



Improving Classification Performance of Fetal Umbilical Cord Using Combination of SMOTE Method and Multiclassifier Voting in Imbalanced Data and Small Dataset

Gede Angga Pradipta^{1,2} Retantyo Wardoyo^{3*} Aina Musdholifah³
 I Nyoman Hariyasa Sanjaya⁴

¹Doctoral Program Department of Computer Science and Electronics,
 Faculty of Mathematics and Natural Science, Universitas Gadjah Mada, Yogyakarta, Indonesia

²Department of Information Technology, Faculty Computer and Informatics,
 Institut Teknologi dan Bisnis STIKOM Bali, Bali, Indonesia

³Department of Computer Science and Electronics, Faculty of Mathematics and Natural Science,
 Universitas Gadjah Mada, Yogyakarta, Indonesia

⁴Department of Obstetrics and Gynecology,
 Faculty of Medicine Udayana University/Sanglah General Hospital, Bali-Indonesia

* Corresponding author's Email: rw@ugm.ac.id

Abstract: The umbilical cord is one of the important organs on the growth and development of the fetus in the womb. Umbilical cords are associated with an adverse perinatal outcome such as intrauterine deaths, preterm delivery, repetitive intrapartum fetal heart deceleration, operative delivery for fetal distress, meconium staining, and chromosomal abnormalities. Initial screening stages of the fetal umbilical cord are carried out by analyzing the coiling pattern of two umbilical arteries. In this study, we propose a relevant feature extraction for classifying this organ based on texture and morphological approach. However, this study is facing an imbalanced class problem, which leads to the inability of the traditional classifier to predict data in the minority class. To deal with the emerging issues, this study proposed a model by optimizing data and algorithmic levels using a combination SMOTE method and Multiclassifier Voting. At the data level, the SMOTE method is used to generate new synthetic data and to balance the skewed data distributions directly. Subsequently, the classification uses a multiclassifier method that combines several traditional classifier methods in making final decisions based on voting schemes. The first experiment was conducted on imbalanced and small size data with a total of 62 umbilical cord images from 3 classes namely hypercoiling, hypocoiling, and normalcoiling. The results showed the multiclassifier voting method was able to achieve the best results with an accuracy of 73%, average recall of 72%, and ROC of 48% compared to other classifier methods such as SVM, Random Forest, KNN, Naïve Bayes, and Decision Tree (C.45). However, all classifiers failed to predict the normocoiling class because of the limited amount of normocloing data in the training phase. Then the second experiment was carried out by adding synthetic data using the SMOTE method with the total data increasing to 111 images spread evenly in each class. The results show a combination of multiclassifier voting and SMOTE methods ultimately leading and produced higher performance than other classifiers, which yielded and accuracy of 81.4%, average recall of 80%, average precision of 81.5%, and ROC of 89.1%.

Keywords: Umbilical cord, Feature extraction, Imbalanced data, SMOTE, Multiclassifer voting.

1. Introduction

The umbilical cord is a connective tissue between the placenta and fetus which aims to maintain viability, remove the residual compounds, and

transport oxygen, nutrients, and antibodies in the womb [1]. The umbilical vessels are covered by mucopolysaccharide also known as Wharton's jelly, which consists of one or two arteries connecting the fetal to the placenta. The length of an umbilical cord

is between 30-100 cm, with an average of 44-55 cm, while the average number of coils in 9.2 cm is 5-6 cm, with a strong and flexible structure [2]. It is classified in three forms namely Normocoiled, Hypocoiled (Umbilical Index below 10th percentile), and Hypercoiled (Umbilical Index above 90th percentile) [3]. At delivery, the density of umbilical vascular coiling was assessed quantitatively using the UCI, determined by dividing the number of complete vascular coils in a given cord by the cord's length in centimeters [4]. Hypercoiling slows down the food intake of the fetus due to excessive coiling. This condition tends to disrupt the growth and development of the fetus, which requires immediate intervention from the doctor. Hypocoiling occurs when there is no coil in the veins from the arteries, thereby, leading to a knotted umbilical cord condition, and blocking the blood flow in the vessels. Hypocoiled umbilical cords are associated with intra-uterine fetal death (IUFD), fetal growth restriction, fetal distress, low Apgar scores, fetal congenital anomalies, and abnormal insertion of the umbilical cord [5]–[7]. It is important to solve the problem associated with the umbilical cord through the early provision of adequate information. Coiling patterns are determined with the use of an ultrasonography machines in Doppler mode, which shows the flow of blood vessels in blue and red color. Unfortunately, USG machine do not have the ability to provide information on the umbilical cord category due to its effect on the flow of blood, oxygen, antibodies, and nutrients needed by the fetus. Doctors perform manual analysis to determine the category of the umbilical cord so quite time-consuming.

This study aims to classify umbilical cord on gestational into three types using the combination of morphological features and texture feature. We observe texture feature using Gray Level Co-occurrence Matrix (GLCM) in grayscale and also RGB color space. The first contribution in this paper is to identify features that can be used to distinguish three types of umbilical cord. Proper use of features that greatly affect the performance of the classification model. The second contribution, we proposed model combines the data level and algorithmic level to improve the predictive performance of the classification method, especially in minority class data. The combination of the Synthetic Minority Oversampling Technique (SMOTE) used to create synthetic data and Multiclassifier voting used to improve the prediction performance of a model.

In fact, a data set with an unequal number of instances for different classes is called imbalanced data set [8]. This skewness in the data underlying

distribution causes many problems for typical machine learning algorithms. In particular, important and crucial information is in the minority class. incorrectly classifying examples of minority classes is the main problem in processing this data. Simply said, the main point of learning is to develop a classifier that will provide high accuracy for minority classes without reducing performance in the majority class.

The rest of the paper is organized as follows: Section 2 reviews several feature extraction method and model to handling imbalance class problem, Section 3 presents the proposed classification model, Section 4 focuses on evaluation method, Section 5 provides experimental results and analysis, and Section 6 elaborates the conclusion of the study.

2. Related work

Several studies have proposed methods for classifying objects from ultrasound images using various features. This section describes several types of features used in ultrasound image classification and several methods for tackling imbalanced class issues.

The texture characteristics of the combination between the GLCM method and the Two Dimensional Discrete Wavelet Transform (2D-DWT) implemented to obtain the characteristics of MRI images from breast cancer more accurately with the multiresolution analysis concept proposed in research [9]. This concept then is analyzed by zooming the image in and out, with the Wavelet method used to decompose it into several sub-images with varying resolutions. A similar model was also proposed by [10] through the combination of the 2D-DWT method with GLCM. The difference lies in the division of images into several parts called blocks with a size of 64x64 which become the Region of Interest (ROI). Furthermore, [12] breast cancer was categorized into Triple-Negative (TN) and Non-Tiple-Negative (NTN), with its texture, shape, and vascularity used to determine the best testing combination. Subsequently, [11] classified breast cancer into benign and malignant categories by extracting 21 morphological features such as Area, Perimeter, Elliptic Normalized Circumference, Elliptic-Normalized Skeleton (ENS), Long axis to Short axis ratio (LS Ratio), Form Factor, Roundness, Solidity, Convexity, Extent, Tumor area to circle area (TCA), Tumor perimeter to ellipse perimeter (TEP), TEP difference (Tumor perimeter - Ellipse perimeter), Tumor perimeter to circle perimeter (TCP), TCP difference (Tumor perimeter - Circle perimeter), AP Ratio, Normalized residual value (NRV),

Eccentricity, Compactness, Elongation Circularity and combining them with the texture features generated by the GLCM method and fractal features. The research conducted by [12] detects and monitors ovarian follicles on two-dimensional ultrasonography images to segment and calculate the number of follicles contained in the ovary during its health check. the feature extraction was used to detect and differentiate follicles from other objects in the ovary using Area, Perimeter, Centroid, Major and Minor Axis Length, and Compactness. Furthermore, [13] conducted research with a combination of features based on shape, color space, and Gabor for early identification of nucleus in the diagnosis of cervical cancer. The shape features used include Tortuosity, Circularity, and Elliptically Normalized Skeleton which aims to distinguish stacked nuclei from single nuclei cells. To predicted fetal weight in pregnancy period, research conducted by [14] used shape information of abdomen circumferences and biparietal diameter as a feature to identify fetal growth.

