Received: August 27, 2023. Revised: October 30, 2023.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Combining CNN Feature Extractors and Oversampling Safe Level SMOTE to Enhance Amniotic Fluid Ultrasound Image Classification

Putu Desiana Wulaning Ayu[1]* Gede Angga Pradipta[1] Roy Rudolf Huizen[1]
Kadek Eka Sapta W[2] I Gede Edy Artana[2]

[1]Department of Magister Information System,
Institut Teknologi dan Bisnis STIKOM Bali, Raya Puputan No.86, Denpasar 80234, Bali, Indonesia
[2]Magister Program, Department of Magister Information System,
Institut Teknologi dan Bisnis STIKOM Bali, Raya Puputan No.86, Denpasar 80234, Bali, Indonesia
* Corresponding author's Email: wulaning_ayu@stikom-bali.ac.id

Abstract: Amniotic fluid is known to play a crucial role in nurturing fetal growth and development during pregnancy by protecting the fetus from shocks, impacts, and pressure on the abdomen of the mother. One of the parameters often examined by obstetricians regarding amniotic fluid is the volume, which should match the pregnancy stage. Investigations focused on identifying amniotic fluid volume has continued to evolve. Previous examinations used image processing techniques for the single deep pocket method implementation but the results achieved were deemed improvable. Therefore, this study 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 XGBoost (classifier). The proposed model was comprehensively tested with an analysis of feature selection and oversampling effects. Based on a 5-fold cross-validation examination, the proposed model demonstrated superior accuracy performance compared to previous studies, achieving 96.5% accuracy.

Keywords: Amniotic fluids, Pretrained network, Feature selection, Oversampling, XGBoost.

1. Introduction

The objective of amniotic fluid examinations is to diagnose the volume and turbidity levels during pregnancy. This fluid serves to protect the fetus from impacts against the uterine wall and pressure on the umbilical cord in case of contractions [1. 2]. The volume increases with the progress of the pregnancy reaching approximately ± 50 ml, ± 350-400 ml, and 1000 ml at 12 weeks, 20 weeks, and 35-38 weeks respectively [3]. The method used to measure amniotic fluid volume (AFV) involves the single deepest vertical (SDP) technique with accuracy depending on the experience and technical expertise of the operator. To categorize the volume of amniotic fluid, the doctor marks the length of the single deep pocket (SDP) by drawing a line between two vertically aligned points, originating from the fluid pocket and intersecting with the uterus [4. 5]. Based on the indicated length of SDP, the volume is categorized into 3 classes namely normal, oligohydramnios (reduced), and polyhydramnios (excessive). Oligohydramnios contributes to perinatal morbidity, including an increased risk of cesarean delivery due to fetal distress, low Apgar scores, neonatal intensive care unit admissions, low birth weight, and meconium aspiration syndrome [6]. Meanwhile, polyhydramnios can lead to fetal structural abnormalities, genetic disorders, fetal anemia, placental tumors, multiple pregnancies, and maternal diabetes, which is mostly idiopathic [7]. Considering these anomalies, a tool capable of accurately measuring amniotic fluid volume is needed.

Antenatal AFV estimation is a fundamental measurement required during antenatal ultrasound examinations [6]. Two-dimensional amniotic fluid

ultrasound images (US) are essential non-radiation imaging techniques utilized for detecting and categorizing amniotic fluid volume. Ayu et al., classified amniotic fluid conditions into 6 classes namely oligohydramnios clear, oligohydramnios echogenic, polyhydramnios clear, polyhydramnios echogenic, normal clear, and normal echogenic [5]. The analysis used a classification model that combined rule-based and random forest methods. The rule-based method was employed to classify AFV based on SDP features, while random forest was predicated on first order statistical (FOS) and gray level co-occurrence matrix (GLCM) features, with an accuracy performance of 90.52% [5]. This study stated that there were still shortcomings in detecting amniotic volume, which was greatly influenced by segmentation results. Images with high noise levels and obstacles will result in poor segmentation results. Moreover, Ayu et al. categorized amniotic fluid conditions based on the single deep pocket method and texture features (FOS and GLCM) into 6 classes using SVM (RBF kernel) with an accuracy performance of 81.4% [8]. Namun hasil ini masih dirasa kurang optimal karena pada kelas minoritas hasil akurasi masih rendah.

Further studies including [9], classified AFV into 2 classes (normal and abnormal) based on transfer learning with MobileNet, achieving an accuracy of 0.94. However, multiclass classification was not implemented in the current study due to the small number of patients with oligohydramnios and polyhydramnios.

The examination published by [10] employed fuzzy techniques to measure and divide AFV into oligohydramnios, borderline, normal, and polyhydramnios, with an accuracy of 0.925. However, the proposed work did not work for high-risk patients.

Several studies related to 2D amniotic fluid images focused on the segmentation process, for instance, a previous study by [6] employed a deep learning model called AF-Net, an extended version of U-Net. The proposed method combined various ideas, including dilated convolutions, multiscale side-input, and side-output layers, using a dataset of 435 ultrasound images and 5-fold cross-validation. The performance, as measured by the dice similarity coefficient (DSC), was 0.877. Another study [11] implemented a deep learning (DL) network comprising AF-Net and an auxiliary network, which used a dataset of 2380 ultrasound images, achieving a DSC performance of 0.8559. Furthermore, [12] performed segmentation on 50 datasets using the proposed pixel classification method and achieved a DSC performance of 0.814 by comparing local window techniques. A previous study [13] also proposed pixel classification with a specified window size limit and in combination with several feature extractions such as gray-level, gray-level-local variance, and distance angle pixel to identify the amniotic fluid area, yielding a DSC performance of 0.876. Another investigation related to amniotic fluid ultrasound image segmentation [1] used the U-Net method and tuned hyperparameters. The best U-Net segmentation performance was obtained by combining RMSprop optimizer parameters, a binary entropy loss function, a learning rate of 0.00001, and 33 epochs, with a DSC performance of 0.88.

