Multi Distance and Angle Models of the Gray Level Co-occurrence Matrix (GLCM) to Extract the Acute Lymphoblastic Leukemia (ALL) Images

Arif Muntasa[1]   Muhammad Yusuf[1]*

[1]Universitas Trunojoyo Madura, Indonesia
* Corresponding author's Email: muhammadyusuf@trunojoyo.ac.id

Abstract: Acute Lymphoblastic Leukemia (ALL) is one of the most dangerous cancer diseases. Therefore, it is necessary to classify this disease accurately. We proposed a method to classify ALL using multi-distance models of the Gray Level Co-occurrence Matrix (GLCM). We employed sixteen distance models to obtain several features of the main object. We applied three channels of the image's enhancement results, where each channel has been extracted using second-order statistic models, which are homogeneity, entropy, energy, and contrast. We have obtained one hundred and ninety-two features for each object. Moreover, the feature extraction performances have been classified using Canberra and Chebyshev techniques. We have evaluated our proposed method using the ALL Image database and produced accuracy at 96.97%, where 2.27% false positive and 0.75% false negative. Our developed method has been compared to other approaches, such as Support Vector Machine (SVM)-Linear Binary Pattern (LBP), SVM Shape, Naïve Bayes Deterministic, Fuzzy based Leukemia detection, Color Correlation, Hausdorff SVM-based Leukemia detection, Automated Differential Learning Vector Quantization (LVQ), and Multi-Level Perceptron (MLP). The results show that our proposed method outperformed the others.

Keywords: Multi distance and angle models, Gray level co-occurrence matrix (GLCM), Leukemia detection, Canberra, Chebyshev.

1. Introduction

Acute lymphoblastic leukemia (ALL) is one of the most dangerous cancer diseases. Therefore, it is necessary to detect this disease early and accurately. Therefore, the patient can get medication earlier, quicker, and accurately. An automatic and computerized detection system is important and can be helpful for better, accurate, quick, and valid treatment. From the computer science perspective, various existing research on ALL detection systems, such as a Gray level co-occurrence matrix (GLCM) and a random forest-based, has been applied for ALL detection. The experiment obtained segmentation accuracy at 96.29%, classification accuracy at 99.004% for cytoplasm, and 96 % for nucleus in each detailed [1].

Moreover, the Auto SVM binary classifier and microscopic images and the GLCM texture feature have been implemented and produced accuracy at 72.4 % for the nucleus and 86.7 % for the cytoplasm area. Additionally, the experiment based on shape features achieved accuracy at 72.4 % for the nucleus and 56.1 % for the cytoplasm region. Also, the experiment using the GLCM texture-shape feature obtained the combined classification accuracy at 89.8 % for the nucleus-cytoplasm region [2]. In addition, three channels of GLCM and testing sets using Manhattan and Euclidean Distance have been developed and produced 91.54% of maximum accuracy [3]. Furthermore, [4] proposed Otsu's thresholding, watershed-based segmentation, canny edge detection, and k-mean clustering for segmentation. These algorithms are used to segment ALL datasets. Three parameters have been used in this research, such as precision, accuracy, and sensitivity. In this first pre-processing, the experiment produced 0.8378 and 0.8562 for precision and accuracy, respectively. Furthermore, the second pre-processing achieved higher accuracy at 0.8545

and precision at 0.8418. Moreover, the third pre-processing obtained precision at 0.8361 and accuracy at 0.8543.

