hat{y}_i$ . The second term avoids over-fitting:  $\Omega$  penalizes the complexity of the model:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (13)$$

Where  $\gamma, \lambda$  are constants controlling the regularization degree,  $T$  is the number of leaves in the tree and  $w$  is the weight of each leaf. Gradient boosting (GB) is effective in regression and classification problems. GB was used with the loss function, extended by a second order Taylor expansion, with the constant term removed to produce a simplified objective at step  $t$ , as follows:

$$\begin{aligned} \tilde{L}^{(t)} &= \sum_{i=1}^n \left[ g_i f_i(X_i) + \frac{1}{2} h_i f_i^2(X_i) \right] + \Omega(f_t) = \\ &= \sum_{i=1}^n \left[ g_i f_i(X_i) + \frac{1}{2} h_i f_i^2(X_i) \right] + \\ &= \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 = \\ &= \sum_{j=1}^T \left[ \left( \sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left( \sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T \end{aligned} \quad (14)$$

Table 3. Notation list

Symbol	Description
$F$	F-statistic
$I(X;Y)$	Mutual Information
$X$	Dependency between a feature
$Y$	Target variable
$H(X,Y)$	Joint Entropy of ( $X$ ) and ( $Y$ )
$H(Y X)$	The conditional entropy of ( $Y$ ) given ( $X$ )
$\hat{y}_i$	value of prediction
$n$	the total cases in the training sample
$K$	total number of trees that need to be constructed
$f_k$	member of the group of trees known as the ensemble trees
$\gamma, \lambda$	constants controlling the regularization degree
$T$	number of leaves in the tree
$w$	weight of each leaf

Where  $l_j = \{i|q(x_i) = j\}$  denote the instance set of leaf  $t$  and

$$g_i = \frac{\partial l(\hat{y}_i^{(t-1)}, y_i)}{\partial \hat{y}_i^{(t-1)}} \quad (15)$$

$$h_i = \frac{\partial^2 l(\hat{y}_i^{(t-1)}, y_i)}{\partial (\hat{y}_i^{(t-1)})^2} \quad (16)$$

Are first and second order gradient statistics of the loss function. The optimal weight  $w_j^*$  of leaf  $j$  and the quality of a tree structure  $q$ , for a given tree structure  $q(x_i)$  can be computed:

$$w_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (17)$$

$$\tilde{L}^{(t)}(q) = - \frac{1}{2} \sum_{j=1}^T \frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \quad (18)$$

In practice, the evaluating for split candidates by utilized the score in the instance sets of left  $I_L$  and right  $I_R$  nodes after the split, where  $I = I_R \cup I_L$  then the loss reduction after the split is:

$$L_{split} = \frac{1}{2} \left[ \frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} + \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_i + \lambda} \right] - \gamma \quad (19)$$

### 3.7 Evaluation and validation performance

To evaluate the performance of the classifier, this study used a binary class confusion matrix. Performance validation for classification involved four variables namely accuracy, precision, recall and F1-Score, as indicate in Eqs. (20) and (22)[19].

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (20)$$

$$Precision = \frac{TP}{TP + FP} \quad (21)$$

$$Recall = \frac{TP}{TP + FN} \quad (22)$$

## 4. Experiment result

In this research, several scenarios were used to determine the best model for classifying the types of echogenicity of amniotic fluid. The following testing scenarios were conducted.

### 4.1 Experiment on segmentation

#### 4.1.1. Optimizer performance scenario on semantic network model

In this research, experiments were conducted to evaluate the performance of each model architecture in segmenting the Amniotic Fluid ROI. In this segmentation process, 720 datasets were used, and validation was performed using 5-fold validation. The performance of each CNN architecture was compared using three optimizers: Stochastic Gradient Descent (SGD), Adam, and RMSprop. The

parameters for each optimizer were set as follows: max\_epoch = 10, max\_iteration = 600, minibatch = 16, and learning rate = 0.01. Based on the experiment, the performance accuracy comparison of ResNet18, ResNet50, MobileNetV1, and Xception of ResNet18, ResNet50, MobileNetV1, and Xception models using different optimizers over iterations in segmentation is shown in Table 4.

It is shown in Table 4 that the SGD optimizer is regarded as the overall best optimizer for models, providing consistent high accuracy with stable convergence. The highest accuracy (99.63%) is achieved by ResNet50, followed closely by ResNet18 (99.57%), MobileNetV1 (99.40%), and Xception (99.47%). Good performance is observed in MobileNetV1 and Xception, reaching 98.81% and 97.61%, respectively. However, a drop in accuracy to 90.25% is seen in ResNet18, which is unusual and suggests that overfitting or learning rate issues might

Table 4. Accuracy comparison of ResNet18, ResNet50, MobileNetV1, and Xception models using different optimizers over iterations in segmentation

Iteration	Resnet18			ResNet50			MobileNetV1			Xception		
	SGD	ADAM	RMS Prop	SGD	ADAM	RMS Prop	SGD	ADAM	RMS Prop	SGD	ADAM	RMS Prop
	<b>Accuracy</b>											
1	37.30	57.28	26.78	75.01	39.78	44.91	23.72	42.57	44.65	32.99	17.55	27.71
50	98.02	93.22	90.11	98.83	91.98	90.95	98.11	95.58	92.44	97.25	90.79	93.24
100	98.77	95.12	91.22	99.14	90.0	90.16	98.95	93.36	94.5	98.48	91.75	87.62
150	98.78	96.67	91.84	99.29	93.98	86.74	99.11	93.82	90.32	98.7	95.62	94.95
200	99.23	95.67	92.11	99.55	93.25	91.43	99.22	97.46	93.2	96.16	93.78	92.4
250	99.18	95.75	92.31	99.48	92.19	92.01	99.35	98.63	90.15	99.19	95.52	92.34
300	99.29	96.36	92.42	99.32	92.54	92.78	99.33	98.23	92.39	99.39	97.28	91.65
350	99.14	95.68	93.51	99.54	96.2	90.92	99.49	98.26	95.41	99.46	96.5	94.66
400	99.29	97.05	94.22	99.58	94.54	92.7	99.43	98.31	95.01	99.49	97.52	92.7
450	99.57	90.25	95.12	<b>99.63</b>	95.22	92.13	99.4	98.81	90.24	99.47	97.61	95.05