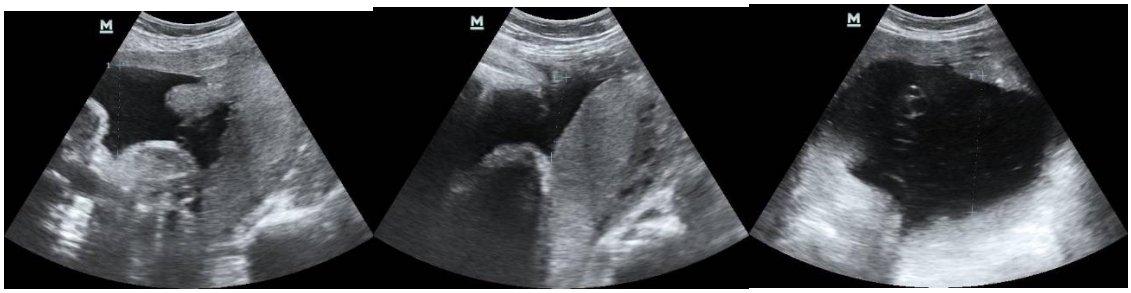
Based on the related studies concerning AF images and amniotic fluids volume classification, the previous results have not reached their maximum performance. Most previous research still relies on object segmentation results to detect amniotic volume. Images with high noise and obstacles sometimes make the region of interest (ROI) less precise. This result measures the deepest area invalid and different from the doctor's results. Apart from that, the condition of small datasets and imbalanced data is one of the problems in the learning process of the classifier used.

Therefore, the novelty of this study proposed a model for classifying AFV by applying feature extraction using the pre-trained deep learning model ResNet101. The application of this method for feature extraction is yet to be carried out on 2D amniotic fluid image objects. The influence of noise and obstruction can be reduced by using a CNN architecture to obtain features that differentiate amniotic volume.

Furthermore, to reduce the dimensions of features that are too high from the feature extraction results and prevent the curse of dimensionality from occurring. Chi-square was employed to select relevant features for AFV, and then oversampling was performed with safe-level SMOTE using k-NN rules to classify extracted feature pools and generate new samples [14].

This simplifies the architecture and reduces processing complexity for small datasets. It is efficient for small-dimensional datasets due to the nonparametric nature and sample classification based on K-NN votes. The classification method utilized a boosting approach using XGBoost [15], which was considered a highly effective technique for medical image classification [15-17].

The remaining part of this research is organized as follows. Section II explains data and method, provides data acquisition and details the proposed method/model. Section III contains experimental results in an analysis of the proposed method, and section IV concludes the research.



(a) (b) (c)

Figure. 1 Amniotic fluid USG image: (a) Normal, (b) Oligohydrmanios, and (c) Poligohydramnios

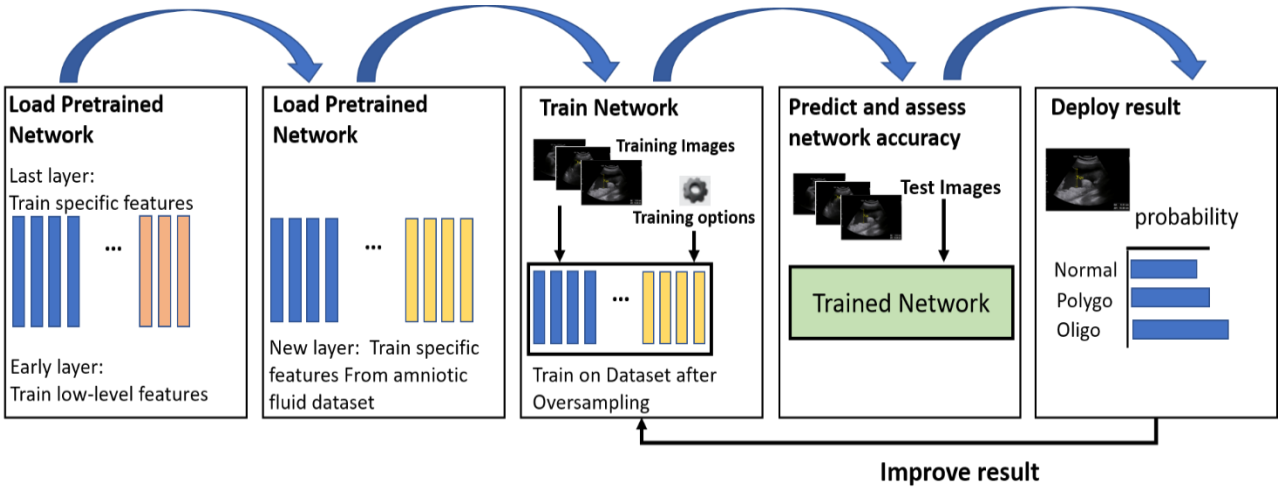


Figure. 2 Pre-trained network concept for amniotic fluid classification

2. Data and method

2.1 Data acquisition

This study, 130 amniotic fluid images were obtained in collaboration with an obstetrics and gynaecology doctor in Kasih Medika Bali Clinic, Denpasar, Indonesia. All data were acquired using an ultrasound machine with Accuvix XG and Transducer specifications, operating at a frequency of 3.5 Hz, lateral resolution of 3 mm, and saved in .jpg format, while the image size was 800 x 600 pixels. Additionally, the data used excluded data included those from pregnant women with obesity, single pregnancies, and gestational ages over 13 weeks. Fig. 1 shows amniotic fluid USG images categorized into three volume classes.

2.2 Extraction of image feature from the deep residual network architecture

Feature extraction was carried out using the pre-trained ResNet-101 model, which involved adapting a previously developed CNN architecture to a new dataset. The architecture consisted of 101 weighted layers, concluding with a global average pooling layer and a 1000-way fully connected layer with softmax classification (He et al., 2016). ResNet-101 was pre-trained using ImageNet, and the initial weights were used as the starting point of classification. Fig. 2 illustrates the steps of the pre-trained network model for classifying different types of amniotic fluid.

