TELKOMNIKA, Vol.17, No.2, April 2019, pp.645~652
ISSN: 1693-6930, accredited First Grade by Kemenristekdikti, Decree No: 21/E/KPT/2018
DOI: 10.12928/TELKOMNIKA.v17i2.9017

Classification of blast cell type on acute myeloid leukemia (AML) based on image morphology of white blood cells

Wiharto Wiharto*, Esti Suryani, Yuda Rizki Putra
Department of Informatics, Universitas Sebelas Maret, Indonesia
*Corresponding author, e-mail: wiharto@staff.uns.ac.id

Abstract

AML is one type of cancer of the blood and spinal cord. AML has a number of subtypes including M0 and M1. Both subtypes are distinguished by the dominant blast cell type in the WBC, the myeloblast cells, promyelocyte, and myelocyte. This makes the diagnosis process of leukemia subtype requires identification of blast cells in WBC. Automatic blast cell identification is widely developed but is constrained by the lack of data availability, and uneven distribution for each type of blast cell, resulting in problems of data imbalance. This makes the system developed has poor performance. This study aims to classify blast cell types in WBC identified AML-M0 and AML-M1. The method used is divided into two stages, first pre-processing, image segmentation and feature extraction. The second stage, perform resample, which is continued over sampling with SMOTE. The process is done until the amount of data obtained is relatively the same for each blast cell, then the process of elimination of data duplication, randomize, classification and performance measurement. The validation method used is k-fold cross-validation with k=10. Performance parameters used are sensitivity, specifyicity, accuracy, and AUC. The average performance resulting from classification of cell types in AML with Random Forest algorithm obtained 82.9% sensitivity, 92.1% specificity, 89.6% accuracy and 87.5% AUC. These results indicate a significant improvement compared to the system model without using SMOTE. The performance generated by reference to the AUC value, the proposed system model belongs to either category, so it can be used for further stages of leukemia subtype AML-M0 and AML-M1.

Keywords*: acute myeloid leukemia, blast cell, oversampling, segmentation, SMOTE, white blood cell*

1. Introduction

Leukemia is a disease of blood and bone marrow cancer. Bone marrow is a spongy tissue in the bone where blood cells are made. Cancerous blood cells will damage blood cells in the bone marrow [1]. Leukemia has several types, namely chronic and acute leukemia. Types of acute leukemia include Acute Lymphoblastic Leukemia (ALL) and Acute Myeloid Leukemia (AML) [2]. AML type leukemia, referring to the French-American-British classification, AML is classified into 8 subtypes including M0, M1, and M2 [3]. AML leukemia is caused by the differentiation of myeloid series cells stopping in blast cells which results in a buildup of the blast in the spinal cord. ALL or AML type leukemia diagnosis has been used to calculate the complete blood cell count. This approach requires relatively expensive energy, time and cost [4]. An alternative that can be done to overcome these problems is using a blood cell image processing approach [2, 5, 6]. The use of blood cell image makes identification process in order to diagnosis can be done by the computerization process. A number of studies on image processing for the diagnosis of leukemia have been carried out. First focused on detecting positive or negative leukemia ALL [7–10], second leukemia ALL or AML [11–13], third is AML subtype.

A number of studies that have used a blood cell image processing approach to the diagnosis of ALL type leukemia are carried out by Devi et al. [14] and Selvaraj et al. [15]. The system model which proposed Devi et al. [14], is divided into several stages: pre-processing, segmentation using otsu thresholding, feature extraction with histogram of oriented gradient(HOG), and classification using adaptive fuzzy inference system. The diagnosis of leukemia based on a fuzzy inference system is also done by Khosrosereshki et al. [16] but uses the mamdani method. Selvaraj et al. [15], using a feature somewhat different from

Received January 21, 2018; Revised November 15, 2018; Accepted December 17, 2018

that of Devi et al. [14]. The features is divided into two groups, namely shape features and densitometric. This features in both groups then used to make conclusions using naive bayesian classification algorithms. The next study was the diagnosis of leukemia subtype AML M2 and M3 by Suryani et al. [17]. The study features extraction done preceded by the segmentation process. The segmentation process uses a watershed distance transform. The extraction feature results in white blood cell (WBC) area, WBC perimeter, WBC roundness, nucleus ratio, WBC mean, and WBC standard deviation. The final conclusion is to determine the AML subtype using the neural network. A similar diagnosis concept is also performed by Harjoko et al. [18], ie the classification of subtypes AML M1, M2 and M3, with feature extraction process preceded by active contour operation.