Furthermore, [5] compared the effect of the saturation, hue saturation value (HSV), green, YCbCr, M, and Cb component of red greed blue (RGB), and cyan magenta yello (CMY) color spaces for nucleus segmentation. The research produced 85% segmentation accuracy for HSV and CMY models, [6] developed a detection system based on a microscopic blood smear image database. The developed approach utilized pre-processing, segmentation, and matching. Otsu and the maximally stable extremely regions (MSER) have been employed for segmenting the leukemia image database. The proposed system is evaluated using false rejection rate (FRR) and false acceptance rate (FAR). The experimental results achieved accuracy at 95.12%, where FRR is 4.75% and FAR is 5.0%, and [7] suggested an approach that uses imaging processing techniques to detect irregular blood cells. The program is trained using a deep convolutional neural network and uploaded into the hadoop distributed file system (HDFS) hadoop framework. The developed system achieved 98.6% accuracy, [8] introduced a hybrid particle swarm optimization (PSO)-contrast stretching (HPSO-CS) method. mean-square error (MSE), peak-signal-to-noise-ratio (PSNR), and absolute mean brightness error (AMBE) have been used to evaluate the proposed method. Furthermore, global enhancement and class limited adaptive histogram equalization (CLAHE) is employed for comparison. The experimental results obtained the lowest MSE at 2067.651, AMBE at 43.51827, and highest PSNR at 14.98671 compared with the global and CLAHE method, [9] developed a detection system for ALL by segmenting the white blood cells (WBC), red blood cells, and platelets. This research also utilized a Support Vector Machine (SVM) classifier and obtained 93.7% accuracy, 92% of sensitivity, as well as 91% of specificity.

Moreover,[10] proposed a region-based image segmentation method using deformable models. This model is employed for leukemia detection. The experiment produced the dice similarity coefficient at 0.92, [11] detected leukemia and determine whether it is acute myeloid leukemia (AML), chronic lymphocytic leukemia (CLL), chronic myeloid leukemia (CML), or ALL, [12] proposed a prediction system for leukemia. Furthermore, Image enhancement, thresholding, mathematical morphology, and labelling are appplied using LabVIEW and MATLAB in this research, [13] developed a leukemia detection method from microscopic white blood cell images. Image enhancement, image segmentation, and feature extraction have been utilized in this research.

Additionally, [14] utilized an image processing approach to analyzed healthy and infected leukemia blood smears, [15] developed a system to identify and classify various types of leukemia based on image processing techniques. This research achieved an accuracy of 80% for the ALL detection separately of one cell, 100% for AML detection separately of one cell, as well as 90% for cell detection throughout many cells, [16] proposed a leukemia prediction method using the histogram of oriented gradients (HOG) feature descriptor and logistics regression. This research employed acute lymphoblastic leukemia – image database (ALL-IDB) dataset as well as a canny edge detector and noise reduction operators for lymphocytes detection. Also, principal component analysis (PCA) has been implemented for image shapes detection. The experiment produced 96% of the maximum average accuracy of the proposed [17] proposed a method to detect a cancer cell. Some important features have been extracted from the blood cell images and learning multiple classifiers in this research. Furthermore, the experimental results show that gradient boosting decision tree classification algorithms obtained better performances than SVM. The proposed system achieved 85.6% of F1 scores on validation data.

Also, [18] developed a deep learning-based leukemia detection system based on the blood smear images. The proposed mutual information conducts the segmentation (MI) based hybrid model. This approach combines the segmentation results of the fuzzy C means algorithm and the active contour model. Moreover, the experimental performances show that the proposed chronological sine cosine algorithm (SCA) based deep convolutional neural network (CNN) classifier achieved 98.7% of the accuracy, [19] developed a detection system using image processing, genetic algorithm, and naïve bayes classifier. The experiment utilized 100 blood images for system evaluation and obtained a detection rate of 90.33%, [20]developed a system using image analytics. The cancer imaging archive (TCIA) repository has been used for the dataset experiment. In this research, k - means clustering produced a good performance for the segmentation stage, and neural networks achieved better results for the classification stage. The experiment achieved calculation accuracy at 0.838, precision at 0.835, area under curve (AUC) at 0.865, and F1 at 0.836 for neural networks, [21] suggested the segmentation process based on hue saturation intensity (HSI) and RGB color space. The developed HSI-based method has been successful for image segmentation. The background noise has been,

and significant features have been preserved, [22] proposed a framework for ALL identification from the microscopic image of WBC. In this research, random forest (RF), SVM, logistic regression (LR), and decision tree (DT) are implemented. The experimental results show SVM has successfully classified and identified the acute lymphoblastic cell, [23] developed two shape features, i.e., hausdorff dimension and contour signature for 108 blood smear images. These features and SVM are implemented for a lymphocytic cell nucleus classification.