ResNet introduced the concept of shortcut connections, where the features from the previous layer were added to the output of the current layer. This approach minimized the loss of important features during convolution. Generally, ResNet-101 consisted of 5 stages of convolution processes followed by average pooling, feature selection using chi-square, oversampling with the safe level SMOTE method, and ending with a fully connected layer as the prediction layer. In this study, the fully connected neural network was replaced with a classifier using XGBoost to focus on learning from the minority data in the amniotic fluid dataset. Fig. 3 illustrates the architecture of the pre-trained network model (ResNet-101) for AFV classification.

2.3 Chi-square feature selection method

The Chi-square numerical test was used to measure the deviation from the expected distribution by considering independent features that were not

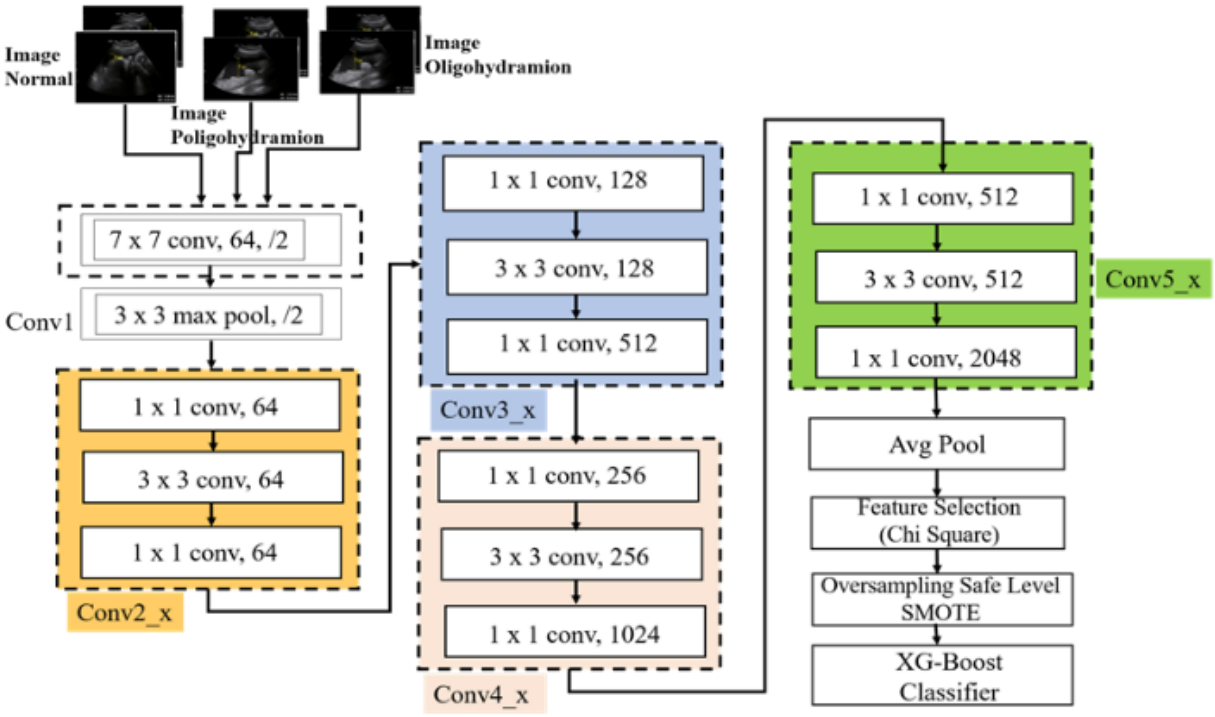


Figure. 3 Proposed pre-trained network (ResNet-101) model for amniotic fluid volume classification

dependent on the class values. The chi-square value was calculated using metrics such as true positive ($t_{p,}$), false positive (f_p), true negative (t_n), false negative (f_n), the probability of the total positive cases (P_{pos}), and the probability of the total negative cases(P_{neg}) [18], shown in Eq. (1).

$$Chi\text{-}square\_metric = t\,(t_{p,}(t_p + f_p)P_{pos}) + t(f_{n,}(f_n + t_n)P_{pos}) + t(f_{p,}(t_p + f_p)P_{neg}) + t(f_{n,}(f_n + t_n)P_{neg}) \quad (1)$$

Where *t* (count, expert) = (count-expect)² / expect.

2.4 Safe level SMOTE

Safe-level SMOTE represented an oversampling method derived from the traditional approach which had significant drawbacks such as overly noisy samples, and blindly generalizing the minority class regionwithout including the majority [19]. This causes overgeneralization problems and increases the possibility of the newly synthesized sample overlapping with the majority class. The safe-level SMOTE method assigned each positive instance to a safe level before creating a synthetic instance. Every synthetic instance was positioned closer to the greatest safety level, and this improved the performance prediction of the classifier in the minority class [20]. Furthermore, safe-level SMOTE applied k-NN rules to classify pool extracted features to generate new samples. This has the advantage of simplifying the architecture and reducing processing complexity for small data sets. The nonparametric nature and the sample classification through K-NN voting [14] contributes to its effectiveness. The steps taken in the Safe-level SMOTE method based on k-NN involve two main processes, as follows [20]:

A. Determine the criteria area to generate the synthetic data

1. Calculate the distance between instances of the minority class in the training set (p) and their nearest neighbors (n) using Euclidean distance, $d(x,y) = \sqrt{\sum_{i=1}^{z}(x_i - y_i)^2}$ .Where x, y represents the two points in Euclidean z -space, x_i, y_i stands for the Euclidean vector starting from the origin of the space (initial point), and z is the z –space.