Most of the research that has been done, especially for the diagnosis of AML subtype leukemia uses stages that are not commonly used by clinicians. This makes it difficult for clinicians to understand each stage of the diagnostic process. Clinicians in diagnosing AML subtypes, predictably identifying the blast cell types present in the WBC. Referring to the type of blast cell contained in WBC that is as knowledge to be used to identify the subtype of AML. Research that has used this approach in diagnosing AML subtypes, is research conducted by Suryani et al. [19]. The study identified the AML M0 and M1 subtypes, with the stages of determining the dominant blast cell type in the WBC. Other studies were also conducted by Suryani et al. [20], but for the diagnosis of AML M1 and M2. Unfortunately, the two studies have weaknesses, namely the low performance of the results of blast cell type classification. The low performance of blast cell type classification, also resulted in low performance in the diagnosis system of AML M0 leukemia, AML M1 and AML M2. The poor performance of blast cell type classification was caused by the lack of availability of white blood cell image data samples, which were identified with leukemia AML M0 subtype, AML M1 and AML M2. These conditions resulted in the distribution of data for each type of blast cell resulting from the feature extraction process in the diagnosis system of leukemia, becoming unbalanced. Imbalances of data can lead to a decline in the performance of classification algorithms [21–25].

Referring to a number of studies that have been done, and with a number of advantages and disadvantages, this study will propose a model of blast cell classification on WBC identified leukemia subtype AML M0 and AML M1. Identification is done by considering the condition of data imbalance. The condition of the data imbalance is overcome by using a combination of resampling, Synthetic Minority Over-sampling Technique (SMOTE), and randomize. System performance is measured using the parameters of sensitivity, specificity, accuracy, and area under the curve (AUC) parameters.

2. Research Method

2.1. Data

This study used data obtained from Dr. Moewardi Hospital, Surakarta Indonesia on Clinical Pathology. The data consisted of 50 white blood cell images that were identified by AML. The data is distributed into the subtype of AML M0 as much 20 images and 30 AML M1 images. Image data using JPG format with size 1600x1200 pixels. Characteristics of blast cell types contained in the WBC for the subtype AML as shown in Table 1, while the images for each blast cell are shown in Figure 1.

Table 1. The Feature of blast cell [26]

Type of blast cell	Feature			
	Cell diameter	The ratio of nucleus: cytoplasm	Nucleus	Cytoplasm
Myeloblast	15 - 20 µm	7:1 - 5:1	Round, there is a nucleus core	Usually blue without granules
Promyelocyte	12 - 24 µm	5:1 - 3:1	Slightly curved	Blue with lots of granules
Myelocyte	10 - 18 µm	2:1 - 1:1	Round to oval, slightly curved	Appears a second granule (pink spots)