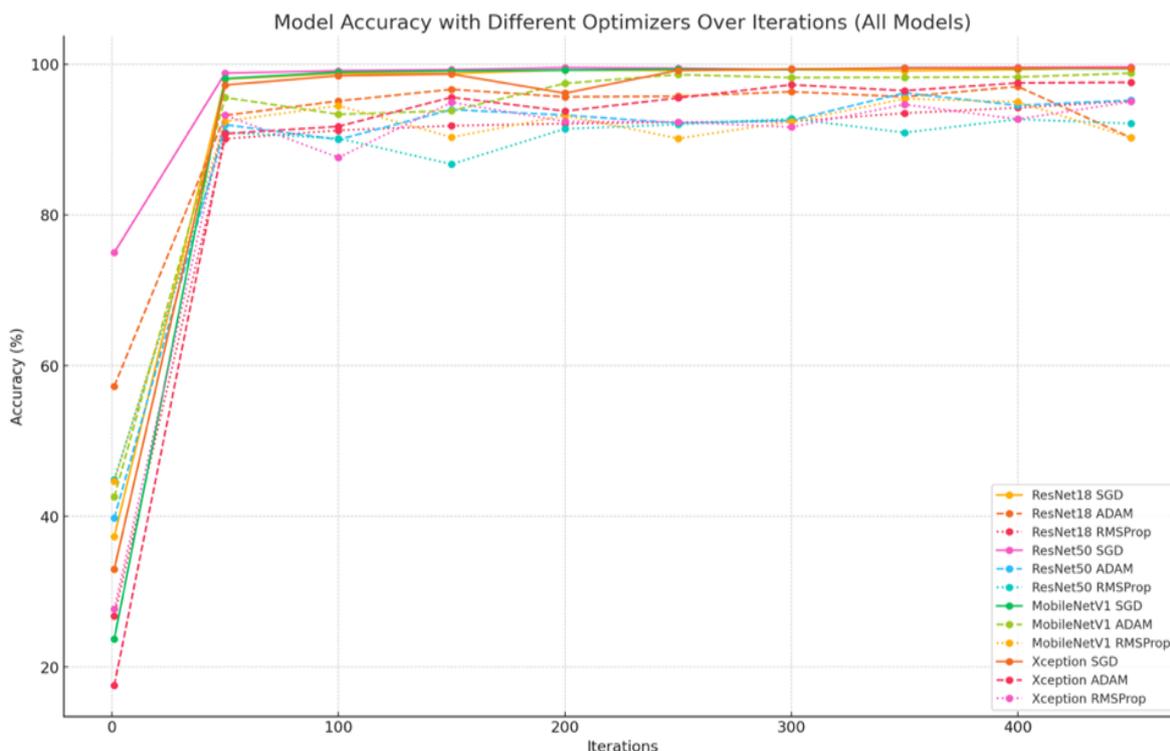


Figure. 7 Plot for accuracy comparison of ResNet18, ResNet50, MobileNetV1, and Xception models using different optimizers over iterations in segmentation

be present. In general, lower final accuracy is achieved by this optimizer compared to the other two optimizers, with significant fluctuations observed, especially in MobileNetV1 (90.24%) and ResNet50 (92.13%). Fig. 7, showed the plot for accuracy comparison of ResNet18, ResNet50, MobileNetV1, and Xception models using different optimizers over iterations in segmentation.

In Fig. 7, the analysis regarding speed and convergence of the three optimizers indicates that all models using SGD are observed to converge quickly, reaching over 98% accuracy by the 50th iteration. This implies that SGD is highly effective at identifying a good minimum early in the process, particularly for deeper models. In terms of stability and fluctuations, the SGD optimizer is shown to

maintain high and stable accuracy across all models, with minimal fluctuations after the initial rapid convergence. ADAM is observed to perform well but exhibits some instability, particularly in ResNet18 during the later iterations. Although ADAM and RMSprop can converge quickly, they may introduce variability, especially if not carefully tuned. In contrast, SGD provides more consistent stability, which is crucial for achieving the highest possible accuracy.

**4.1.2. DSC and IoU performance on semantic segmentation model**

In the second scenario, the performance of several semantic segmentation models was compared based on DSC and IoU values. The DSC and IoU performance of the CNN architecture model in semantic segmentation is shown in Table 5, and the graph is presented in Fig. 8. Segmentation results from the architecture models used are illustrated in Fig. 8. From several examples of segmentation results shown in Fig. 8, it is observed that, while all models are capable of segmenting the echogenicity regions to some extent, ResNet-50 is found to outperform the others in terms of accuracy and detail. The choice of model is determined by the trade-off between computational efficiency and the level of accuracy required in practice. Table 5 shows, for all models, SGD consistently provides the highest DSC and IoU scores, indicating that it is the most effective optimizer for semantic segmentation in these scenarios. The highest scores are achieved by ResNet50 with SGD (DSC: 0.96, and IoU: 0.92. MobileNetV1 with Adam achieves a very competitive DSC of 0.90 and IoU of 0.82, which is close to its performance with SGD.

Table 5. Performance of optimizers on CNN architecture models for the segmentation of amniotic fluid echogenicity

Semantic Segmentation Model	Optimizer					
	Adam		SGD		RMS prop	
	DSC	IoU	DSC	IoU	DSC	IoU
ResNet 18	0,73	0,60	0.95	0.90	0,29	0,22
ResNet 50	0,71	0,56	<b>0.96</b>	<b>0.92</b>	0,33	0,22
Mobile NetV1	0.90	0.82	0.94	0.89	0,53	0,38
Xception	0,69	0.82	0.93	0.89	0,51	0,37