2. Execute the K-NN method with a value of k = 5. Illustration of Safe-Level SMOTE using nearest neighbors with k of 5.

3. Randomly select one of the k nearest neighbors n obtained from the minority class.

4. Recalculate the distance between and its neighbor k using the same Euclidean distance.

5. Randomly select one of the k nearest neighbors n obtained from the minority class.

6. Calculate the safe level for p and n, Sl_p and Sl_n using Eqs. (2) and (3).

$$safe\ level\ at\ instance\ p, sl_p = the\ number\ of\ positive\ instance\ in\ k - nearest\ neighbors\ for\ p \quad (2)$$

Table 1. Rule of synthesis minority instance in safe-level SMOTE [20]

Case	sl_ratio	sl_p	Synthesis at a range between p and n, gap	Description
1	$= \infty$	0	Do not produce a positive synthetic instance	Both p and n instances are noise
		$0 \neq$	$gap = 0$	The instance is noise, therefore, synthesis is close as possible to the location of the p instance, and the synthetic data will be generated far from n by duplicating p
2	=1		$0 \leq gap \leq 1$	The safe level of p instance is the same as with n instance. The synthetic data will be generated along the line between p and n because p is as safe as n
3	>1		$0 \leq gap \leq 1/sl\_ratio$	The p instance is safer than the n instances, hence, synthesis is closer to the p position. The synthetic data will be generated closer to p at a distance [0.1/SLR]
4	< 1		$(1- sl\_ratio) \leq gap \leq 1$	The n instance is safer than the p instance, therefore, synthesis is closer to the n position. The synthetic data will be generated closer to n at a distance [1-SLR,1]

$$safe\ level\ at\ instance\ n, sl_n = the\ number\ of\ positive\ instance\ in\ k - nearest\ neighbors\ for\ n \quad (3)$$

7. Calculate the safe-level ratio for p and n based on Eq. (4), then the ratio was categorized into four cases such as sl_ratio =∞, sl_ratio = 1, sl_ratio < 1, and sl_ratio > 1. Each rangesl_ratio determines the 'gap' location range, enabling the synthesis of new samples. After determining the case for each positive instance, new synthetic instances (s_i) were generated in coordinates s_i as represented by Eq. (5), where *pi* represents the coordinate of the chosen positive instance. The gap value was calculated based on the corresponding positive instance case, as determined by the safe-level ratio. The variable $diff$ represents the distance between instance p and the selected nearest neighbor n. However, synthetic minority instances were not generated when the safe-level ratio was infinity and the safe level of instance p was zero.

$$safe - levelratio, sl_{ratio} = sl_p / sl_n \quad (4)$$

$$si = pi + (gap\ x\ diff) \quad (5)$$

B. Generate the synthetic dataset based on a safe level SMOTE.

Generate synthetic data based on safe-level ratios, the rules outlined in Table 1 were followed [20].

1. Determine the difference between *p* and *n*.

2. Refer to the range of random numbers based on the safe level ratio (*sl_ratio*) obtained in Table 1.

3. The difference obtained in Step 1 was multiplied by a random number obtained in Step 2.

4. The result obtained from Step 3 was added to *p* to generate the new instances.

5. The steps were repeated until the number of minority class observations was

2.5 Classification with XG-Boost

Extreme gradient boosting (XGBoost) is a relatively new and increasingly popular tree-based algorithm for data classification, which has proven to be a very effective method [16]. It is a highly scalable end-to-end tree-boosting system used in machine learning for classification and regression tasks [21]. This method can accept null values, scale imbalanced data, learn from previous mistakes, and fine-tune hyperparameters. Furthermore, the use of XGBoost improves the performance of weak learners in classification and regression tasks. When the previously predicted values for the data provided as input are considered, then this method can create a new tree to optimize the increased predictions generated. In the proposed model, the XGBoost classifier replaced the fully connected layer (FCL) in ResNet 101. This substitution arises from the use of non-medical related images in the original FCL classifier on the ImageNet dataset [15]. XGBoost uses an ensemble of *K* classification and regression trees (CARTs), each of which has $K_E^i | i \in 1 \ldots K$ nodes. The final prediction is the sum of the prediction scores for each tree shown in Eq. (6)-(14) [22].

$$\hat{y}_1 = \varphi(X_i) = \sum_{k=1}^{K} f_k(X_i), f_k \in F, \quad (6)$$

Where 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:

$$L(\varphi) \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (7)$$

The first term, *l*, represents the differentiable loss function, which measures the difference between target *yi* and the prediction $\hat{y}_i$. The second term avoids overfitting: Ω penalizes the complexity of the model:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} w_j^2 \quad (8)$$

Where γ, λ 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. It 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:

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

$$\tilde{L}^{(t)} = \sum_{i=1}^{n} \left[g_i f_i(X_i) + \frac{1}{2} h_i f_i^2(X_i) \right] + \Omega(f_t) = \sum_{i=1}^{n} \left[g_i f_i(X_i) + \frac{1}{2} h_i f_i^2(X_i) \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} w_j^2 = \sum_{j=1}^{T} \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T \quad (9)$$

$$g_i = \frac{\partial l\left(\hat{y}_i^{(t-1)}, y_i\right)}{\partial \hat{y}_i^{(t-1)}} \quad (10)$$

$$h_i = \frac{\partial^2 l(\hat{y}_1^{(t-1)}, y_i)}{\partial \left(\hat{y}_i^{(t-1)}\right)^2} \quad (11)$$

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:

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_i} h_i + \lambda'} \quad (12)$$

$$\tilde{L}^{(t)}(q) = -\frac{1}{2} \sum_{j=1}^{T} \frac{\left(\sum_{i \in I_i} g_1\right)^2}{\sum_{i \in I_i} h_i + \lambda} + \gamma T \quad (13)$$

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:

$$L\ split = \frac{1}{2} \left[\frac{\left(\sum_{i \in I_L} g_i\right)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{\left(\sum_{i \in I_R} g_i\right)^2}{\sum_{i \in I_R} g_{i+\lambda}} + \frac{\left(\sum_{i \in I} g_i\right)^2}{\sum_{i \in I} h_i + \lambda} \right] - \gamma \quad (14)$$

2.6 Evaluation and validation performance

To evaluate the performance of the classifier, this study used a multiclass confusion matrix. Typically, the minority and majority class labels were set as positive and negative respectively. Performance validation for classification involved four variables namely accuracy, precision, recall, and F1-score. Meanwhile, parameters for measuring the performance of AFV classification employed accuracy, precision, and recall, as indicated in Eqs. (15)-(17) [12].