However, there are some issues associated with studying the data patterns such as small sample size, imbalanced class distribution, and overlapping or complexity due to difficulty in data collection [15]. The imbalanced data conditions complicate the classifiers in minority classes, therefore poor accuracy results are obtained in its prediction [15–17]. This condition tends to occur when the amount of data between classes in the dataset are not balanced. Classes with small amounts of data are called minority, while the large ones are called the majority. The condition of imbalanced data has become a challenge and a problem in data mining because it often appears in classifications of real-world data. Similarly, in some cases, data on medical diagnosis like breast cancer specimens [18, 19], Moffit cancer [20] and Melanoma diagnosis [21]. some of these cases require a data preprocessing to balance the amount of data in each class so that the classifier is able to generalize data characteristics better. In different cases the imbalanced data problem also appears on Fraud detection in the transaction [22], forecasting natural disaster [23], biological anomalies [24], facing challenges where important information is precisely in the minority class and demands high accuracy not only in the majority class but also in minority. According to [16] three approaches have been used and applied in various studies, namely, the optimization of the data, algorithm, and cost-sensitive learning model. [25] stated that the SMOTE oversampling method has been proven in many studies to possess the ability to overcome imbalanced data by making new synthetic

data based on the nearest neighbors' points. Research on the classification of cervical cancer [26–29], diabetes prediction [8], Lung cancer [30] used the SMOTE method in the preprocessing data stage. This method is proven to be able to improve the accuracy of classifier prediction compared with the original dataset.

Algorithmic or model-based solutions such as cost-sensitive methods, ensemble learning algorithms, and one-class learning are among the proposed ways of dealing with these problems [15]. The research was also carried out to detect faces with histogram features and ensemble classifiers [31]. The combination methods used are Artificial Neural Network (ANN), K Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest (RF), REP tree, Naive Bayes, CART. Petintrin [32] predicted the bioactive molecule using the majority voting method. The research conducted several experiments using a combination of single classifiers such as Support Vector Machine (SVM), Naive Bayes (NB), Decision Tree (DT), k-Nearest Neighbor (k-NN), and Random Forest (RF) to determine the highest accuracy. The majority voting method is used with the same weight for all classifiers. A research conducted by [33] used the weighting method to calculate the results of majority voting. This study, however, compares the filtering method, single classifier (biased and adaptive filter), and majority voting for a dataset from the UCI machine learning repository with imbalanced dataset characteristics. The dataset is divided into smaller sub-data and a classifier is provided to predict the class label of an instance. The Ensemble learning model in the research carried out by [34], used the Multiobjective Differential Evolution to weigh each classifier by measuring the results of class predictions (single classifier) of an instance. The parameters F-measure, Precision, and Recall at each classifier are used as a reference to determine the weight of the final decision on voting.

Based on this review, specific research using fetal umbilical objects has never been done before in the realm of pattern recognition or object classification. This study aims to classify three types of umbilical cord using a classification model that robust to imbalance class problems. The feature extraction stages refer to ultrasonography images previously studied using texture and shape approach. GLCM method in grayscale and RGB images used to obtain the right coiling pattern. Subsequently in the classification stage, this research proposes a novel model with a combination between the SMOTE oversampling method and the multiclassifier voting model. Optimization at both the data and algorithm

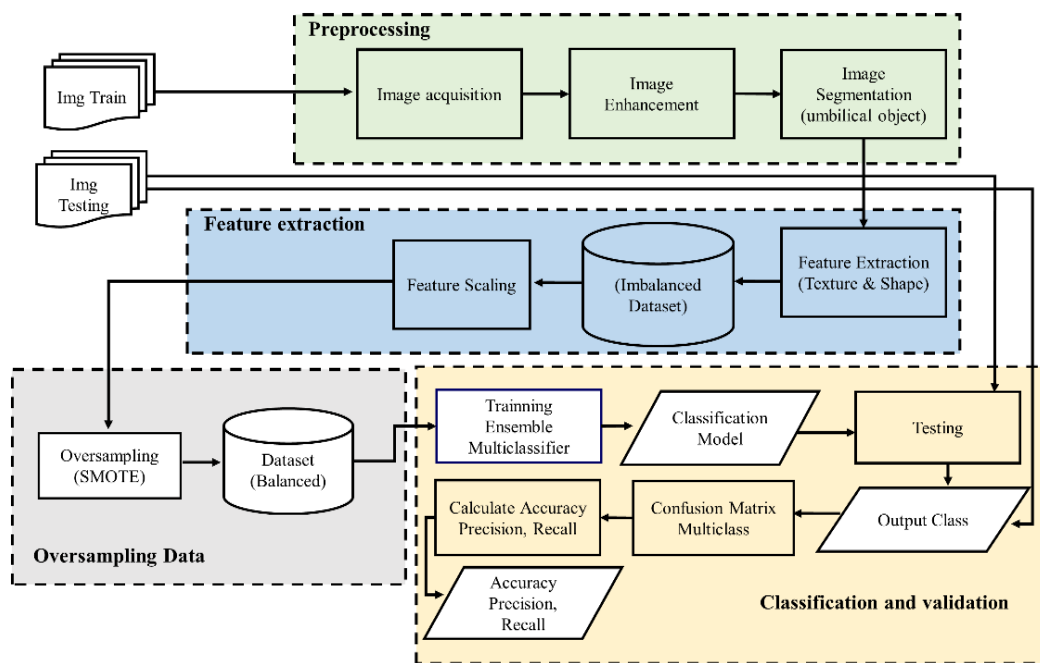


Figure. 1 Proposed classification model

level is expected to improve accuracy and precision compared to traditional classifier methods.

3. Research method

3.1 Proposed model

The proposed model in this study is divided into four processes namely image preprocessing, feature extraction, data oversampling, classification, and validation as shown in Fig. 1. Image data is divided into training and testing with a ratio of 75:25. The preprocessing stage starts with image enhancement to clarify the color contrast of the object, therefore it eases the segmentation stage. The search for region of interest (ROI) is carried out by segmenting umbilical cord objects from the background with the threshold method based on the HSV color space. After obtaining the segmentation results, the feature extraction with texture and shape approaches was performed to produce a vector dataset of each image.