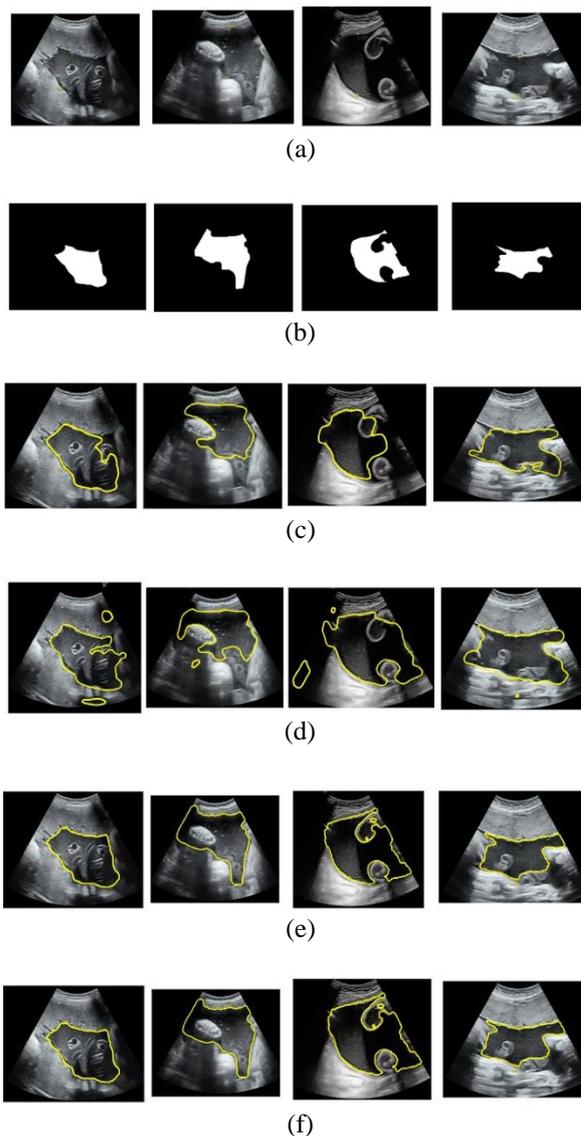


Figure. 8 Segmentation result based on CNN architecture model: (a) original dataset, (b) ground truth image, (c) segmentation result with ResNet-18, (d) segmentation result with ResNet-50, (e) segmentation result with MobilenetV1, and (f) segmentation result with Exception model

**4.1.3. Comparative analysis in segmentation model**

In this section, the efficiency of the proposed semantic segmentation model is compared with several previous studies that also focused on amniotic fluid segmentation. The segmentation model in this study applies a modified input and output approach to the ResNet18, ResNet50, MobileNetV1, and Xception models. We compared the proposed model with several previous studies that focused on amniotic fluid segmentation. Yan Li et al. [29] developed an autoencoder-decoder network model, achieving a DSC of 0.781 and an IoU of 0.544. In

addition, our previous studies have also initiated research on amniotic fluid starting in 2020. The pixel segmentation model using the Local Gray Level window [30] applies a texture-based approach limited to the window area. This model achieved a DSC score of 0.840 and an IoU of 0.727. The research continued with the proposal of the distance angle pixel [17] to classify each pixel in the image. This model achieved a DSC score of 0.876 and an IoU of 0.768. Then, in 2023, we, Ayu et al. [12], proposed using an unmodified U-Net architecture applied to amniotic fluid images. The U-Net model was able to achieve a DSC score of 0.880 and an IoU of 0.780. Finally, Csillag et al. [31] proposed AmnioML, leveraging deep learning and conformal prediction. This model achieved a DSC score of 0.910. In our study, the modified semantic segmentation model was able to reach a DSC score of 0.961 and an IoU of 0.922. This result improved by 0.5 from the highest segmentation results in previous studies. These findings indicate that the

proposed model can provide an improvement in segmentation accuracy of the amniotic fluid area. Table 6, show the comparative analysis in segmentation.

#### 4.2 Experimental on classification echogenicity type in amniotic fluids

In the first experiment, feature extraction was performed using pre-trained deep learning models. In this study, five pre-trained CNN models were used: ResNet18, ResNet50, ResNet101, Inceptionv3, and Squeeze Net. After the segmentation models were tested, the process continued with feature extraction. The best segmentation result, which was obtained using the ResNet50 model, was used as the input image for the feature extraction process. In this experiment, the training data were divided into two classes: the meconium-stained class (feces), consisting of 360 images, and the caseosa class, consisting of 379 images.

Table 6. Comparative analysis in segmentation amniotic fluid

Authors	Method	Year	Datasets	DSC	IoU
Yan Li <i>et al.</i> [29]	Encoder-Decoder Network	2017	Private Amniotic Fluids Datasets Ultrasound (Waseda University)	0,781	0.54
Ayu <i>et al.</i> [30]	Local Gray Level window	2020	Private Amniotic Fluids Datasets Ultrasound (Indonesia)	0.840	0.72
Sun S <i>et al.</i> [10]	RBV-Net	2021	Private Amniotic Fluids Dataset Ultraosund (Yonsei University)	0.859	-
Cho H <i>et al</i> [11]	AF-Net	2021	Private Amniotic Fluids Dataset Ultraosund (Yonsei University)	0.877	-
Ayu <i>et al</i> [17]	Pixel Wise Classification	2021	Private Amniotic Fluids Datasets Ultrasound (Indonesia)	0.876	0.76
Ayu <i>et al</i> [12]	U-Net (Rms Prop)	2023	Private Amniotic Fluids Datasets Ultrasound (Indonesia)	0,880	0,78
Csillag <i>et al</i> [31]	Amnio ML	2023	Private Amniotic Fluids Datasets MRI	0.910	-
<b>Proposed Model</b>	Semantic Segmentation (Modified ResNet 50)	2024	Private Amniotic Fluids Datasets Ultrasound (Indonesia)	<b>0.961</b>	<b>0.92</b>

Table 7. Performance of the classifier and feature selection on the pretrained ResNet50 model

XG-Boost Classifier	Chi-Square		ANOVA		Mutual Information		No Feature Selection	
	Feces	Vernix	Feces	Vernix	Feces	Vernix	Feces	Vernix
Accuracy	0.93	0.93	0.93	0.92	0.93	0.93	0.93	0.93
Precision	0.93	0.93	0.94	0.93	0.92	0.96	0.93	0.94
Recall	0.94	0.93	0.94	0.93	0.96	0.90	0.95	0.91
F1-score	0.93	0.92	0.94	0.93	0.94	0.93	0.94	0.92
Best Number of Feature	695		500		120			
Best accuracy	0.92		0.93		0.93		0.93	

Table 8. Performance of the classifier and feature selection on the pretrained ResNet18 model