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (15)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (16)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (17)$$

TP is true positive (a positive label predicted as an actual label), FP is false positive (a negative label but predicted as a positive label), TN is true negative (negative data that is predicted correctly), and FN is false negative (a positive label but predicted as a negative label).

3. Result and discussion

3.1 Experimental setting

The feature extraction results from amniotic fluid images yielded data with 2048 dimensions. The experiments were conducted using k-fold cross-validation, with k set to 5. Before experimentation, the data was divided into training and testing sets at a ratio of 70:30. Testing was then carried out with two observations using non-feature-selected data and applying feature selection. The experiments compared the results of previous state-of-the-art methods in oversampling development, such as adaptive synthetic oversampling (ADAYSN) [23], borderline-SMOTE [24], and safe-level SMOTE [14]. Additionally, this study compared the results with previous examinations related to amniotic fluid classification with the proposed model.

Table 2. Performance metrics on amniotic fluid testing data without feature selection step

Metric Performance	None	XGboost-SMOTE	XGboost-ADAYSN	XGboost-Borderline	XGboost-Safelevel
Accuracy	58.2	90.7	81.5	96.3	88.2
Precision	60.1	91.2	83.2	96.2	87.9
Recall	59.3	90.7	81.2	96.2	88.2
F1-Measure	59.2	91.4	82.1	95.4	88.1

Table 3. Performance metrics on amniotic fluid testing data with feature selection step

Metric Performance	None Oversampling + Selection Feature	XGboost-Selection Feature+SMOTE	XGboost-Selection Feature+ADAYSN	XGboost-Selection Feature+Borderline	XGboost-Selection Feature+Safe-level
Accuracy	65.2	92.1	86.2	93.5	**96.5**
Precision	64.3	92.5	86.2	93.2	96.3
Recall	65.3	92.6	86.2	93.1	97.3
F1-Measure	66.1	91.2	86.2	93.3	96.4

Table 4. Comparison of classification results with previous state-of-the-art models

Authors	Object	Class	Methods	Accuracy %	Precision %	Recall %	F1-Measure %
Ayu et al[5]	Volume Amniotic Fluid	Normal, Oligo, Poligo	SDP	76.9	76.8	76.9	77.1
khan et al[9]	Volume Amniotic Fluid	Abnormal and Normal	MobileNet	94	96	94	95
Ayu et al[8]	Volume and Echogenicity	Normal Echogenic, Normal clear, Oligohydramnios echogenic, Oligohydramnios normal, Polyhydramnios normal	GLCM, FOS, and SDP feature	81.4	80.8	81.8	81
Amuthadevi et al [10]	Volume Amniotic Fluid	oligohydramnios, borderline, normal, and hydramnios	fuzzy logic algorithm	92.5	-	-	-
Proposed Model	**Volume Amniotic Fluid**	**Normal, Oligohydramnios, Poligohydramnios**	**CNN Feature extractor, Chi-Square selection feature, Safe Level SMOTE, and XGBoost**	**96.5**	**96.3**	**97.3**	**96.4**

3.2 Experiment on amniotic volume classification

The first experiment analyzed the performance of feature-extracted data without selection as shown in Table 2. In this test, the amniotic fluid data had a dimensionality of 2048 and was evaluated using four oversampling methods. The results indicated that without feature selection, the XGBoost model had the lowest performance, achieving an accuracy of 58.2%,

precision of 60.2%, recall of 59.3%, and an F1-Measure of 59.2%. The best result from this test was obtained with the combination of XGBoost and Borderline-SMOTE, yielding an accuracy of 96.3%, as well as precision, recall, and F1-Measure of 96.2%, 96.2%, and 95.4% respectively. These results showed that the Borderline-SMOTE approach of filtering data samples for creating synthetic data assisted the XGBoost algorithm in finding its optimal decision boundary. However, the high dimensionality affected the computational time.The next experiment assessed the impact of adding feature selection using the Chi-Square method. Feature selection involved selecting the top 5 features with the best chi-values from the 2048-dimensional data. Table 3 shows the results of XGBoost performance when combining feature selection and oversampling.

The results indicated a significant improvement in performance for each model with improved accuracy, precision, recall, and F1-measure of 7%, 4.2%, 6%, and 6.9% for the non-oversampled testing data. The most substantial performance enhancement occurred with the Safe-Level SMOTE oversampling method, resulting in values of 8.3%, 8.4%, 9.1%, and 8.3% respectively. These results showed that the effectiveness of feature selection using Chi-Square and the Safe-Level SMOTE oversampling method yielded a machine learning model capable of classifying amniotic fluid data with optimal performance. The challenges posed by imbalanced data were mitigated through the implementation of oversampling methods that exhibited greater resilience to noisy data conditions and overlapping region classes.