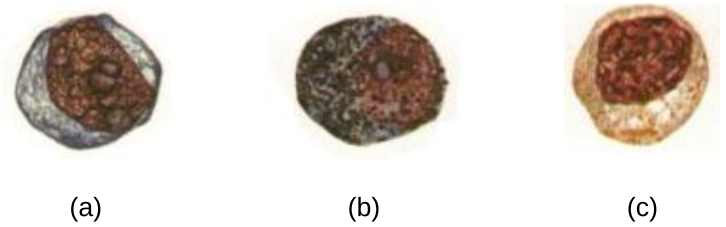

Figure 1. (a) Myeloblast, (b) Promyelocyte, (c) Myelocyte

2.2. Proposed Method

The system model for identification of blast cell types in the process of diagnosis of leukemia disease can be shown in Figure 2. The system model is divided into 4 parts, namely image processing, oversampling, classification and evaluation of system performance. Image processing, including pre-processing, segmentation [19, 27] and feature extraction. Feature extraction produces three features, namely WBC diameter, nucleus ratio, and nucleus roundness [19]. These three variables are a feature of blast cells in the WBC. In this study, blast cells were observed in 3 types of blast cells with, as shown in Table 1. The oversampling section includes resample process, SMOTE, deletion of redundant data, and randomize. The resampling process is done to retrieve the re-samples from the existing data, for the next SMOTE process. The results of the SMOTE process then performed the same data deletion and finished by the randomizing process.