XG-Boost Classifier	Chi-Square		ANOVA		Mutual Information		No Feature Selection	
	Feces	Vernix	Feces	Vernix	Feces	Vernix	Feces	Vernix
Accuracy	0.93	0.93	0.93	0.92	0.93	0.93	0.93	0.93
Precision	0.93	0.93	0.94	0.93	0.92	0.96	0.93	0.94
Recall	0.94	0.93	0.94	0.93	0.96	0.90	0.95	0.91
F1-score	0.93	0.92	0.94	0.93	0.94	0.93	0.94	0.92
Best Number of Feature	695		500		120			
Best accuracy	0.92		0.93		0.93		0.93	

Table 9. Performance of the classifier and feature selection on the pretrained ResNet101 model

XG-Boost Classifier	Chi-Square		ANOVA		Mutual Information		No Feature Selection	
	Feces	Vernix	Feces	Vernix	Feces	Vernix	Feces	Vernix
Accuracy	0.93	0.93	0.93	0.92	0.93	0.93	0.93	0.93
Precision	0.93	0.93	0.94	0.93	0.92	0.96	0.93	0.94
Recall	0.94	0.93	0.94	0.93	0.96	0.90	0.95	0.91
F1-score	0.93	0.92	0.94	0.93	0.94	0.93	0.94	0.92
Best Number of Feature	695		500		120			
Best accuracy	0.92		0.93		0.93		0.93	

Table 10. Performance of the classifier and feature selection on the pretrained InceptionV3 model

XG-Boost Classifier	Chi-Square		ANOVA		Mutual Information		No Feature Selection	
	Feces	Vernix	Feces	Vernix	Feces	Vernix	Feces	Vernix
Accuracy	0.92	0.92	0.93	0.92	0.93	0.93	0.91	0.91
Precision	0.92	0.93	0.92	0.93	0.95	0.94	0.90	0.93
Recall	0.94	0.90	0.94	0.90	0.93	0.90	0.95	0.86
F1-score	0.93	0.91	0.93	0.91	0.93	0.92	0.92	0.90
Best Number of Feature	650		630		725			
Best accuracy	0.92		0.92		0.93		0.91	

A total of 720 features with a dimension of 2048 were generated by each pre-trained model. In the subsequent experiment, three feature selection methods—Chi-Square, ANOVA, and Mutual Information—were tested. The purpose of this testing

was to evaluate the best performance in supporting the Extreme Boosting Classifier method for classifying the types of echogenicity in amniotic fluid. The results of the tests are shown in the Tables 7-11.

Table 11. Performance of the classifier and feature selection on the pretrained Squeeze Net model

XG-Boost Classifier	Chi-Square		ANOVA		Mutual Information		No Feature Selection	
	Feces	Vernix	Feces	Vernix	Feces	Vernix	Feces	Vernix
Accuracy	0.94	0.94	0.92	0.93	0.94	0.94	0.91	0.91
Precision	0.95	0.94	0.93	0.96	0.93	0.96	0.92	0.91
Recall	0.94	0.93	0.96	0.92	0.96	0.92	0.93	0.89
F1-score	0.92	0.94	0.95	0.94	0.95	0.94	0.92	0.90
Best Number of Feature	315		375		220			
Best accuracy	<b>0.94</b>		<b>0.92</b>		<b>0.94</b>		0.91	

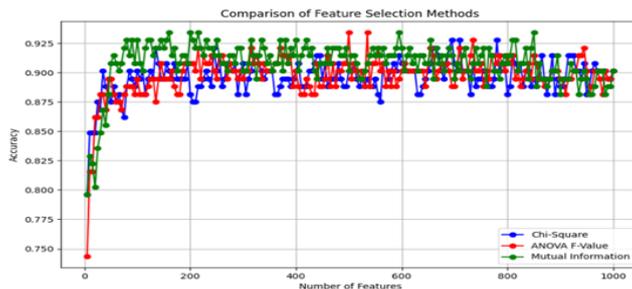
The performance metrics for the XG-Boost classifier with four different feature selection methods Chi-Square, ANOVA, Mutual Information and No Feature Selection are shown in Tables 7-11. The metrics provided include Accuracy, Precision, Recall, and F1-score for two classes, feces and Vernix, as well as the best number of features and best accuracy obtained with each method. In summary, the pretrained model with the Squeeze Net architecture in Table 11 was found to deliver the best performance compared to the other pretrained models. It is demonstrated in Table 10 that Mutual Information emerges as the most effective feature selection method, providing the highest accuracy while requiring the fewest features (220), making it both accurate and computationally efficient. Chi-Square also performs well, achieving high accuracy and balanced performance for both classes, although it requires more features (315). ANOVA shows strong precision and recall, particularly for the Vernix caseosa class, but it necessitates the largest number of features (375) to reach optimal performance. No Feature Selection significantly underperforms in comparison to the other methods, with lower accuracy and F1-scores, highlighting the importance of feature selection in improving classifier performance.

In Fig. 9, it is shown that Mutual Information is the most reliable and efficient feature selection method, particularly when fewer features are selected. It is observed to outperform the other methods early on and maintain stable performance as more features are added. Chi-Square is shown to exhibit variability, especially with a smaller number of features, but its performance improves as additional features are selected. ANOVA is consistently shown to provide strong results. In performance convergence observed across all charts that, as the number of features increases, the performance of all feature selection methods tends to converge. This indicates that with a sufficient number of features, the choice of feature selection method may become less significant.

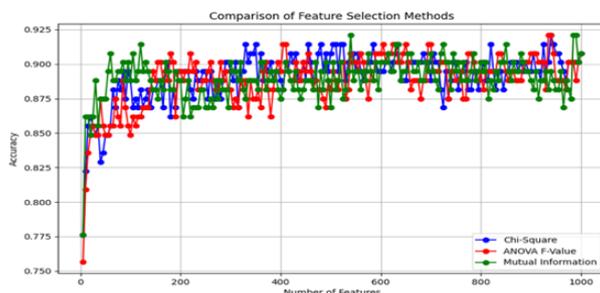
In Fig. 10 shows four Receiver Operating Characteristic (ROC) curves, where the performance of different feature selection methods applied to the Squeeze Net architecture for two classes, Meconium-stained (Feces) and Vernix caseosa, is compared. ROC curves are commonly used to illustrate the trade-off between the True Positive Rate (Sensitivity) and the False Positive Rate of a classifier at various threshold levels. The Area Under the Curve (AUC) is provided as an aggregate measure of performance across all classification thresholds. Fig. 10 (a) Chi-Square + Squeeze Net showed ROC curve for the Meconium-stained (Feces) class (Blue Line) is shown to reach an AUC of **0.96**, indicating strong performance. Similarly, the ROC curve for the Vernix caseosa class (Green Line) is observed to achieve an AUC of **0.96**, demonstrating that the classifier performs equally well for both classes when Chi-Square is used for feature selection.