The ROC (receiver operating characteristic) value was also measured to assess the impact of adding oversampling and feature selection steps to the proposed model. Fig. 4 illustrates the ROC curve results for each tested model. Based on the ROC measurement results, the models with the addition of oversampling and Chi-Square feature selection steps showed significantly improved performance compared to the data without these methods. This demonstrated the enhanced capability of the XGBoost model when supplied with synthetic data and feature dimensions derived based on relevance and correlation with the class label. The ROC values for each class in the depicted figures were all above 0.85 and even reached the maximum point of 1.00 for class 2 (oligohydramnios). In contrast, without oversampling and feature selection, the performance for each class only achieved values of 0.52, 0.62, and 0.54 for class 0 (normal), 1 (polyhydramnios), and 2 (oligohydramnios) respectively. The ROC curve shows a graphical representation of the relationship between sensitivity and 1-specificity. In medical studies involving ultrasonography images of amniotic fluids, the ROC curve is widely used to illustrate diagnostic accuracy and determine the optimal cut-off value. Diagnostic accuracy was derived from the area under the ROC curve, and the optimal cut-off was utilized to identify positive and negative conditions in diagnosis. Finally, the proposed model was compared with previous detection models for AFV, as shown in Table 4.

The comparison we present in Table 4 is by comparing previous studies related to the classification of amniotic fluids for level of echogenicity and volume. In our previous research [5], we proposed the single deep pocket method to find the deepest line of the amniotic sac. This line search is based on the constituent pixel points vertically without being cut off by other objects. This study classifies amniotic fluid into three classes, namely normal, oligo, and poligo. The dataset used is still the same as in the current research. This study classifies amniotic fluid into three classes, namely normal, oligo, and poligo. The dataset used is still the same as in the current research. From the analysis results, this method is still vulnerable and very dependent on the results of the segmentation of the amniotic fluid. When the segmentation results are not good, it will have a big impact on decreasing the model's accuracy. Then, our research continued by developing a classification into six classes consisting of echogenicity level and amniotic fluid volume [8].

In this study, we propose a texture feature extraction method using the gray level co-occurrence metrics (GLCM) and first order statistics (FOS) methods to obtain the characteristics of amniotic fluid. Amniotic fluid turbidity is divided into two levels, namely echogenic and clear. Echogenic conditions give rise to a grainy texture and a higher grey level, whereas clear levels tend to be darker. This turbidity classification model and volume classification [5] in previous research were then combined with a rule-based method for each output produced from the two models so that there were six labels, namely normal echogenic, normal clear, oligohydramnios echogenic, oligohydramnios normal, and polyhydramnios normal. This research achieved an accuracy value of 81.4%, precision of 80.8%, recall of 81.8%, and F1-Measure of 81%.

We tried to find other studies that also used amniotic fluids as a research object. Khan et al. [9] conducted research by classifying amniotic fluids into two classes, abnormal and normal; this labelling is more general compared to our previous research and the proposed model in the current research. The dataset used consisted of 166 US images of pregnant

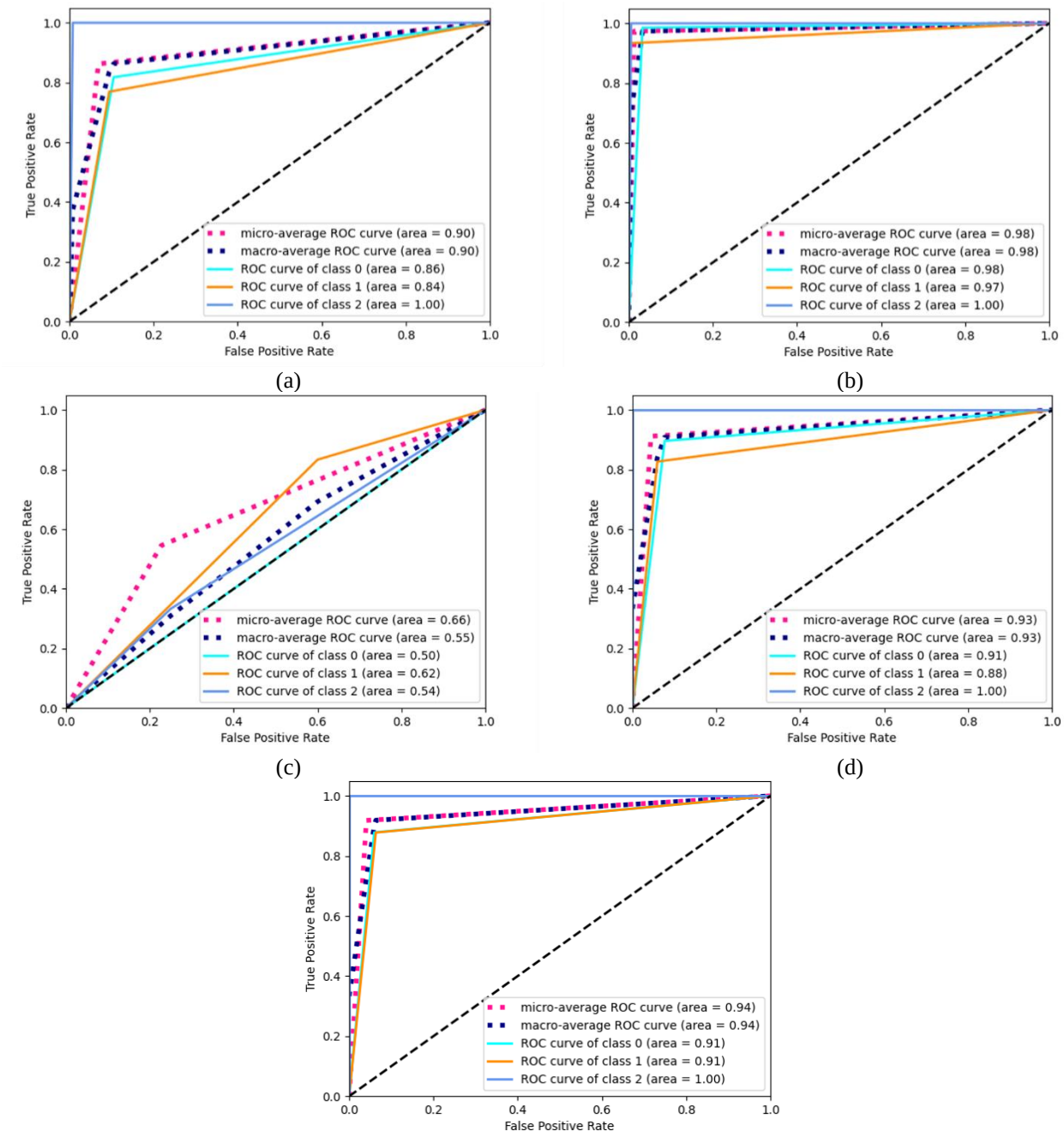


Figure. 4 ROC curve for: (a) None oversampling and feature selection, (b) Borderline oversampling and chi-square feature selection, (c) ADAYSN oversampling and chi-square feature selection, (d) Safe-level SMOTE oversampling and chi-square feature selection, and (e) SMOTE oversampling and chi-square feature selection