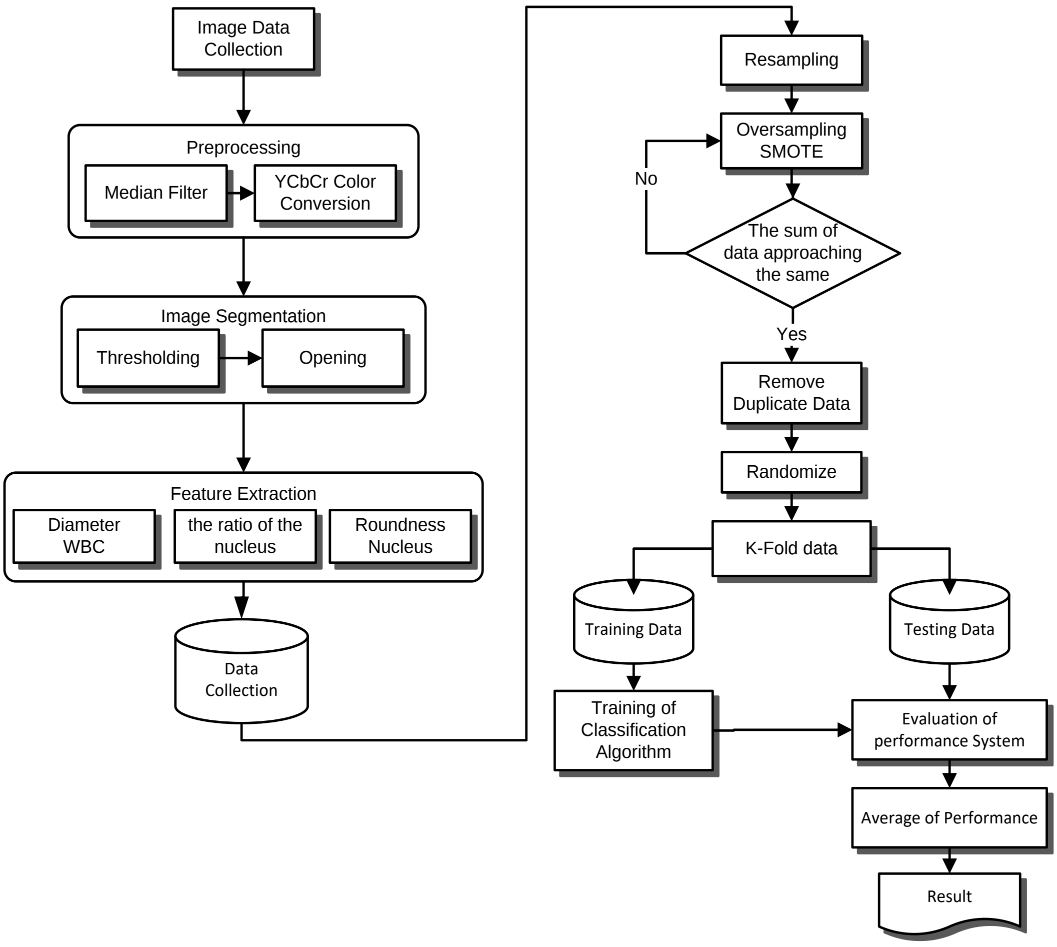

Figure 2. The proposed method

The third part, the process of classification using Random Forest algorithm, in addition to Random Forest also tested k-NN algorithm. The fourth part is performance evaluation. Performance evaluation is done by using 3-dimensional confusion matrix, as shown in Table 2. Referring to Table 2, it can be derived into a 2-dimensional confusion matrix. The descending process for myeloblast cell types can be shown in (1-4), using the same concept is also used for cell types promyelocyte and myelocyte. Referring to (1-4), it can be used to derive the equation of performance parameters. The performance parameters are the sensitivity, specificity, accuracy, and area under the curve (AUC). The system performance is validated by k-folds cross-validation method with value k=10. The method will divide the data into k-groups, with k-1 groups as training data, and 1 group for testing, and performed repeatedly so that all data groups have been used for testing.

Table 2. Confusion Matrics [20]

Actual	Classified		
	Myeloblast	Promyelocyte	Myelocyte
Myeloblast	A	B	C
Promyelocyte	F	E	D
Myelocyte	G	H	I

$$\mathrm{TP} = \mathrm{A} \tag{1}$$

$$\mathrm{TN} = \mathrm{E} + \mathrm{D} + \mathrm{H} + \mathrm{I} \tag{2}$$

$$\mathrm{FP} = \mathrm{F} + \mathrm{G} \tag{3}$$

$$\mathrm{FN} = \mathrm{B} + \mathrm{C} \tag{4}$$

2.3. Synthetic Minority Over-sampling Technique (SMOTE)

The unbalanced data can be solved by using some sampling techniques. One of the sampling techniques to overcome the imbalanced data is by using the method of Synthetic Minority Over-Sampling Technique (SMOTE) [28]. The SMOTE method over-samples minority classes by creating synthetic samples that operate more on feature space than in data space so that the data distribution of each class becomes balanced. The SMOTE technique creates a synthetic sample by exploring samples of existing minority classes with random samples obtained from k-nearest neighbors.

2.4. Classification Algorithms

Classification algorithms can be grouped into two by looking at the approaches they used, ie black-box and non-black-box [29]. In this research use both approaches, that is for black-box is using the k-NN algorithm, while for non-black-box use random forest (RF) algorithm. The random forest classification algorithm is an improvement of the CART classification algorithm. Improvements were made by applying the bootstrap aggregating (bagging) method and Random feature selections [30, 31]. In Random forest will use a number of decision trees, with each decision tree has been trained using a sample of data, and each attribute is broken into the selected tree between the subset attribute, which is random. The classification process is done by taking majority votes from the set of trees that are formed, for each tested data.

The k-Nearest Neighbor (k-NN) classification algorithm is a classification algorithm based on Euclidean distance [32]. The precision of the k-NN algorithm is determined by the presence or absence of irrelevant features, or if the feature weight is not equivalent to its relevance to the classification. Another factor that affects the performance of k-NN is the value of k used. The k value is too high it will decrease the noise effect on the classification process, but will cause the boundary between each class to be blurred. A good k value can be done by determining the optimum parameters, for example by using the feature selection method.

3. Results and Analysis

The blast cell classification model on the WBC identified subtype AML M0 and subtype AML M1, can be shown the results for each stage. The first stage, namely image processing. At this stage the image segmentation process in which the result of WBC image segmentation as shown in Figure 3. The second stage, with reference to the process of image segmentation process, then performed feature extraction process. The feature extraction process produces 3 attributes. The WBC diameter which in units of μm, the nucleus ratio and the nuclear roundabout. The WBC diameter attribute has a value that is much different from the other attributes, so it needs normalization to have the same attribute value equal to the other attribute. The normalization method used is Min-Max [33].