Fig. 10 (b) ANOVA + Squeeze Net showed ROC curve for the Meconium-stained (Feces) class (Blue Line) is observed to achieve an AUC of **0.96**, indicating that the classifier performs well. Similarly, the ROC curve for the Vernix class (Green Line) is shown to reach an AUC of **0.96**. Fig. 10 (c) Mutual Information + Squeeze Net showed the ROC curve for the Meconium-stained (Feces) class (Blue Line) is seen to achieve an AUC of 0.96, indicating excellent performance. The ROC curve for the Vernix caseosa class (Green Line) is also observed to reach an AUC of 0.96. Fig 10 (d) No Selection + Squeeze Net showed the AUC for the Meconium-stained (Feces) class (Blue Line) is shown to drop slightly to **0.94**, indicating slightly worse performance compared to the feature selection method. The AUC for the Vernix caseosa class (Green Line) is also observed to drop to **0.94**.

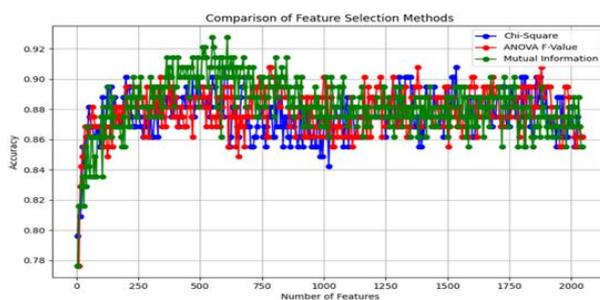
Although the curves for both classes still indicate good classification performance, it is noted that without feature selection, the classifier does not perform as well as it does when feature selection



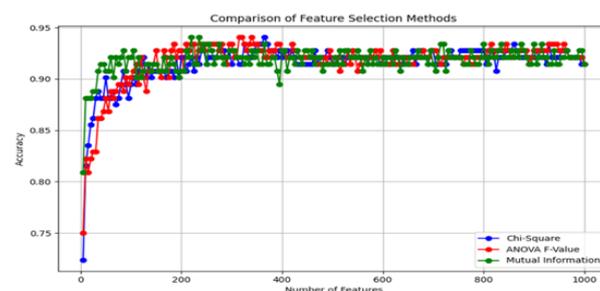
(a)



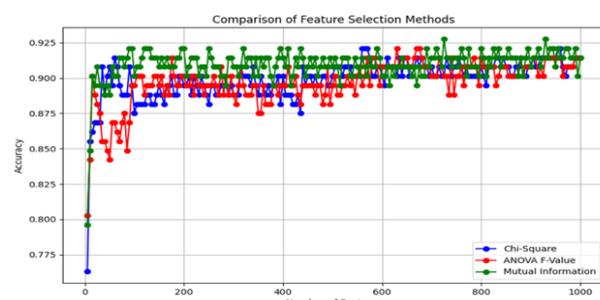
(b)



(c)

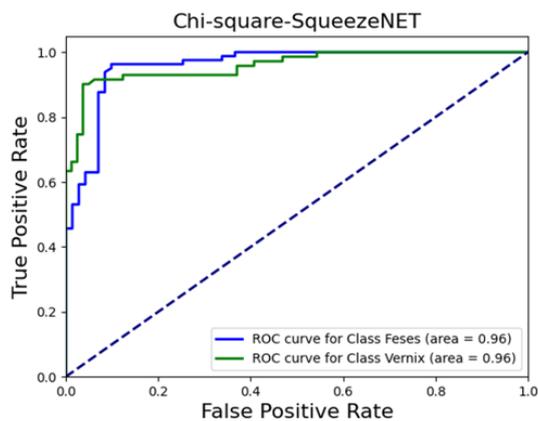


(d)

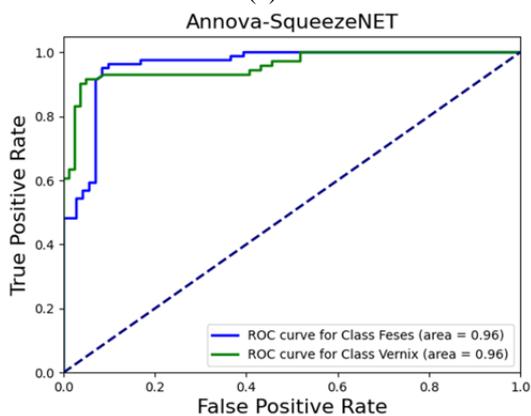


(e)

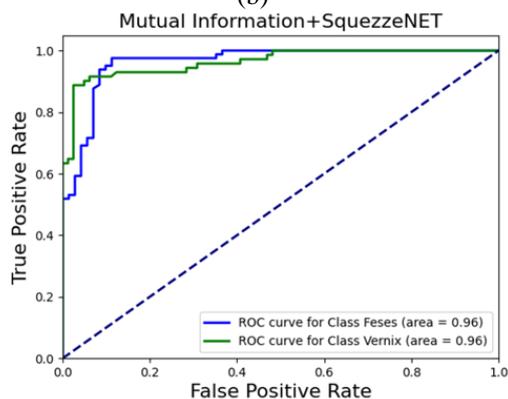
Figure. 9 Comparison of feature selection method for different pretrain CNN architecture: (a) ReNet150, (b) ResNet18, (c) ResNet101, (d) InceptionV3, and (e) Squeeze Net



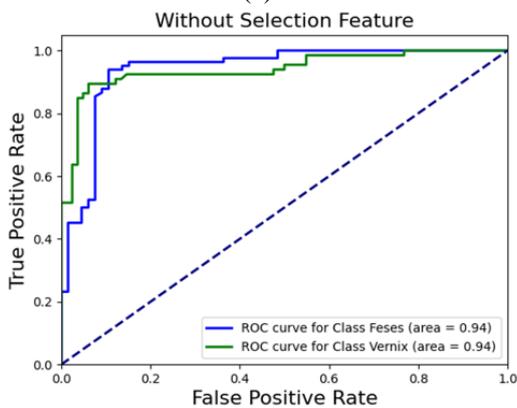
(a)