The next process is oversampling to overcome the imbalanced data problems using the SMOTE method. The main idea was to overcome the overfitting rendered by simply oversampling by replication and assist the classifier to improve its generalization on the testing data. At the classification stage, to further improve performance results, conducted by trained several classifiers on training data and their evaluations are aggregated to produce the final

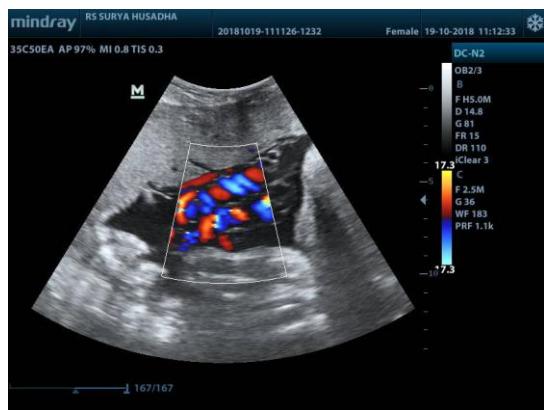
classification decision. To determine the performance of the proposed model, testing is carried out by comparing the accuracy, precision, and recall of the proposed model with the single classifier model.

3.2 Image segmentation

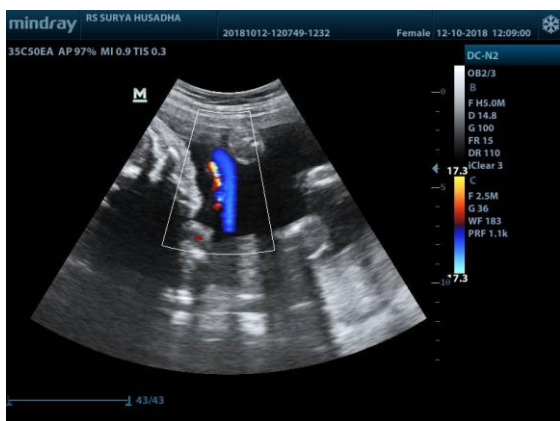
The screening on the fetal umbilical cord consists of red and blue colors, with the grayish foreground color. The various stages of the image segmentation are based on differences in color between the foreground and background. Fig. 2 shows sample images of (a) Hypercoling (b) Normal (c) Hypocoiling.

In digital imagery, the red, orange, yellow, green, blue, and purple colors of the visible spectrum are represented by the Hue value. Therefore, the umbilical cord image segmentation process is carried out in the HSV (Hue, Saturation Value) color space. The first step in this process is to crop the USG image to eliminate the information in the form of text obtained from the machine. Cropping is conducted using a rectangle according to the coordinates of the object to the size of 570 x 427.

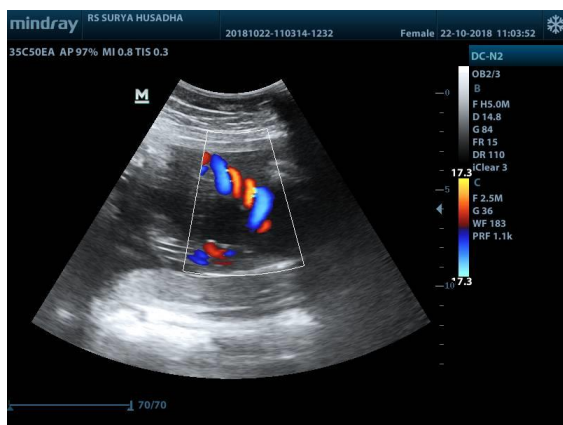
This is followed by transforming the image color space that was originally in the RGB format to HSV, accepted by human vision. Next is the segmentation process by using the thresholding approach in the Hue (H) and Saturation (S) color spaces as shown in



(a)



(b)

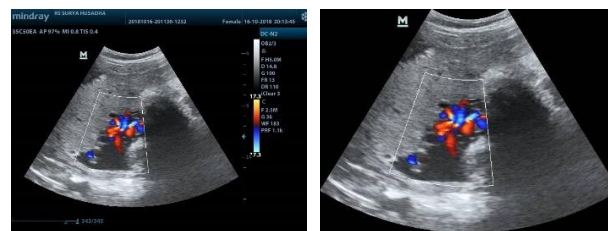


(c)

Figure. 2 Umbilical cord categories: (a) hypercoloring, (b)normal, and (c) hypocoiling

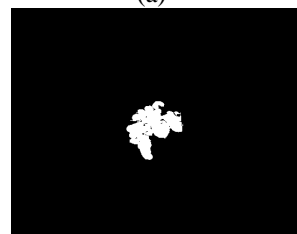
Eq. (1). The midpoint value of image segmentation is called thresholding, while the pixel values represent the background and others the object points. T is the threshold value. x, y are coordinates of the threshold value point and $f(x, y)$ are points the gray level image pixels. Threshold image $b(x, y)$ can be define:

$$b(x, y) = \begin{cases} 1, & \text{if } f(x, y) \geq T \\ 0, & \text{if } f(x, y) < T \end{cases} \quad (1)$$

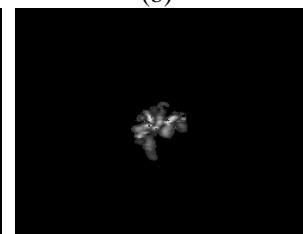


(a)

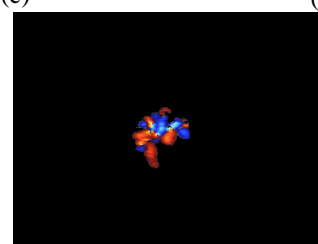
(b)



(c)



(d)



(e)

Figure. 3 Umbilical cord preprocessing result: (a) original images, (b) images after cropping, (c) segmented object (binary), (d) segmented object (grayscale), and (e) segmented object (RGB)

The segmentation results are refined at the edges with the removal of small objects using the opening and closing morphological operation method. This is followed by masking which is the combination of the segmented and original images with values of 1 and 0, respectively. Therefore, the masking results produce the fetal umbilical cord and black background as shown in Fig. 3.

3.3 Feature extraction

3.3.1. Gray level co-occurrence matrix (GLCM)

Gray Level Co-Occurance Matrix (GLCM) is included in the statistical method with a gray degree distribution (histogram) by measuring the level of contrast, granularity, and roughness of an area from the relationship between pixels in the image. This statistical paradigm has unlimited usage ability, therefore, it is suitable for natural textures. It is also used to obtain texture features by calculating the probability of a neighborhood relationship between two pixels at a certain angular orientation distance [35]. GLCM is a matrix with the same number of rows and columns similar to the G gray levels in the image. Matrix elements $P(i, j|\Delta x, \Delta y)$ are the relative frequencies between two pixels separated by

space $(\Delta x, \Delta y)$, with gray level probability intensity of i and j , with distance of d and a certain angular orientation of θ .