Also, many methods have been developed by researchers, which are SVM-LBP and SVM Shape[24]. In this research, the experiment utilized geometric features and LBP texture features for the lymphocytes images detection as well as SVM as the classifier. The research performance shows that the LBP texture features have better results than shape features. Furthermore, naïve bayes deterministic (NBD) has been employed for lymphocyte classification and produced 75% accuracy [25]. Also, correlation-based feature selection has been utilized to lymphocyte cells with sixteen features on 260 blood lymphocyte images. This experiment obtained 92.30% accuracy [26]. In addition, fuzzy-based leukemia detection (Fuzzy-BLD), which are hausdorff dimension and contour signature, are employed to classify a lymphocytic cell nucleus and SVM for classification. The experiment result produced 93% accuracy [27].

Moreover, hausdorff dimension and contour signature are utilized for lymphocytic cell classification. The research produced 95% accuracy for 108 blood smear images[28]. Also, automated differential learning vector quantization (LVQ), multi-level perceptron (MLP), as well as SVM-shape have been used and achieved 83.33%, 89.74%, 91.03% accuracies [29]. However, improvements in similarity measurement methods depend on the segmentation and feature extraction results as well. We have proven that the segmentation and feature extraction process improvement can increase classification accuracy, though the similarity measurement methods use are simple. Therefore, based on the existing research, there is still limited research using the advanced gray level co-occurrence matrix (GLCM), such as multi-distance and angle models.

Furthermore, this research proposes a method to classify the ALL by using multi-distance models of the GLCM. We employed sixteen distance models to obtain several features of the main object. We applied three channels of the image's enhancement results, where each channel has been extracted using second-order statistic models, which are homogeneity, entropy, energy, and contrast. We have obtained one hundred and ninety-two features for each object. Moreover, the feature extraction results have been classified using canberra and chebyshev methods. Additionally, we proposed to improve the segmentation and feature extraction processes.

Moreover, we enhanced source images for each channel except the blue channel. Enhancement of the blue channel will cause the main object and backgrounds to be distinguished. The segmentation results can give impact the other process, which is feature extraction. We proposed multi-distance models to represent leukemia image samples to view based on several distances. Furthermore, the strong points of our proposed method are the method applied modified GLCM with some models representing each channel, using various angles to represent the objects, and representing objects with many distances. Further explanation will be presented in section 3.3.

Therefore, this research contributes by providing a method to classify the ALL by using multi-distance models of the GLCM. We employed sixteen distance models to obtain several features of the main object. The feature extraction results have been classified using canberra and chebyshev methods. There are some novelties in this research. Firstly, we employed all of the distances to build a gray level co-occurrence matrix. It is used to obtain the object characteristics based on the distance. Secondly, we also utilized all of the object channels to obtain the features. A combination of the models has been used to distinguish the infected image from and healthy one.

Furthermore, this paper consists of four sections: Introduction, proposed method, experimental results, and discussion, as well as conclusion. The introduction section contains the background and importance of the leukemia research, literature review of the existing research, gap, aim, novelty, contribution, and brief structure of the paper. The proposed method section contains stages of the experiments. The experimental results and discussion present the experimental database used, results of the novel model, and comparison with other methods. Lastly, the conclusion section captures the summary and future research.

2. Proposed method

We proposed six main processes to classify the ALL images: pre-processing, segmentation, building GLCM for each channel, feature extraction, image classification, and accuracy calculation. Firstly, we conducted image pre-processing, where all of the images for both the training and the testing sets are enhanced to obtain a clearer image. The images are

360

more easily distinguished between the main object and the background. Secondly, we need to calculate the best threshold value to segment the main object and background. The segmentation results are used as a template to obtain the original image based on the binary image. Thirdly, we built GLCM from the previous image results to reduce the image dimension.