women. Dataset obtained from US images from King Fahd Hospital of the University (KFHU) and Elite Clinic in Dammam, KSA. The US device used to collect the images was the GE Voluson P6, and we collected 166 US images, among which 100 cases belonged to normal AF levels and 66 cases belonged to abnormal AF levels. The dataset access resource is closed and not to be shared with the public. Five transfer learning models, namely, Xception, Densenet, InceptionResNet, MobileNet, and ResNet, were applied to obtain the characteristics of normal and abnormal amniotic US images. From the test results, the MobileNet architecture produces the best performance, namely with an accuracy of 94%, precision of 96%, recall of 94% and F1-Measure of 95%.

The next research that still uses amniotic fluids as an object is to use the fuzzy logic algorithm method [10]. This study classifies amniotic fluids into four classes, namely oligohydramnios, borderline, normal, and hydramnios, based on volume level. The dataset originates from 2D ultrasound images of pregnant

women (gestational age is 27 to 40 weeks), which are collected and stored in the database. The images are in the DICOM format, but the dataset access resource is closed and not to be shared with the public. The shape templates are to be developed with deformable methods. After feature extraction, the contour points were identified to find the AFI. Then, prediction will be done by creating a fuzzy inference mechanism. The results of the tests found that the percentage of classification by the proposed inference system ranges from 91 to 94%. The average is 92.5%.

In the current study, classifying amniotic fluids is based on volume level, where the same cases and data are found in studies [5] and [8]. This method uses a pre-trained network Resnet 101 to extract features from images and replaces the neural network classifier with the XG-Boost algorithm. This method showed an increase in performance from previous research by achieving an accuracy value of 96.5%, precision of 96.3%, recall of 97.3%, and F1-Measure of 96.4%. These results show the effectiveness of using the XG-Boost method as a classifier and the effect of feature selection, as shown in Table 3 and Table 4, which reduces feature dimensions from 2048 to 5. Also, creating synthetic data/data augmentation also increases model performance. More data makes it easier for models to learn patterns and be smarter at generalizing new data.

4. Conclusion

In conclusion, this study successfully developed a machine-learning model capable of identifying amniotic fluids based on ultrasonography images. Amniotic fluids a crucial role in fetal growth and development during pregnancy, and the focus of this study was to measure the volume parameter for identification. The proposed model suggested a classification method for AFV using a pre-trained deep-learning network as a feature extractor. The process was accompanied by feature selection using the Chi-Square method to obtain the top 5 features most relevant to the data label. For the imbalanced data issue, specifically in the oligohydramnios class, the Safe-Level SMOTE oversampling method was used to generate new synthetic data, ensuring that this class was not excluded from the learning process and enabling better learning of data patterns. XGBoost was proposed as the classifier offering a parallel tree boosting algorithm (also known as GBDT, GBM) that provided efficient and accurate solutions to various data science problems. The proposed model was tested using 5-fold cross-validation. The testing involved three schemes namely, using non-feature-selected, and feature-selected data, as well as testing with oversampling methods. Furthermore, three oversampling methods were evaluated, including SMOTE, Borderline SMOTE, and Safe-Level SMOTE. The final testing phase compared the results with previous studies that utilized different approaches. Based on the conducted tests, the addition of oversampling and feature selection steps was found to improve the performance of the model. The impact of feature selection in reducing and selecting features was significant for the overall tested model. The handling of imbalanced data conditions through oversampling significantly enhanced performance, with precision, recall, and F1-measure improving alongside increased accuracy. The Safe-Level SMOTE oversampling method exhibited the highest performance enhancement among the oversampling methods. The use of sample data filtering effectively generated synthetic data that avoided overlapping with other class regions. The comparison of performance between this study and previous studies also indicated improvement. In this study, the characteristics of amniotic fluid were extracted using a CNN algorithm, ultimately achieving the highest accuracy performance of approximately 96.5%.

Future studies on amniotic fluid should focus on detecting fluid turbidity and identifying small or opaque objects that may affect fetal health. The combination of amniotic fluid volume identification and pouch detection provides a unified approach that can assist obstetricians in diagnosing patients.

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Author contributions

In this research article, the author's contributions are as follows: “Conceptualization, Putu Desiana Wulaning Ayu; Methodology, Putu Desiana Wulaning Ayu and Gede Angga Pradipta; Software, Putu Desiana Wulaning Ayu; Validation, Putu Desiana Wulaning Ayu and 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 and Gede Angga Pradipta; Writing-review and editing, Putu Desiana Wulaning Ayu; Supervision, Putu Desiana Wulaning Ayu; Visualization, Roy Rudolf H, Kadek Eka Sapta, Gede Edy Artana; Project Administrator, Roy Rudolf H, Kadek Eka Sapta; Funding acquisition, Putu Desiana Wulaning Ayu.

Acknowledgments

The authors are grateful to the Directorate of Research, Technology, and Community Service (DPRM) Indonesia for funding this study through Program Funding (Fundamental Reguler) for the 2023 Fiscal Year.