For example the adjacency matrix of $M \times N$ is an input image consisting of a G gray level from 0 to $G-1$, with $f(m, n)$ as the sample intensity of m columns and n rows for each adjacency. The GLCM matrix element is defined with the following Eqs. (2) - (5):

$$P(i, j|\Delta x, \Delta y) = WQ(i, j|\Delta x, \Delta y) \quad (2)$$

with

$$W = \frac{1}{(M - \Delta x)(N - \Delta y)} \quad (3)$$

$$Q(i, j|\Delta x, \Delta y) = \sum_{m=1}^{M-\Delta y} \cdot \sum_{n=1}^{N-\Delta x} A(m, n, j, \Delta x, \Delta y) \quad (4)$$

and

$$A = \begin{cases} 1 & \text{if } f(m, n) = \\ & i \text{ and } f(m + \Delta x, n + \Delta y) = j \\ 0 & \text{if other else} \end{cases} \quad (5)$$

The steps to obtain the value of texture features are initiated by the transformation of the image into grayscale, followed by the creation of a co-occurrence matrix, as well as the formation and normalization of the symmetrical matrix. Each element $p(i, j)$ represents the sum of the frequency of occurrence of that pixel in i and j , and this can be denoted as $p(i, j|d, \theta)$. The symbol d is the distance between pixel i and j while θ symbolizes the orientation. GLCM is calculated with four different orientations at 0° , 45° , 90° , and 135° to produce four texture features as follows:

1) Contrast:

Contrast is a feature that represents the different levels of color or grayscale that appears on an image. Contrast will be 0 if the neighboring pixels have the same value.

$$\sum_i \sum_j (i - j)^2 p(i, j) \quad (6)$$

2) Correlation:

Correlation represents the linear relationship of the degree of gray image and has a range of values

between -1 to 1. μ_i represent mean probability occurrence of each row i in matrices and μ_j represents mean probability occurrence of each column j . σ_i represent GLCM variance in row/horizontally and σ_j is variance for column/vertically in matrices.

$$\sum_i \sum_j \frac{(i - \mu_i)(j - \mu_j)p(i, j)}{\sigma_i \sigma_j} \quad (7)$$

Where

$$\mu_i = \sum_i \sum_j i p(i, j)$$

$$\mu_j = \sum_i \sum_j j p(i, j)$$

$$\sigma_i = \sqrt{\sum_i \sum_j (i - \mu_i)^2 p(i, j)}$$

$$\sigma_j = \sqrt{\sum_i \sum_j (j - \mu_j)^2 p(i, j)}$$

3) Energy:

Energy represents a measure of uniformity in images. The higher image will be the higher value of energy.

$$\sum_i \sum_j p(i, j)^2 \quad (8)$$

4) Homogeneity:

Homogeneity represents a measure of similarity. Homogeneity will be of high value if all pixels have uniform values. This also often referred to as "Inverse Difference Moment" which means the inversion of contrast. This affects the calculation of the weight of the increased element will result in increased diagonal elements.

$$\sum_i \sum_j \frac{p(i, j)}{1 + |i - j|} \quad (9)$$

3.3.2. Morphological feature

The information related to the shape and geometric features of the umbilical cord object is extracted from the normal, hypercoiling, hypocoiling categories of the fetal umbilical cord. The shape features used include:

a) *Perimeter*,

This represents the size of the circumference or edge length of an object. In fetal umbilical objects, the hypercoiling class visually consists of more coiling or rotation with greater perimeter value.

b) *Convexity*

This feature is a comparison between the circumference of the convex hull produced for the umbilical cord object and its perimeter shown in Eq. (10).

$$Convexity = \frac{Convex_Perimeter}{Perimeter_Object} \quad (10)$$

c) *Solidity*

This feature is a comparison between the area of the convex hull and the outside as shown in Eq. (11).

$$Solidity = \frac{Convex_Area}{Area_Object} \quad (11)$$

d) *Bounding Rectangle*

Bounding Rectangle is a form of a minimum rectangle with the ability to accommodate all objects of the umbilical cord. From this rectangle measurement, two features are generated, namely the ratio of the object and the bound[36].

$$Extent = \frac{Area}{Bounding_Rectangle} \quad (12)$$

Furthermore, the difference between the Rectangle Area and the object is calculated as follows.

$$Diff\ Area = Area - Area\ rectangle \quad (13)$$

3.3.3. Synthetic minority oversampling technique (SMOTE)

The SMOTE algorithm was first carried out by [37] with oversampling and undersampling in the minority and majority classes, respectively. This was achieved by using several samples from the class and making a synthetic data along the lines from the k points of nearest minority data. Oversampling is the process of adding new data to a class by resampling

the minority, while undersampling reduces the data in a class until there is a balance. However, with the SMOTE oversampling method approach, the amount of data in the minority class is added to the desired ratio. Therefore, the number of k-nearest neighbors randomly chosen and commonly used is 5. The synthetic samples are created by calculating the difference in distance between the selected feature vectors and their closest neighbors. Furthermore, multiplication is performed with random numbers between 0 and 1 before it is added to the previously selected feature vector. The flow chart of the SMOTE algorithm is shown in Fig. 4. This method creates a new data point along the interpolation line between the sample data point and its nearest neighbor. In addition, the shortest distance is calculated from a sample data point using the euclidean distance measurement method.

This is followed by the process of generating synthetic data on the interpolation line between the sample point and k *nearsneighbors* with the following Eq. (14).

$$x_{new} = \tilde{P} + rand(0,1) \times (\tilde{T} - \tilde{P}) \quad (14)$$

where x_{new} is a new feature vector, \tilde{P} is a sample point and \tilde{T} is a data point from one of the nearest neighbors. The next step is to check when new data is successfully created, then whether the total amount equals the amount multiplied by the sampling rate. However, assuming the iteration is complete, it means that the amount of new data is in accordance with the sampling rate, else a random selection of new sample data is made in the minority class and re-calculated till the amount of data is equal to the specified sampling rate.

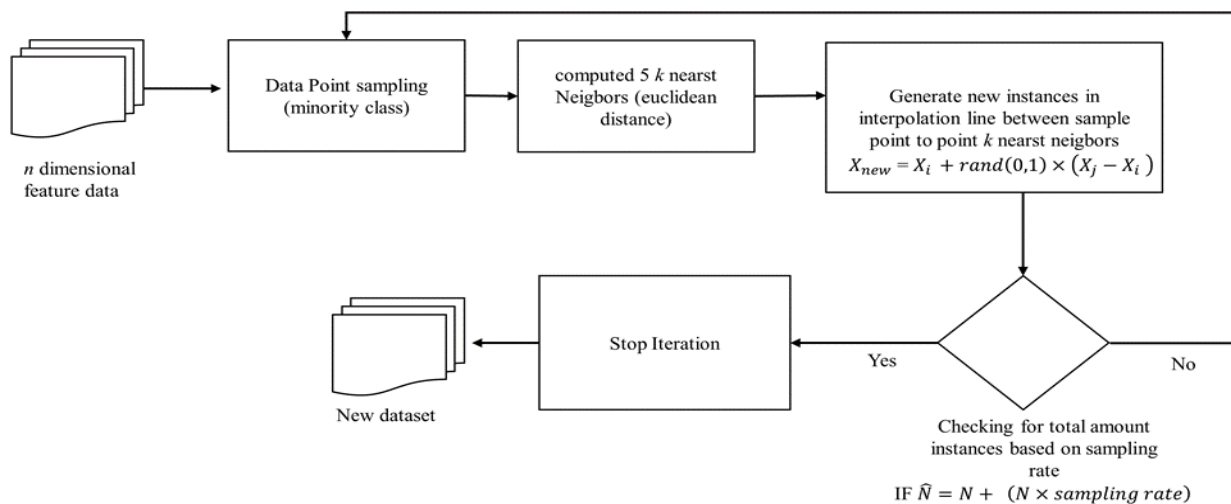


Figure. 4 SMOTE algorithm workflow diagram

3.3.4. Multiclassifier voting

The Multiclassifier voting classification method is included in the ensemble learning category. The use of several classifiers tends to accurately predict the label output more accurately. The proposed model uses the majority voting method which means that each classifier has the same weight value in determining the final decision and output label based on input data. Fig. 5 shows a schema of multiclassifiers with five different types of traditional classifiers. This study uses a classification method with learning characteristics different from one another. It is carried out to determine the prediction results based on various methods and expected to provide output class labels with good accuracy. This study applies 5 types of classification methods namely Support Vector Machine (SVM), Decision Tree, Random Forest (RF), Naive Bayes, and Multilayer Neural Network based on [31-32, 34, 38]. Each classifier has the same weight used in predicting the instance test class while each class also has a weight according to the selected classifier weights. The class with the highest resultant weight was expected to be proposed for the instance as determined by the ensemble. This, therefore, means an unlabeled instance was classified according to the class with the highest weight of vote as shown in the Eq. (15).