(b)



(c)



(d)

Figure. 10 ROC curve comparison of squeeze net with different selection method: (a) chi-square, (b) ANOVA, (c) mutual information, and (d) no selection feature method

methods like Chi-Square, ANOVA, or Mutual Information are applied. Strong classification results are achieved by all feature selection methods (Chi-Square, ANOVA, and Mutual Information), with an AUC of 0.96 for both classes, indicating their effectiveness in enhancing the classifier's ability to distinguish between the Meconium-stained (feces) and Vernix caseosa classes. Across all methods, both

classes show consistent performance, with no significant differences in AUC, demonstrating that the classifier handles both classes equally well. While all feature selection methods perform similarly well (AUC = 0.96), they consistently outperform the model without feature selection, making their application preferable to omitting feature selection altogether.

Table 12. Comparative analysis proposed model in this study and previous state-of-the-art works for amniotic classification

Authors	Methods	Datasets	Amniotic Fluid Class	Accuracy %	Precision %	Recall %	F1-Score %
Amuthadevi et al. 2019 [20]	Fuzzy logic	Private 2D ultrasound images	Normal, borderline, oligohydramnios and hydramnios	92.5	-	-	-
Ayu et al. 2021 [14]	Graylevel Co-occurrence matrix (GLCM), First order statistical (FOS), and Single deep pocket (SDP)	Private 2D ultrasound images	Normal echogenic, normal clear, oligohydramnios echogenic, oligohydramnios clear, and polyhydramnios clear	81.4	80.8	81.4	81.0
Ayu et al. 2021 [15]	Combination Rule based model and texture analysis	Private 2D ultrasound images	Normal echogenic, normal clear, oligohydramnios echogenic, oligohydramnios clear, and polyhydramnios clear	90.52	95.72	75.57	81.51
Khan et al 2022 [19]	Transfer learning models	US images from King Fahd Hospital of the University (KFHU) and Elite Clinic in Dammam, KSA	Abnormal AF and normal AF	94.0	96.0	94.0	95.0
Ayu et al. 2023 [18]	Single Deepest Vertical Detection (SDVD) algorithm	Private 2D ultrasound images	Normal, oligohydramnios, and polyhydramnios	<b>92.63</b>	<b>85.23</b>	<b>95.6</b>	-
<b>Proposed Method</b>	<b>Semantic segmentation and Squeeze Net as feature extractor</b>	<b>Private 2D ultrasound images</b>	<b>Amniotic Echogenicity Meconium-stained (feces) and Amniotic Echogenicity Vernix caseosa</b>	<b>94.1</b>	<b>95.1</b>	<b>94.1</b>	<b>94.1</b>

### 4.3 Comparative analysis in classification amniotic fluids

In this section, we discuss the comparison and performance evaluation of methods used to identify and classify types of amniotic fluids. Amniotic fluids are categorized into two classes (normal and abnormal), four classes (oligohydramnios, borderline, normal, and hydramnios), and five classes (normal echogenic, normal clear, oligohydramnios echogenic, oligohydramnios clear, and polyhydramnios clear) in various studies. In our current research, we focus on identifying the causes of echogenicity, which are divided into two categories: meconium-stained and caseosa. Amuthadevi et al. 2019 [14, 20] proposed contour points to find the AFI. After that, features are extracted and fuzzy logic algorithm is used to classify the given image into one of the four categories such as oligohydramnios, borderline, normal and hydramnios state. This approach achieves an average prediction accuracy of up to **92.5%**. Ayu et al [14] utilized the Gray Level Co-occurrence Matrix (GLCM), First Order Statistics (FOS), and Single Deep Pocket (SDP) as feature extraction methods. Their model was capable of attaining average values of **81.4%** for accuracy, **80.8%** for precision, **81.4%** for recall, **81%** for F-measure, and **0.88** for the Receiver Operating Characteristic (ROC) curve, respectively. Subsequently, the study was expanded by experimenting with a combination of rule-based models applied to volume and echogenicity values using texture analysis [15].

The resulting model was able to improve performance, achieving an accuracy of **90.52%**, precision of **95.72%**, recall of **75.57%**, and an F-measure of **81.51%**. Subsequently, Ayu et al. [18] further developed an algorithm to measure the depth of the amniotic sac by implementing the Single Deepest Vertical Detection (SDVD) algorithm. This approach automatically measures the longest vertical line in accordance with medical rules and regulations. The model achieved average accuracy, precision, and recall of **92.63%**, **85.23%**, and **95.6%**, respectively, for amniotic fluid classification. The approach was further developed by employing Convolutional Neural Networks (CNN) as a feature extractor [13]. Each extracted feature was then selected using the chi-square algorithm, and the learning process was conducted using the XG-Boost machine learning method. With this approach, an accuracy of approximately **96.5%** was achieved.

In this study, we employed a two-stage approach consisting of semantic segmentation for detecting areas of amniotic fluids and Squeeze Net as a feature

extractor for identifying the causes of echogenicity in amniotic fluids. The proposed model achieved an average accuracy of **94%**, precision of **95.1%**, recall of **94.1%**, and an F1-Measure of **94.1%**. Table 12 presents a comparison between the proposed model in this study and previous state-of-the-art works.

## 5. Conclusion and future work

In this study, a model was developed to detect and classify types of echogenicity present in amniotic fluid. In the initial stage, during data acquisition, three augmentation techniques rotation, translation, and reflection were applied to increase the amount of data. As a result, the total dataset used in this study consisted of 720 images, which were divided into two classes: Meconium-stain (feces) and vernix-caseosa. A segmentation model was implemented and developed to obtain the Region of Interest (ROI) from the amniotic fluid, particularly focusing on meconium-satin and vernix caseosa. A semantic segmentation model was used, utilizing five CNN model architectures: ResNet18, ResNet50, MobileNetv1, and Xception. Modifications were made to the final Dense Layer to classify the data into two classes (amniotic and background). During the segmentation stage, several optimizers, including Adam and SGD, were tested, with parameters set as RMSprop, max\_epoch = 10, max\_iteration = 600, minibatch = 16, and a learning rate of 0.01. The test results showed that the best semantic segmentation model was ResNet50, optimized using SGD, achieving an accuracy of 99.63%.