Moreover, the GLCM results are applied to extract the features. Furthermore, we classified all of the testing sets by using canberra and chebyshev methods. Lastly, the classification results have computed the accuracy, sensitivity, and specificity using a confusion matrix.

2.1 Pre-processing

Firstly, we have to enhance the image source to obtain the best results when the segmentation process is conducted, as shown in Fig. 1. Firstly, we split the image source into the red, green, and blue channels. We enhanced the red and green channel by increasing high pixel intensity and decreasing low pixel intensity, whereas the blue channel is not necessary to be improved. The image enhancement results of the red and green channels are combined with the original blue channel. The channel combination results will be transformed into an HSV channel. Furthermore, we applied a median filter to enhance the hue channel.
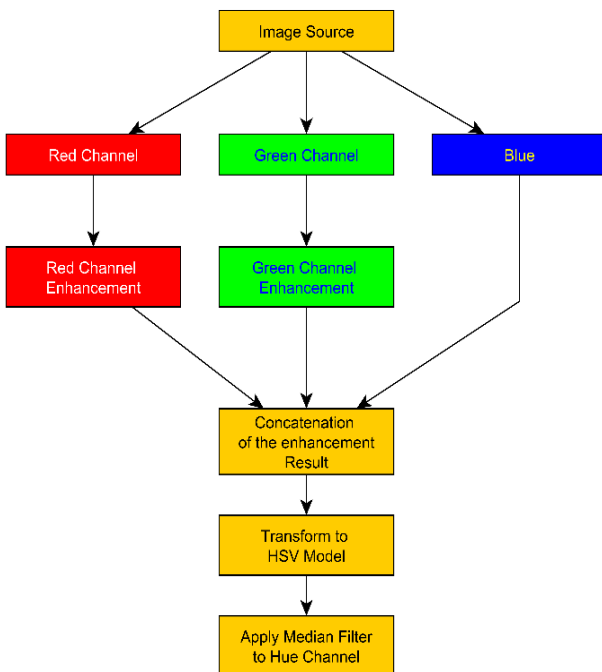

Figure 1. Pre-processing image

2.2 Image segmentation

We proposed two main stages in this process: calculated the threshold and converting it into a binary image. The pre-processing result is used to calculate the threshold value. In this case, we apply Otsu to obtain the best value as the threshold. The best value is utilized to distinguish the main object and the background. The threshold value can be calculated by using the Otsu algorithm can be seen as follows:

Calculate histogram image
Calculate probabilities for each intensity
For each index of $t \leftarrow 1$ to Maximum intensity
Update the values of ω_k and μ_k.
Calculate the variance value of $\sigma_b^2(t)$.
The best threshold value is the maximum of $\sigma_b^2(t)$

$$\sigma_w^2(t) = \sum_{j=0}^{1} \omega_j(t) \times \sigma_j^2(t) \tag{1}$$

$$\omega_0(t) = \sum_{k=0}^{t-1} p(t) \tag{2}$$

$$\omega_1(t) = \sum_{k=t}^{L-1} p(t) \tag{3}$$

$$\sigma_b^2(t) = \sigma^2 - \sigma_w^2 \tag{4}$$

$$\sigma_b^2(t) = \omega_0(t) \times \omega_1(t) \times [\mu_0(t) - \mu_1(t)]^2 \tag{5}$$

$$\mu_0(t) = \frac{\sum_{j=0}^{t-1} j \times P(j)}{\omega_0(t)} \tag{6}$$

$$\mu_1(t) = \frac{\sum_{j=t}^{L-1} j \times P(j)}{\omega_1(t)} \tag{7}$$

$$\mu_T = \sum_{j=0}^{L-1} j \times P(j) \tag{8}$$

The symbols ω_k is weight, μ_k represents the average value, histogram image is L, and P represents class probability. Based on the best threshold value, we can distinguish the main and background image

$$g(x,y) = \begin{cases} 0, & f(x,y) < Threshold \\ 1, & f(x,y) \geq Threshold \end{cases} \tag{9}$$

Symbols of $g(x, y)$ represent binary conversion result of a new image, $f(x, y)$ describes an original image, whereas x and y are the pixel position of an image pixel. Furthermore, we need to determine the region of interest of an image. An image segmentation results have to be cropped to remove noise and reduce image size. The cropping result is a region of interest of an image.