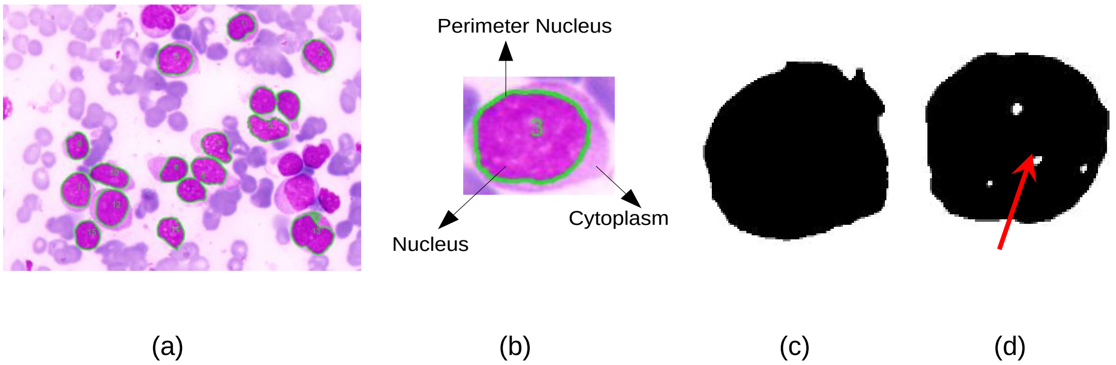

Figure 3. Image segmentation (a) segmentation WBC (b) segmentation cell (c) WBC (d) nucleuse

The WBC image used in this study amounts to 50, which is distributed to 20 AML M0 indefinable imagery and 30 AML M1 images. The 50 WBC image data from the feature extraction feature, with the WBC diameter feature, the nucleus ratios, and the nucleate roundabout obtained 165 blast cell data, as shown in Table 3, and the data was distributed into myeloblast blast cells 97, promyelocyte 31, and myelocyte 37. The distribution of the feature extraction data for each type of blast cell shows unbalanced data, for myeloblast cell types almost 3 times the number of promyelocyte and myelocyte cells. This indicates an imbalance in the distribution of data.

Table 3. The Data Sample of Result Feature Extraction

No	WBC Diameter (μm)	Nucleus Ratio	Nucleus Roundabout	Cell Type
1	16.037	0.669	0.513	Myelocyte
2	17.075	0.666	0.573	Promyelocyte
3	15.554	0.729	0.605	Myeloblast
4	18.814	0.679	0.564	Promyelocyte
5	16.468	0.748	0.602	Myeloblast
6	15.513	0.796	0.467	Myeloblast
7	15.513	0.558	0.625	Myeloblast
8	14.450	0.728	0.577	Promyelocyte
9	15.797	0.707	0.591	Promyelocyte
10	13.159	0.762	0.641	Myelocyte

The next step in the proposed system model is to perform resample, SMOTE and remove duplicate data. The results of these steps resulted in data of 244, distributed into 64 myeloblast cells, 84 promyelocyte cells, and 96 myelocyte cells. The results of the stages are then classified by the Random Forest (RF) classification algorithm, and k-NN [32] with the k-folds cross-validation validation method, where k=10. The test results are as shown in Table 4 and Table 5.

Table 4. The Performance of the Proposed System Model (RF)

Cell Type	Sensitivity		Specificity		Accuracy		AUC	
	RF	RF+SMOTE	RF	RF+SMOTE	RF	RF+SMOTE	RF	RF+SMOTE
Myeloblast	0.845	0.672	0.544	0.939	0.721	0.869	0.695	0.805
Promyelocyte	0.484	0.929	0.925	0.925	0.842	0.926	0.705	0.927
Myelocyte	0.405	0.885	0.906	0.899	0.794	0.893	0.656	0.892
Mean	0.578	0.829	0.792	0.921	0.786	0.896	0.685	0.875

Table 5. The Performance of the Proposed System Model (k-NN)

Cell Type	Sensitivity		Specificity		Accuracy		AUC	
	kNN	kNN+SMOTE	kNN	kNN+SMOTE	kNN	kNN+SMOTE	kNN	kNN+SMOTE
Myeloblast	0.722	0.609	0.500	0.933	0.630	0.848	0.611	0.771
Promyelocyte	0.323	0.917	0.866	0.900	0.764	0.906	0.594	0.908
Myelocyte	0.378	0.885	0.852	0.899	0.745	0.893	0.615	0.892
Mean	0.474	0.804	0.739	0.911	0.713	0.883	0.607	0.857