In the next testing scenario, the performance of the semantic segmentation models was compared in terms of DSC and IoU metrics. It was found that the ResNet50 model with the SGD optimizer achieved a DSC score of 0.96 and an IoU score of 0.92. The next step involved classifying the types of echogenicity in the amniotic fluid. The first step taken was feature extraction using a pre-trained CNN model. Five architectures were used: ResNet18, ResNet50, ResNet101, Inceptionv3, and Squeeze Net. Pre-trained feature extraction resulted in features with dimensions of 720x2048. In the second stage, feature selection was performed by applying three models: Chi-Square, ANOVA, and Mutual Information, with the aim of determining which combination of pre-trained models and feature selection methods produced the best performance when used with the Extreme Boosting Classifier to classify echogenicity types in the amniotic fluid. It was found that the combination of the Squeeze Net pre-trained model and Mutual Information achieved the best accuracy score, both in terms of class accuracy (feces and

vernix) and global accuracy. The combination of Squeeze Net and Mutual Information resulted in a global accuracy of 0.94, with an ROC score of 0.96. Additionally, a comparison was made between the performance with and without feature selection. It was found that without feature selection, the best accuracy achieved was 0.93, obtained using the ResNet101 pre-trained model. Thus, it was concluded that the combination of augmentation models, semantic segmentation, pre-trained models, feature selection, and the XG-Boost classifier achieved an accuracy of 0.94 or 94% in classifying echogenicity types in amniotic fluid. Based on the results of the model developed in this study, future research will focus on exploring feature fusion and developing a more advanced classifier model.

### Conflicts of Interest

The author declares no conflict of interest.

### Author Contributions

Conceptualization, Putu Desiana Wulaning Ayu; Methodology, Putu Desiana Wulaning Ayu and Gede Angga Pradipta; Software, Putu Desiana Wulaning Ayu and Made Liandana; Validation, Putu Desiana Wulaning Ayu, Gede Angga Pradipta; Formal analysis, Putu Desiana Wulaning Ayu and Gede Angga Pradipta; Data Curation, Putu Desiana Wulaning Ayu; Writing-Original Draft Preparation, Putu Desiana Wulaning Ayu, Gede Angga Pradipta and I Made Darma Susila; Writing-review and editing, Putu Desiana Wulaning Ayu; Supervision, Putu Desiana Wulaning Ayu; Visualization, Darma Susila, Dandy Pramana H, Made Liandana; Project Administrator, Made Darma Susila, Dandy Pramana; Funding acquisition, Putu Desiana Wulaning Ayu.

### Acknowledgments

The authors are grateful to the Directorate Jenderal Akademik of Research, Technology, and Community Service (DPRM) KEMENDIKBUDRISTEK Indonesia for funding this study through Program Funding (Fundamental Reguler) for the 2024 Fiscal Year.

### References

- [1] D. M. Gallo, R. Romero and M. Bosco, "Meconium-stained amniotic fluid", *American Journal of Obstetrics & Gynecology*, Vol. 228, No. 5, 2023, doi: 10.1016/j.ajog.2022.11.1283.
- [2] M. F. Meda, E. M. Elhelw, and H. E. E. Hamed, "Significance of echogenic amniotic fluid at term pregnancy and its association with meconium", *Al-Azhar International Medical Journal*, Vol. 4, No. 3, 2023, doi: 10.58675/2682-339x.1716.
- [3] S. Argyridis and S. Arulkumaran, "Meconium stained amniotic fluid", *Obstetrics, Gynaecology and Reproductive Medicine*, Vol. 26, No. 8, 2016, doi: 10.1016/j.ogrm.2016.05.001.
- [4] G. N. Buyuk, Z. A. Oskovi-Kaplan, S. Kahyaoglu, and Y. Engin-Ustun, "Echogenic particles in the amniotic fluid of term low-risk pregnant women: does it have a clinical significance?", *J Obstet Gynaecol (Lahore)*, Vol. 41, No. 7, pp. 1048-1052, 2021, doi: 10.1080/01443615.2020.1834520.
- [5] L. Avagliano, V. Massa, and G. Bulfamante, "Meconium-stained amniotic fluid and histologic signs of fetal distress in stillbirths", *European Journal of Obstetrics and Gynecology and Reproductive Biology*, Vol. 266, pp. 55-62, 2021, doi: 10.1016/j.ejogrb.2021.09.016.
- [6] Z. M. J. Al-Sattam, "Role of Ultrasound in Antenatal Detection of Echogenic Amniotic Fluid and Pregnancy Outcome", *Al-Rafidain Journal of Medical Sciences*, Vol. 6, No. 1, pp. 112-116, 2024, doi: 10.54133/ajms.v6i1.500.
- [7] A. Kaluarachchi, G. R. M. UdaraGanthika PeirisJayawardena, A. P. Ranaweera, and M. M. Rishard, "Hyperechoic amniotic fluid in a term pregnancy", *J Family Med Prim Care*, Vol. 7, No. 3, p. 635, 2018, doi: 10.4103/jfmpc.jfmpc\_83\_18.
- [8] D. W. Ayu, S. Hartati, and A. Musdholifah, "Amniotic Fluid Segmentation by Pixel Classification in B-Mode Ultrasound Image for Computer Assisted Diagnosis", In: *Proc. of Soft Comput. Data Sci*, p. 59-70, 2019, doi: 10.1007/978-981-15-0399-3\_5.
- [9] P. D. W. Ayu and S. Hartati, "Pixel Classification Based on Local Gray Level Rectangle Window Sampling for Amniotic Fluid Segmentation", *International Journal of Intelligent Engineering and Systems*, Vol. 14, No. 1, pp. 420-432, 2021, doi: 10.22266/IJIES2021.0228.39.
- [10] S. Sun, J. Y. Kwon, Y. Park, H. C. Cho, C. M. Hyun, and J. K. Seo, "Complementary Network for Accurate Amniotic Fluid Segmentation from Ultrasound Images", *IEEE Access*, Vol. 9, pp. 108223-108235, 2021, doi: 10.1109/ACCESS.2021.3098844.
- [11] H. C. Cho, Sun S "Automated ultrasound assessment of amniotic fluid index using deep