$$\text{class}(x) = \underset{c_i \in \text{dom}(y)}{\text{argmax}} \sum_k g(y_k(x), c_i) \quad (15)$$

where y_k represents the classification of k^{th} classifier and $g(y, c)$ denotes the indicator function defined as Eq. (16).

$$g(y, c) = \begin{cases} 1, & y = c \\ 0, & y \neq c \end{cases} \quad (16)$$

4. Evaluation matrix

In this research, the umbilical cord image data is categorized into three types namely Normal, Hypocoiling, and Hypercoiling, therefore a multi-class confusion matrix is needed. This matrix consists of more than 2 classes or $n \geq 2$, therefore the measure is equal to $n \times n$. The standard form of a multi-class confusion matrix is shown in Table 1.

Evaluation conducted by calculates the accuracy, precision, and recall values of each model. X_{ij} is data in the target of i (input) and j (output) class with values of 1, 2, and 3. The diagonal in Table 1 shows the amount of data in j class output that is correctly classified according to the target of i in the proposed model. The accuracy in the proposed model is based on the multi-class confusion matrix by comparing the amount of data available on the main diagonal with the total data used. Number of classes = n where $i, j = 1, 2, \dots, n$, therefore, the accuracy value is formulated in the following Eq. (17).

$$\text{Accuracy} = \frac{\sum_{i=1}^n X_{ii}}{\sum \sum X_{ij}} \times 100\% \quad (17)$$

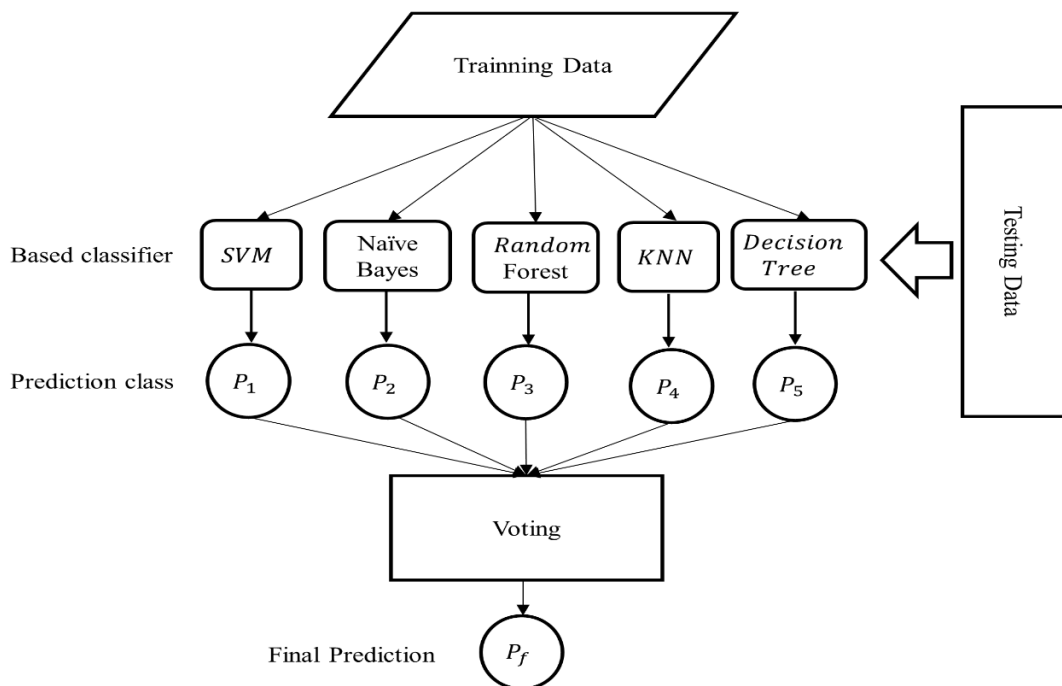


Figure. 5 Multiclassifier model

Table 1. Precision and recall matrix calculation

Target Class	Prediction Class			
	C1	C2	C3	Precision
C1	X_{11}	X_{12}	X_{13}	$\sum_{j=1}^n X_{1j} \frac{X_{11}}{\sum_{j=1}^n X_{1j}}$
C2	X_{21}	X_{22}	X_{23}	$\sum_{j=1}^n X_{2j} \frac{X_{22}}{\sum_{j=1}^n X_{2j}}$
C3	X_{31}	X_{32}	X_{33}	$\sum_{j=1}^n X_{3j} \frac{X_{33}}{\sum_{j=1}^n X_{3j}}$
Recall	$\sum_{i=1}^n X_{i1}$	$\sum_{i=1}^n X_{i2}$	$\sum_{i=1}^n X_{i3}$	
	$\frac{X_{11}}{\sum_{i=1}^n X_{i1}}$	$\frac{X_{22}}{\sum_{i=1}^n X_{i2}}$	$\frac{X_{33}}{\sum_{i=1}^n X_{i3}}$	

This is followed by visualizing the model performance created using Receiving Operating Characteristics (ROC). The graphical plot of a ROC curve is made up by plotting the false positive rate (FP) on the x-axis and true positive rate (TP) on the y-axis. Each threshold value forms a pair of measurements of FP versus TP.

5. Experimental result and discussion

5.1 Datasets

The umbilical cord dataset consists of 19 features namely contrast_gray, correlation_gray, energy_gray, homogeneity_gray, contrast_red, correlation_red, energy_red, homogeneity_red, contrast_green, correlation_green, energy_green, homogeneity_green, contrast_blue, correlation_blue, energy_blue, homogeneity_blue, convexity, solidity, extent. The imbalanced ratio (IR) is 6.3% with a total data amount of 63 as shown in Table 2. In this study, oversampling data uses the SMOTE method with the concept of one versus all in a multiclass classification where positive and negative classes are minority and majority data, respectively.

Table 2. Descriptive information of dataset

Dataset	Attr	IR Ratio	Hyper coiling	Hypo coiling	Normo coiling
Before SMOTE	19	14.5	37	21	4
After SMOTE	19	0	37	37	37

Table 3. Parameters value for classifier

Algorithm	Parameters
SVM	Kernel = polynomial, gamma = auto
Decision Tree	Criterion = gini, max_depth = None
Random Forest	n_estimators = 10, criterion = 'entropy', random_state = 0
KNN	n_neighbors = 5, metric = minkowski,
Naïve Bayes	Priors = none
Ensemble Multiclassifier	Voting = hard, weight = none, based classifier = logistic regression, random forest, naïve bayes, SVM, KNN, decision tree

The parameters in the SMOTE method are the number of k neighbors used which equals 5 and the sampling rate of 100%. The results of this oversampling data make the imbalanced ratio become 1 with 37 data in each class.