The model of the blast cell classification system in WBC identified leukemia AML M0 and AML M1 proposed to provide better performance. The performance is as shown in Table 4 and Table 5. In Table 4 and Table 5 it can be seen, when the distribution of data for each blast cell is unbalanced, giving an average performance of AUC in poor category, whereas when done the process of SMOTE gives the performance in good category (in the range 80-90%) [34]. In Table 4, particularly for the sensitivity performance parameters of the Random Forest classification algorithm without SMOTE, the classification of myeloblast blast cells is better than that of the other cells. This is caused by the amount of data myeloblast blast cells more than other cells (3xmore). The condition indicates an imbalance of data.

The proposed system model, as compared with previous research, as did Suryani et al. [19] suggests that the proposed system model is better. This is shown from the test of significance by using the t-test, with 95% confidence level. The results of the tests are shown in Table 6. In a study conducted by Suryani et al. [19], the classification algorithm used is k-NN. Differences in the use of classification algorithms without oversampling with SMOTE showed no significant differences, such as when k-NN compared with Random Forest, where the p-value>0.05. The performance of the proposed system model when replaced with the classification algorithm does not use Random Forest, also provides better performance, such as when using the k-NN algorithm, where the p-value is <0.05. It shows that the use of a combination of resampling, SMOTE and remove duplicate data, is able to provide better performance.

Further comparison with research conducted by Suryani et al. [20]. The study diagnosed AML M1 and AML M2 leukemia by using blast cell types present in AML M1 and M2 leukemia as parameters in making decisions. Blast cells used in the study were myeloblast, promyelocyte, myelocyte and metamyelocyte. The difference is because it is used to detect leukemia AML M2. The problem that occured in this study is imbalancing data, so that the poor performance in detecting blast cell types. When compared with the proposed blast cell type detection model, the proposed model has a much better performance, namely by showing the p-value <0.05. Complete comparison as shown in Table 6.

Table 6. The Comparison of Proposed System Models with Previous Research

Algorithms	Suryani et al. [19]	Suryani et al. [20]
	p-value	
RF	0.072089	0.884358
RF+SMOTE	0.011811	0.016583
k-NN+SMOTE	0.018726	0.027933

The proposed blast cell type classification model, capable of delivering performance in a good category, when referring to the AUC value. The performance is obtained by using 3 attributes that are a feature of each type of blast cell. The three attributes can also be analyzed to find out how much the influence of performance the proposed system model. How much influence can be seen using some feature selection filter type algorithms, such as information gain [35], the gain ratio [7, 36] and ReliefF [37]. The ranking can be shown in Table 7. Referring to Table 6, it can be shown that all algorithms give the same conclusion, that is by sequence WBC rank, Nucleus roundness and Nucleus ratio. These results indicate that the feature nucleus ratio gives the least effect compared to other features. The result of friction by using the three algorithms shows that oversampling process with SMOTE does not affect the attribute rank. It also shows that the oversampling process with SMOTE does not affect the degree of influence of each attribute in identifying blast cell lines in the identified WBC AML M0 and AML M1.

Table 7. Ranking Feature of Blast Cells

Feature	Information Gain		Gain Ratio		ReliefF	
	Non-SMOTE	SMOTE	Non-SMOTE	SMOTE	Non-SMOTE	SMOTE
WBC Diameter	0.135	0.447	0.164	0.214	0.0309	0.0618
Nucleus Roundness	0.107	0.209	0.116	0.204	0.0136	0.0290
Nucleus Ratio	0.000	0.000	0.000	0.000	0.0083	0.0186

4. Conclusion

The blast cell type classification model on the WBC identified AML M0 and AML1, using a combination of resampling, SMOTE and remove duplicate data, capable of delivering performance in either category. This is indicated by the value of the under the curve area worth 87.5%, with Random Forest classification algorithm and the validation use 10-folds cross-validation. The most dominant WBC feature in determining the blast cell type is the WBC diameter, followed by the nucleus roundness and nucleus ratio. The proposed model also has better performance compared with some previous studies. Further development of this research is to identify leukemia AML M0 and AML M1 subtypes, by referring to the identified blast cell types.