- learning”, *Med Image Anal*, Vol. 69, p. 101951, 2021, doi: 10.1016/j.media.2020.101951.
- [12] P. D. Wulaning Ayu and G. A. Pradipta, “U-Net Tuning Hyperparameter for Segmentation in Amniotic Fluid Ultrasonography Image”, In: *Proc. of 2022 4th International Conference on Cybernetics and Intelligent System, ICORIS 2022*, 2022, doi: 10.1109/ICORIS56080.2022.10031294.
- [13] P. D. W. Ayu, G. A. Pradipta, R. R. Huizen, E. S. W. Kadek, and I. G. E. Artana, “Combining CNN Feature Extractors and Oversampling Safe Level SMOTE to Enhance Amniotic Fluid Ultrasound Image Classification”, *International Journal of Intelligent Engineering and Systems*, Vol. 17, No. 1, pp. 251-262, 2024, doi: 10.22266/ijies2024.0229.24.
- [14] P. D. W. Ayu, S. Hartati, A. Musdholifah, and D. S. Nurdiati, “Amniotic fluid classification based on volume and echogenicity using single deep pocket and texture feature”, *ICIC Express Letters*, Vol. 15, No. 7, pp. 681-691, 2021, doi: 10.24507/icicel.15.07.681.
- [15] D. S. N. Ayu, P D W, Sri hartati, Aina Musdholifah, “Amniotic Fluids Classification Using Combination of Rules-Based and Random Forest Algorithm”, *Soft Computing in Data Science*, p. 15, 2022, doi: 10.1007/978-981-16-7334-4\_20.
- [16] P. D. W. Ayu and S. Hartati, “Pixel Classification Based on Local Gray Level Rectangle Window Sampling for Amniotic Fluid Segmentation”, *International Journal of Intelligent Engineering and Systems*, Vol. 14, No. 1, pp. 420-432, 2021, doi: 10.22266/IJIES2021.0228.39.
- [17] P. D. W. Ayu, S. Hartati, A. Musdholifah, and D. S. Nurdiati, “Amniotic fluid segmentation based on pixel classification using local window information and distance angle pixel”, in *Applied Soft Computing*, 2021, doi: 10.1016/j.asoc.2021.107196.
- [18] P. D. Desiana Wulaning Ayu, G. Angga Pradipta, R. Rudolf Huizen, K. W. Eka Sapta, and I. Gede Edy Artana, “Modified of Single Deepest Vertical Detection (SDVD) Algorithm for Amniotic Fluid Volume Classification”, *Jurnal Informatika*, Vol 11, No 2, 2023, doi: 10.30595/juita.v11i2.18435.
- [19] I. U. Khan and Aslam N “Deep Learning-Based Computer-Aided Classification of Amniotic Fluid Using Ultrasound Images from Saudi Arabia”, *Big Data and Cognitive Computing*, Vol. 6, No. 4, 2022, doi: 10.3390/bdcc6040107.
- [20] C. Amuthadevi and G. M. Subarnan, “Development of fuzzy approach to predict the fetus safety and growth using AFI”, *Journal of Supercomputing*, Vol. 76, No. 8, pp. 5981-5995, 2020, doi: 10.1007/s11227-019-03099-8.
- [21] C. Shorten and T. M. Khoshgoftaar, “A survey on Image Data Augmentation for Deep Learning”, *J Big Data*, Vol. 6, No. 1, 2019, doi: 10.1186/s40537-019-0197-0.
- [22] K. He, “Deep Residual Learning for Image Recognition”, In: *Proc. of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 770-778, 2016, doi: 10.1109/CVPR.2016.90.
- [23] A. G. Howard, Zhu M, Adam H “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications”, *arXiv:1704.04861*, 2017, doi: 10.48550/arXiv.1704.04861.
- [24] F. Chollet, “Xception: Deep learning with depthwise separable convolutions”, In: *Proc. of 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, Vol. 2017-Janua, pp. 1800-1807, 2017, doi: 10.1109/CVPR.2017.195.
- [25] I. Sumaiya Thaseen and C. Aswani Kumar, “Intrusion detection model using fusion of chi-square feature selection and multi class SVM”, *Journal of King Saud University - Computer and Information Sciences*, Vol. 29, No. 4, pp. 462-472, 2017, doi: 10.1016/j.jksuci.2015.12.004.
- [26] N. Sevani, K. Azizah, and W. Jatmiko, “A Feature-based Transfer Learning to Improve the Image Classification with Support Vector Machine”, *International Journal of Advanced Computer Science and Applications*, Vol. 14, No. 6, pp. 291-301, 2023, doi: 10.14569/IJACSA.2023.0140632.
- [27] S. Liu and M. Motani, “Improving Mutual Information based Feature Selection by Boosting Unique Relevance”, *arXiv:2212.06143*, pp. 1-13, 2022, doi: <http://arxiv.org/abs/2212.06143>.
- [28] S. Thongsuwan, S. Jaiyen, A. Padcharoen, and P. Agarwal, “ConvXGB: A new deep learning model for classification problems based on CNN and XGBoost”, *Nuclear Engineering and Technology*, Vol. 53, No. 2, pp. 522-531, 2021, doi: 10.1016/j.net.2020.04.008.
- [29] Y. Li, R. Xu, J. Ohya, and H. Iwata, “Automatic fetal body and amniotic fluid segmentation from fetal ultrasound images by encoder-decoder network with inner layers”, In: *Proc. of IEEE Engineering in Medicine and Biology*

- Society*, pp. 1485-1488, 2017, doi: 10.1109/EMBC.2017.8037116.
- [30] P. D. W. Ayu and S. Hartati, "Pixel Classification Based on Local Gray Level Rectangle Window Sampling for Amniotic Fluid Segmentation", *International Journal of Intelligent Engineering and Systems*, Vol. 14, No. 1, pp. 420-432, 2021, doi: 10.22266/IJIES2021.0228.39.
- [31] D. Csillag, Paes L and Orenstein P "AmnioML: Amniotic Fluid Segmentation and Volume Prediction with Uncertainty Quantification", In: *Proc. of the 37th AAAI Conference on Artificial Intelligence, AAAI 2023*, Vol. 37, pp. 15494-15502, 2023, doi: 10.1609/aaai.v37i13.26837.