5.2 Experimental procedure

The classification model is carried out by comparing the traditional single classifier and ensemble multi-classifier towards the original dataset and after oversampling (SMOTE). The parameters of each classifier are as shown in Table 3 with a sampling rate parameter of 100% and k = 5.

The second experiment is measuring the effects of selected features on classification performance. This testing phase is to prove the proposed features can provide good performance on classification.

5.3 Result and discussion

This section presents an evaluation of umbilical cord classification performance to obtain the performance result of the proposed model is carried out by analyzing the accuracy, precision, and recall results of each class. Each classifier is tested based on a dataset before and after oversampling using the SMOTE method in order to determine the various performance. The testing of each model measures its performance towards relatively small datasets and imbalanced conditions. Comparison results of classification performance in accuracy and ROC parameter are shown in Table 4.

The results show without the SMOTE method in the data preprocessing stage, individual classifier such as Decision tree, Naïve bayes, SVM, KNN, and Random forest relatively failed to achieve satisfactory results on the testing data. Subsequent results show that an increase in accuracy and ROC performance occurs when the SMOTE method is used in the original data training. Combination of Multiclassifier voting and SMOTE method can achieve

Table 4. Performance on accuracy and ROC

	Accuracy (%)	ROC
SVM	68.7	0.50
KNN	68.7	0.50
Decision tree (CART)	56.2	0.50
Random Forest	56.2	0.46
Naïve Bayes	50.0	0.39
Ensemble Multiclassifier	73.0	0.48
SVM + SMOTE	71.4	0.79
KNN + SMOTE	75.4	0.91
Decision tree (CART) + SMOTE	64.0	0.70
Random Forest + SMOTE	79.2	0.86
Naïve Bayes + SMOTE	64.2	0.80
Ensemble Multiclassifier + SMOTE (Proposed Method)	81.4	0.89

Table 5. Performance on precision

	Prc Hyper (%)	Prc Hypo (%)	Prc Normo (%)	AVG 'recision (%)
SVM	100	67.0	0	67.2
KNN	57.0	78.0	0	63.0
Decision tree (CART)	43.7	67.1	0	52.0
Random Forest	50.0	71.0	0	57.0
Naïve Bayes	50.3	67.0	0	54.0
Ensemble Multiclassifier	79.2	74.0	0	73.0
SVM + SMOTE	80.0	100	61.4	80.1
KNN + SMOTE	60.2	100	79.0	83.4
Decision tree (CART) + SMOTE	67.2	50.1	83.5	70.0
Random Forest + SMOTE	84.0	74.0	79.0	79.0
Naïve Bayes + SMOTE	50.5	58.1	80.0	65.4
Ensemble Multiclassifier + SMOTE (Proposed Method)	84.2	74.0	82.0	81.5

Table 6. Performance on recall

	Rcl Hyper (%)	Rcl Hypo (%)	Rcl Normo (%)	AVG Recall (%)
SVM	25.3	100	0	69.5
KNN	100	70.0	0	69.0
Decision tree (CART)	75.2	60.1	0	56.7
Random Forest	100	50.0	0	56.3
Naïve Bayes	50.0	60.5	0	50.0
Ensemble Multiclassifier	68.0	84.0	0	72.0
SVM + SMOTE	67.0	45.0	100	71.4
KNN + SMOTE	100	36.0	100	75.3
Decision tree (CART) + SMOTE	73.8	83.0	45.0	64.0
Random Forest + SMOTE	86.0	68.0	84.0	79.3
Naïve Bayes + SMOTE	50.0	64.2	73.0	64.1
Ensemble Multiclassifier + SMOTE (Proposed Method)	84.2	70.0	86.4	80.0

best result in accuracy of 81.4% and ROC of 0.89. This proves that the imbalanced data condition and small dataset needs an appropriate modification model. Moreover, after data oversampling, the minority classes started to be recognized as indicated by increased precision and recall values for each class that shows in Table 5 and Table 6.

According to the measurements in Table 5 and Table 6, the precision results of each classifier in the Normocoiling class show a significant increase performance. The combination of multiclassifier and SMOTE method showed satisfactory results where the precision value in each class was balanced with 84.2% of Hypercoiling, 74% of Hypocoiling, and 82% of Normocoiling. The recall value also shows a relatively good increase in value. Combination Multiclassifier and SMOTE achieved the most balanced performance for each class compared to other models with 84.2% of Hypercoiling, 70% of Hypocoiling, and 86.4% of Normocoiling. The results achieved indicate that the SMOTE method is able to improve the ability to generalize data from

Table 5. Classification performance of multiclassifier method by combination features

No	Features	Number of features	Average Accuracy (%)	Average Recall (%)	Average-F Measure (%)	Average Precision (%)	ROC Area
1	GLCM Grayscale	4	69.1	68.3	68.0	67.9	0.78
2	GLCM RGB	12	76.2	75.0	74.6	74.6	0.83
3	GLCM Red Channel	4	68.0	68.2	69.0	69.3	0.76
4	GLCM Green Channel	4	71.4	70.3	70.0	69.6	0.79
5	GLCM Blue Channel	4	66.0	65.6	65.6	66.0	0.76
6	GLCM Grayscale + RGB	16	78.9	74.6	74.6	75.0	0.83
7	Shape Eccentricity, Perimeter, Extent	3	75.0	74.6	75.0	75.3	0.81
8	Texture + Shape (Proposed Method) GLCM Grayscale + GLCM RGB + Eccentricity + Perimeter + Extent	19	81.4	80.0	80.0	79.3	0.89

each classifier and with a combination of multiclassifier voting the results achieved can be further increased. Furthermore, the comparison results of umbilical cord based on the selected features is shown in Table 5. Based on the global performance in Table 5, it can be observed that combining feature types produces higher accuracy than using only one type of feature. GLCM in the grayscale image can achieve 69.1% of accuracy and GLCM in RGB color space can achieve 76.2% of accuracy. When we combine GLCM in grayscale and RGB features, accuracy improves to 78.9%. The morphological features produce a relatively stable performance in accuracy, recall, precision, F-measure, and ROC area. The combination of eccentricity, perimeter, and extent of umbilical cord yields an accuracy 75%, average recall 74.6%, average precision 75.3%, F-measure 75%, and ROC area 0.81. Finally, the combination of texture (GLCM) and shape features achieve the best result performance with an average improvement of 2.5% of accuracy, 5.0 % of recall, 5.0 % of F-measure, 4.0 % of precision, 0.069 of ROC area. Thus, by combine shape and texture contain important information that contributes to the improvement of classification accuracy.