References

[1] Singh G, Bathla G, Kaur SP. *A Review to Detect Leukemia Cancer in Medical Images. International Conference on Computing*. Computing, Communication and Automation (ICCCA), 2016 International Conference. Noida. 2016: 1043–1047.

[2] Kasmin F, Prabuwono AS, Abdullah A. Detection of Leukemia In Human Blood Sample Based On Microscopic Images: A Study. *Journal of Theoretical & Applied Information Technology.* 2012; 46(2): 579–586.

[3] Theml H, Diem H, Haferlach T.Color Atlas of Hematology-Practical Microscopic and Clinical Diagnosis*.* 2nd Edition. Stuttgart-New York: Thieme. 2004.

[4] Houwen B. The differential cell count. *Laboratory Hematology.* 2001; 7: 89–100.

[5] Khobragade S, Mor DD, Patil CY. *Detection of leukemia in microscopic white blood cell images*. International Conference on Information Processing (ICIP). Pune. 2015: 435–440.

[6] Bagasjvara RG, Candradewi I, Hartati S, Harjoko A. *Automated detection and classification techniques of Acute leukemia using image processing: A review*. International Conference on Science and Technology-Computer (ICST). Yogyakarta. 2016: 35–43.

[7] Vogado LHS, Veras RMS, Araujo FHD, Silva RRV, Aires KRT. Leukemia diagnosis in blood slides using transfer learning in CNNs and SVM for classification. *Engineering Applications of Artificial Intelligence.* 2018; 72: 415–422.

[8] Rawat J, Singh A, Bhadauria HS, Virmani J. *Computer Aided Diagnostic System for Detection of Leukemia Using Microscopic Images*. 4th International Conference on Eco-friendly Computing and Communication Systems. Kurukshetra. 2015; 70: 748–756.

[9] Shafique S, Tehsin S. Computer-Aided Diagnosis of Acute Lymphoblastic Leukaemia. *Computational and mathematical methods in medicine.* 2018: 1–13.

[10] Putzu L, Di Ruberto C. *Investigation of Different Classification Models to Determine the Presence of Leukemia in Peripheral Blood Image*. International Conference on Image Analysis and Processing, Berlin. 2013; 8156: 612–621.

[11] Su J, Liu S, Song J. A segmentation method based on HMRF for the aided diagnosis of acute myeloid leukemia. *Computer methods and programs in biomedicine* 2017; 152: 115–123.

[12] Fatonah NS, Tjandrasa H, Fatichah C. Automatic Leukemia Cell Counting using Iterative Distance Transform for Convex Sets. *International Journal of Electrical and Computer Engineering*. 2018; 8(3): 1731–1740.

[13] Rawat J, Singh A, Hs B, Virmani J, Devgun JS. Computer assisted classification framework for prediction of acute lymphoblastic and acute myeloblastic leukemia. *Biocybernetics and Biomedical Engineering.* 2017; 37(4): 637–654.

[14] Devi SD, Sharada R, Shankari R, Tamilarasi T, Priya G. Automatic Diagnosis of Acute Lymphoblastic Leukemia Using Duplex Method. *Int. J. Healthc. Sci.* 2017; 5(1): 14–21.

[15] Selvaraj S, Kanakaraj B. Naïve Bayesian Classifier for Acute Lymphocytic Leukemia Detection. *ARPN Journal of Engineering and Applied Sciences.* 2015; 10(16): 6888–6892.

[16] Khosrosereshki MA, Menhaj MB. *A fuzzy based classifier for diagnosis of acute lymphoblastic leukemia using blood smear image processing*. 5th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS). Qazvin. 2017: 13–18.

[17] Suryani E, Wiharto W, Palgunadi S, Nurcahya Pradana T. *Classification of Acute Myelogenous Leukemia (AML M2 and AML M3) using Momentum Back Propagation from Watershed Distance Transform Segmented Images*. Journal of Physics: Conference Series. 2017; 801.