Notation list

p	:	Minority class in trainning set
n	:	Nearest neighbour
x, y	:	Two points in euclidean distance
pi	:	coordinate of the chosen positive instance
s_i		new synthetic instances
sl_p	:	Safe-level ratio minority class in trainning set
sl_n		Safe-level ratio nearest neighbour
$diff$	:	distance between instance p and the selected nearest neighbor n
X_i	:	members of the training set
y_i	:	corresponding class labels
f_k	:	leaf score for the k^{th} tree
F	:	set of all *K* scores for all CARTs
l		represents the differentiable loss function
$\hat{y}_i$	:	prediction
γ, λ	:	constants controlling the regularization degree
T	:	number of leaves
w_j^*	:	optimal weight

References

[1] P. D. W. Ayu and G. A. Pradipta, "U-Net Tuning Hyperparameter for Segmentation in Amniotic Fluid Ultrasonography Image", *2022 4th Int. Conf. Cybern. Intell. Syst. ICORIS 2022*, 2022.

[2] X. L. Tong, L. Wang, T. B. Gao, Y. G. Qin, Y. Q. Qi, and Y. P. Xu, "Potential function of amniotic fluid in fetal development-novel insights by comparing the composition of human amniotic fluid with umbilical cord and maternal serum at mid and late gestation", *J. Chinese Med. Assoc.*, Vol. 72, No. 7, pp. 368–373, 2009.

[3] G. Tam and T. A. Dughaishi, "Case report and literature review of very echogenic amniotic fluid at term and its clinical significance", *Oman Med. J.*, Vol. 28, No. 6, pp. 28–30, 2013.

[4] G. Demiris, "Physical Principles of Medical Ultrasonics, 2nd edition [Book Review]", *IEEE Eng. Med. Biol. Mag.*, Vol. 23, No. 4, pp. 81–82, 2004.

[5] D. S. N. Ayu, P. D. W. S. Hartati, and A. Musdholifah, "Amniotic Fluids Classification Using Combination of Rules-Based and Random Forest Algorithm", in *Soft Computing in Data Science*, 2022, p. 15.

[6] H. C. Cho, S. Sun, C. Min Hyun, J. Kwon, B. Kim, Y. Park, and J. K Seo, "Automated ultrasound assessment of amniotic fluid index using deep learning", *Med. Image Anal.*, Vol. 69, p. 101951, 2021.

[7] E. Yefet and E. D. Spiegel, "309: Polyhydramnios with normal detailed ultrasound examination – short and long term outcomes", *Am. J. Obstet. Gynecol.*, Vol. 212, No. 1, p. S166, 2015.

[8] 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 Lett.*, Vol. 15, No. 7, pp. 681–691, 2021.

[9] I. U. Khan, A. Alowayed, and F. M Anis., "Deep Learning-Based Computer-Aided Classification of Amniotic Fluid Using Ultrasound Images from Saudi Arabia", *Big Data Cogn. Comput.*, Vol. 6, No. 4, 2022.

[10] C. Amuthadevi and G. M. Subarnan, "Development of fuzzy approach to predict the fetus safety and growth using AFI", *J. Supercomput.*, Vol. 76, No. 8, pp. 5981–5995, 2020.

[11] 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.

[12] P. D. W. Ayu and S. Hartati, "Pixel Classification Based on Local Gray Level Rectangle Window Sampling for Amniotic Fluid Segmentation", *Int. J. Intell. Eng. Syst.*, Vol. 14, No. 1, pp. 420–432, 2021, doi: 10.22266/ijies2021.0228.39.

[13] 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", *Appl. Soft Comput.*, Vol. 107, 2021.

[14] C. Bunkhumpornpat, K. Sinapiromsaran, and C. Lursinsap, "Safe-Level-SMOTE : Safe-Level-Synthetic Minority Over-Sampling TEchnique", *Pacific-Asia Conf. Knowl. Discov. data Min.*, pp. 475–482, 2009.

[15] X. Y. Liew, N. Hameed, and J. Clos, "An

investigation of XGBoost-based algorithm for breast cancer classification", *Mach. Learn. with Appl.*, Vol. 6, No. August, p. 100154, 2021.
[16] J. Parashar, Sumiti, and M. Rai, "Breast cancer images classification by clustering of ROI and mapping of features by CNN with XGBOOST learning", *Mater. Today Proc.*, No. xxxx, 2021.
[17] S. H. Karaddi and L. D. Sharma, "Automated multi-class classification of lung diseases from CXR-images using pre-trained convolutional neural networks", *Expert Syst. Appl.*, Vol. 211, No. June 2022, p. 118650, 2023.
[18] I. S. Thaseen and C. A. Kumar, "Intrusion detection model using fusion of chi-square feature selection and multi class SVM", *J. King Saud Univ. - Comput. Inf. Sci.*, Vol. 29, No. 4, pp. 462–472, 2017.
[19] Z. Jiang, T. Pan, C. Zhang, and J. Yang, "A new oversampling method based on the classification contribution degree", *Symmetry (Basel).*, Vol. 13, No. 2, pp. 1–13, 2021.
[20] S. N. S. S. Daud, R. Sudirman, and T. W. Shing, "Safe-level SMOTE method for handling the class imbalanced problem in electroencephalography dataset of adult anxious state", *Biomed. Signal Process. Control*, Vol. 83, No. February, p. 104649, 2023.
[21] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system", In: *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, Vol. 13-17-Augu, pp. 785–794, 2016.
[22] S. Thongsuwan, S. Jaiyen, A. Padcharoen, and P. Agarwal, "ConvXGB: A new deep learning model for classification problems based on CNN and XGBoost", *Nucl. Eng. Technol.*, Vol. 53, No. 2, pp. 522–531, 2021.
[23] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning", In: *Proc. Int. Jt. Conf. Neural Networks*, No. 3, pp. 1322–1328, 2008.
[24] H. Han, W. Wang, and B. Mao, "Borderline-SMOTE : A New Over-Sampling Method in", *Int. Conf. Intell. Comput.*, pp. 878–887, 2005.