Finally, to validate the result of proposed model is also conducted by analyzing the results of graph based on the ROC curve. ROC value measurements are carried out in each class to determine the performance of the model. Fig. 6 show the ROC curve performance of each class of multiclassifier without SMOTE method. It can be seen that the performance of the classification method fails to

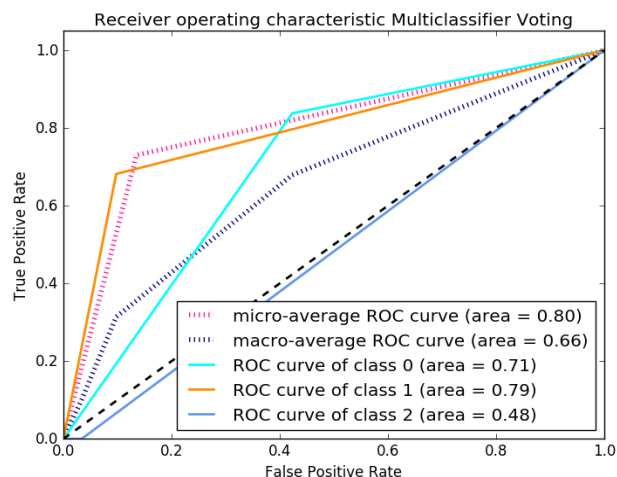


Figure. 6 ROC curve of multiclassifier voting without SMOTE

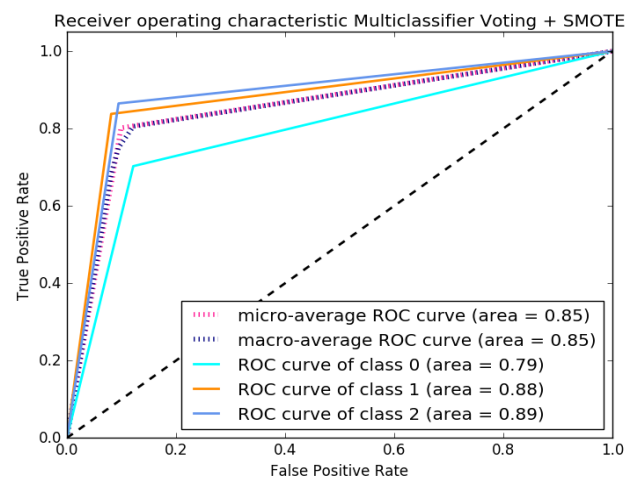


Figure. 7 ROC curve of multiclassifier voting combine with SMOTE

recognize normocoiling classes which are only able to reach 0.48 of ROC. Furthermore, when the multiclassifier was combined with the SMOTE method, there was an increase in the performance of the classification model with the ROC value in the normocoiling class reaching 0.89 that shown in Fig. 7.

6. Conclusion

This study started by extracting features based on texture and shape characteristics represented in vector form. This research, specifically, focused on obtaining features of umbilical cord and its classification in imbalanced data conditions. The performance of the traditional classifier in the experimental section shows unsatisfactory results. This indicates that the classifier has difficulty determining decision boundaries because of the small and imbalanced amount of data. In addition, the distribution of data points in the dataset produced uneven and overlapping characteristics. This shows that the dataset has low variance and standard deviation, thereby, leading to unsatisfactory performance. To address these issues, we proposed a classification model through the combination of the SMOTE data oversampling and multiclassifier voting method. Experimental results show a significant increase in classification performance combines with SMOTE method.

Based on the result of this study, we can conclude that using combination shape and texture GLCM feature types yields increased accuracy. Measurement parameters can be seen in the value of accuracy, precision, and recall that increases in predicting each class. The proposed model achieves the best result with 81.4% of accuracy, 80% of recall, 80% of F-measure, 79.3% of precision, and 0.89 of ROC area. These results showed that the proposed model is better compared to other single classification methods in dealing with imbalanced data conditions. For future work, SMOTE method will be modified in order to minimize overlapping data between classes and eliminate noise, thereby, achieving better distribution results, with improved classification performance. Therefore, implemented features that are in line with medical science is also used to represent the class of umbilical cords.

Conflicts of Interest

The authors declare no conflict of interest

Author Contributions

Conceptualization, Angga, Retantyo, Aina; methodology, Angga, Retantyo, Aina; software, Angga; validation, Angga, Retantyo, Aina, and Hariyasa; formal analysis, Hariyasa; resources data, Hariyasa; writing—original draft preparation, Angga, Retantyo, Aina, and Hariyasa; writing—review and editing, Angga, Retantyo and Aina; supervision, Retantyo, Aina, and Hariyasa; funding acquisition, Retantyo”,

Acknowledgments

The authors would like thank to the Research Directorate of Universitas Gadjah Mada in the Rekognisi Tugas Akhir (RTA) 2020 scheme.

References

- [1] S. Gupta, M. Faridi, and J. Krishnan, “Umbilical coiling index”, *Journal Obstetric Gynecology of India*, Vol. 56, No. 4, pp. 315–319, 2006.
- [2] L. Raio, F. Ghezzi, E. D. Naro, R. Gomez, M. Franchi, M. Mazor, & H. Brühwiler, “Sonographic measurement of the umbilical cord and fetal anthropometric parameters”, *European Journal of Obstetrics and Gynecology and Reproductive Biology*, Vol. 83, No. 2, pp. 131–135, 1999.
- [3] T. H. Strong, D. L. Jarles, and J. S. Vega, “The umbilical coiling index”, *American Journal of Obstetrics and Gynecology*, Vol. 170, No. 1, pp. 29–32, 1994.
- [4] Y. Qin, T. K. Lau, and M. S. Rogers, “Second-trimester ultrasonographic assessment of the umbilical coiling index”, *Ultrasound in Obstetrics and Gynecology*, Vol. 20, No. 5, pp. 458–463, 2002.
- [5] Y. Ohno, M. Terauchi, and K. Tamakoshi, “Perinatal outcomes of abnormal umbilical coiling according to a modified umbilical coiling index”, *Obstetrics and Gynaecology Research.*, Vol. 42, No. 11, pp. 1–7, 2016.
- [6] M. Predanic, “Sonographic assessment of the umbilical cord”, *The International Journal of Continuing Education and Current Awareness*, Vol. 5, No. 2, pp. 105–110, 2005.
- [7] J. Rana, G. Ebert, and K. Kappy, “Adverse perinatal outcome in patients with an abnormal umbilical coiling index”, *Obstetrics and Gynecology*, vol. 85, no. 4, pp. 573–577, 1995.
- [8] N. Nnamoko and I. Korkontzelos, “Efficient Treatment of Outliers and Class Imbalance for Diabetes Prediction”, *Artificial Intelligence in Medicine*, Vol. 104, No. 101815, 2020.