[18] Harjoko A, Ratnaningsih T, Suryani E, Wiharto W, Palgunadi S, Prakisya N. *Classification of Acute Myeloid Leukemia Subtype M1, M2 and M3 Using Active Contour without Edge Segmentation and Momentum Back P5ropagation Artificial Neural Network*. The 2nd International Conference on Engineering and Technology for Sustainable Development, Yogyakarta. 2017; 154: 1–6.

[19] Suryani E, Wiharto W, Palgunadi S, Putra YR. *Cells Identification of Acute Myeloid Leukemia AML M0 and AML M1 using k-N earest Neighbour Based on Morphological Images*. International Conference on Data and Software Engineering (ICoDSE), Palembang, Indonesia. 2017: 1–6.

[20] Suryani E, Wiharto W, Palgunadi S, Putra AP. Identification Of Acute Myeloid Leukemia M1 And M2 Through White Blood Cell Morphological Imaging Using Naïve Bayes' Classifier. *Journal of Information and Communication Technology (JICT)*. 2018.

[21] Choi JM. A selective sampling method for imbalanced data learning on support vector machines. PhD Disertation. Ames: Iowa State University; 2010.

[22] Ali A, Shamsuddin SM, Ralescu AL. Classification with class imbalance problem: A Review. *Int. J. Adv. Soft Comput. Its Appl.* 2015; 7(3): 176–204.

[23] Bi J, Zhang C. An empirical comparison on state-of-the-art multi-class imbalance learning algorithms and a new diversified ensemble learning scheme. *Knowledge-Based Systems.* 2018; 158: 81–93.

[24] Haixiang G, Yijing L, Shang J, Mingyun G, Yuanyue H, Bing G. Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications.* 2017; 73: 220–239.

[25] Rao RR, Makkithaya K. Learning from a Class Imbalanced Public Health Dataset: a Cost-based Comparison of Classifier Performance. *International Journal of Electrical and Computer Engineering*. 2017; 7(4): 2215–2222.

[26] Perkins S et al. 2018 Hematology and Clinical Microscopy Glossary. College of American Pathologists. 2018.

[27] Nazlibilek S, Karacor D, Ercan T, Sazli MH, Kalender O, Ege Y. Automatic segmentation, counting, size determination and classification of white blood cells. *Measurement*. 2014; 55: 58–65.

[28] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research.* 2002; 16: 321–357.

[29] Marateb HR, Goudarzi S. A noninvasive method for coronary artery diseases diagnosis using a clinically-interpretable fuzzy rule-based system. *Journal of research in medical sciences.* 2015; 20(3): 214–223.

[30] Liaw A, Wiener M. Classification and regression by randomForest. *R News*. 2002; 2(3): 18–22.

[31] Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and Regression Tree. New York: Chapman And Hall. 1984.

[32] Arieshanti I, Purwananto Y, Ramadhani A, Nuha MU, Ulinnuha N. Comparative study of bankruptcy prediction models. *TELKOMNIKA Telecommunication Computing Electronics and Control*. 2013; 11(3): 591–596.

[33] Jayalakshmi T, Santhakumaran A. Statistical Normalization and Back Propagationfor Classification. *International Journal of Computer Theory and Engineering.* 2011; 3(1): 89–93.

[34] Gorunescu F. Data Mining Concepts, Models and Techniques, Intelligent Systems Reference Library. *Berlin: Springer*. 2011.

[35] Hssina B, Merbouha A, Ezzikouri H, Erritali M. A comparative study of decision tree ID3 and C4.5. *International Journal of Advanced Computer Science and Applications.* 2014; 4(4): 13–19.

[36] Karegowda AG, Manjunath AS, Jayaram MA. Comparative study of attribute selection using gain ratio and correlation based feature selection. *International Journal of Information Technology and Knowledge Management.* 2010; 2(2): 271–277.

[37] Jensen R, Shen Q. Computational Intelligence and Feature Selection : Rough and Fuzzy Approaches. *USA: IEEE Press*. 2008.