- [9] S. Beura, B. Majhi, and R. Dash, "Neurocomputing Mammogram classification using two dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer", *Neurocomputing*, Vol. 154, pp. 1–14, 2015.
- [10] F. Mohanty, S. Rup, B. Dash, B. Majhi, and M. N. S. Swamy, "Digital mammogram classification using 2D-BDWT and GLCM features with FOA-based feature selection approach", *Neural Computing and Applications*, Vol. 32, pp. 7029–7043, 2019.
- [11] T. Prabhakar and S. Poonguzhali, "Automatic Detection and Classification of Benign and Malignant Lesions in Breast Ultrasound Images using Texture Morphological and Fractal Features", In: *Proc. of Biomedical Engineering International Conf. (BMEiCON)*, pp. 1–5, 2017.
- [12] M. Sahrim, U. S. A. Rahman, W. Z. W. Ismail, I. Ismail, J. Jamaludin, and S. R. Balakrishnan, "Automated Feature Description of Follicle Size in Assisted Reproductive Treatment", *International Journal of Integrated Engineering*, Vol. 10, No. 1, pp. 182–185, 2018.
- [13] X. Wang, P., Wang, L., Li, Y., Song, Q., Lv, S., & Hu, "Biomedical Signal Processing and Control Automatic cell nuclei segmentation and classification of cervical Pap smear images", *Biomedical Signal Processing and Control*, Vol. 48, pp. 93–103, 2019.
- [14] G. A. Pradipta and P. D. Wulaning Ayu, "Fetal weight prediction based on ultrasound image using fuzzy C means clustering and Iterative Random Hough Transform", In: *Proc. of 1st Int. Conf. Informatics Computational Science, ICICoS 2017*, Semarang, pp. 71–76, 2018.
- [15] A. Ali, S. M. Shamsuddin, and A. L. Ralescu, "Classification with class imbalance problem: A review", *International Journal of Advances in Soft Computing and Its Applications*, Vol. 7, No. 3, pp. 176–204, 2015.
- [16] M. Galar, A. Fern, E. Barrenechea, and H. Bustince, "Hybrid-Based Approaches", *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, Vol. 42, No. 4, pp. 463–484, 2012.
- [17] A. Fern and S. Garc, "SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary", *Journal of Artificial Intelligence Research*, Vol. 61, pp. 863–905, 2018.
- [18] S. Reza and J. Ma, "Imbalanced Histopathological Breast Cancer Image Classification with Convolutional Neural Network", In: *Proc. of 14th IEEE International Conf. on Signal Processing (ICSP)*, pp. 619–624, 2018.
- [19] K. Wang, B. Makond, and K. Wang, "An improved survivability prognosis of breast cancer by using sampling and feature selection technique to solve imbalanced patient classification data", *BMC Medical Informatics and Decision Making*, Vol. 13, No. 1, p. 1, 2013.
- [20] R. Liu, L. O. Hall, K. W. Bowyer, D. B. Goldgof, R. Gatenby, and K. B. Ahmed, "Synthetic Minority Image Over-sampling Technique: How to Improve AUC for Glioblastoma Patient Survival Prediction", In: *Proc. of IEEE International Conf. on Systems, Man, and Cybernetics (SMC)*, pp. 1357-1362, 2017.
- [21] L. B. Maia, A. Lima, R. M. Pinheiro Pereira, G. B. Junior, J. Dallyson Sousa De Almeida, and A. C. De Paiva, "Evaluation of Melanoma Diagnosis using Imbalanced Learning", In: *Proc. of International Conf. on Systems, Signals, and Image Processing*, Vol. 18, No. 1, 2018.
- [22] P. K. Chan and S.J. Stolfo, "Toward Scalable Learning with Non-uniform Class and Cost Distributions: A Case Study in Credit Card Fraud Detection", In: *Proc. of the Fourth International Conference on Knowledge Discovery and Data Mining*, pp. 164-168, 1998.
- [23] H. Cao, X. Li, D. Y. Woon, and S. Ng, "Integrated Oversampling for Imbalanced Time Series Classification", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 25, No. 12, pp. 2809–2822, 2013.
- [24] W. Choe, O. K. Ersoy, and M. Bina, "Neural network schemes for detecting rare events in human genomic DNA", *Bioinformatics*, Vol. 16, pp. 1062–1072, 2000.
- [25] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique", *Journal of Artificial Intelligence Research*, Vol. 16, pp. 321–357, 2002.
- [26] K. J. Wang, B. Makond, K. H. Chen, and K. M. Wang, "A hybrid classifier combining SMOTE with PSO to estimate 5-year survivability of breast cancer patients", *Applied Soft Computing Journal*, Vol. 20, pp. 15–24, 2014.
- [27] S. F. Abdoh, M. Abo Rizka, and F. A. Maghraby, "Cervical cancer diagnosis using random forest classifier with SMOTE and feature reduction techniques", *IEEE Access*, Vol. 6, pp. 59475–59485, 2018.
- [28] S. W. Purnami, P. M. Khasanah, S. H. Sumartini, V. Chosuvivatwong, and H. Sriplung, "Cervical cancer survival prediction using hybrid of SMOTE, CART and smooth support vector

- machine”, In: *Proc. of AIP Conf. Proceedings*, Vol. 1723, No. 1, 2016.
- [29] S. Fotouhi, S. Asadi, and M. W. Kattan, “A comprehensive data level analysis for cancer diagnosis on imbalanced data”, *Journal of Biomedical Informatics*, Vol. 90, p. 103089, 2019.
- [30] K. J. Wang, A. M. Adrian, K. H. Chen, and K. M. Wang, “A hybrid classifier combining Borderline-SMOTE with AIRS algorithm for estimating brain metastasis from lung cancer: A case study in Taiwan”, *Computer Methods and Programs in Biomedicine*, Vol. 119, No. 2, pp. 63–76, 2015.
- [31] Mehmet Akif Yaman, A. Subasi, and F. Rattay, “ Comparison of random subspace and voting ensemble machine learning methods for face recognition”, *Symmetry*, , Vol. 10, No. 11, p. 651, 2018.
- [32] O. Petinrin, F. Saeed and T. Al-Hadhrami., “Voting-Based Ensemble Method for Prediction of Bioactive Molecules”, In: *Proc. of 2nd International Conf. on Knowledge Engineering and Applications (ICKEA). IEEE*, pp. 118–122, 2017.
- [33] M. R. Smith and T. Martinez, “The robustness of majority voting compared to filtering misclassified instances in supervised classification tasks”, *Artificial Intelligence Review*, Vol. 49, No. 1, pp. 105–130, 2018.
- [34] A. Onan, S. Korukoglu, and H. Bulut, “A Multiobjective Weighted Voting Ensemble Classifier Based on Differential Evolution Algorithm for Text Sentiment Classification”, *Expert Systems With Applications*, Vol. 62, pp. 1–16, 2016.
- [35] V. K. Sudarshan, E. Y. K. Ng, U. R. Acharya, S. M. Chou, R. S. Tan, and D. N. Ghista, “Computer-aided diagnosis of Myocardial Infarction using ultrasound images with DWT , GLCM and HOS methods: A comparative study”, *Computers in Biology and Medicine*, Vol. 62, No. 2015, pp. 86–93, 2015.
- [36] Y. L. Huang, D. R. Chen, Y. R. Jiang, S. J. Kuo, H. K. Wu, and W. K. Moon, “Computer-aided diagnosis using morphological features for classifying breast lesions on ultrasound”, *Ultrasound in Obstetrics and Gynecology*, Vol. 407, No. March, pp. 565–572, 2008.
- [37] N. V Chawla, K. W. Bowyer, L. O. Hall, Kegelmeyer, and W. Philip, “SMOTE: Synthetic Minority Over-sampling Technique”, *Journal of Artificial Intelligence Research*, Vol. 16, pp. 321–357, 2002.
- [38] S. Bashir, U. Qamar, and F. Hassan, “BagMOOV: A novel ensemble for heart disease prediction bootstrap aggregation with multi-objective optimized voting”, *Australasian Physical and Engineering Sciences in Medicine*, Vol. 38, No. 2, pp. 305–323, 2015.