2.3 Building gray level co-occurrence matrix

The region of interest image results will be used to build a GLCM. In this case, we applied some offset values to obtain some GLCMs. The offset values are the distance that built based on the direction of the pixels as shown in Fig. 2, which are

Horizontal, it is described by using [0, D]

The right diagonal is represented by using [-D, D]

Vertical, it is symbolized by using [-D, 0]

The left diagonal is written as [-D, -D]

The symbol of D represents the distance between pixels

We proposed to build multiple GLCMs known as the multiple GLCMs. The GLCM describes the relationship between pixels on the image. We define sixteen offsets to represent relationship pixels on the GLCM. The offset is defined as the direction of the horizontal pixels, the right diagonal, the vertical, and the left diagonal. We proposed to utilized sixteen offsets, as shown in Fig. 3. The offset values for each direction can be described in detail as follows:

Horizontal [0, -D] has offset points [0, 1], [0,2], [0,3], and [0.4].

The right diagonal [-D, D] has offset points [-1, 1], [-2,2], [-3,3], and [-4.4].

Vertical [-D, 0] has offset points [-1, 0], [-2,0], [-3,0], and [-4.0].

The left diagonal [-D, -D] has offset points [-1, -1], [-2,-2], [-3,-3], and [-4.-4].

Furthermore, the GLCM results are normalized. So that the range has 0 as the minimum and 1 as the maximum values as shown in the following equation:

$$P_{i,j} = \frac{V_{i,j}}{\sum_{i,j=0}^{N-1} V_{i,j}} \qquad (10)$$

The symbol of *V* represents the value of the GLCM results, where symbols of *I* and *j* are the row and column index of the GLCM. The normalization process results are illustrated by *P*.
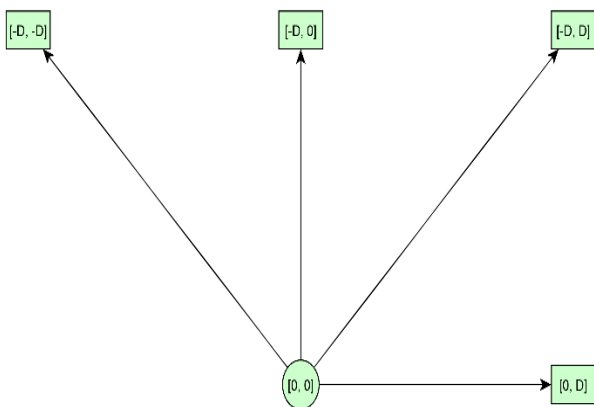

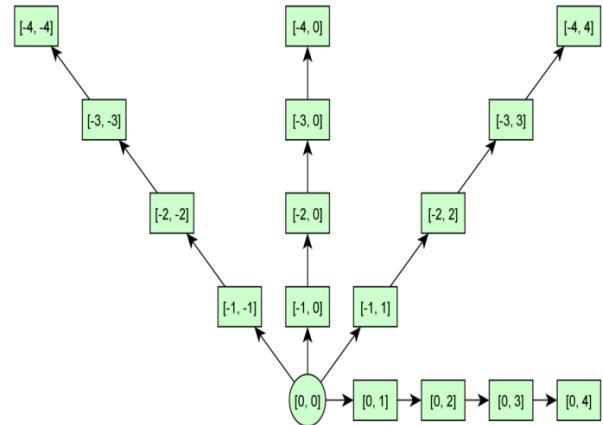Figure 2. Pixels direction of the offset

Figure 3. The values of the offset on the GLCM

2.4 Features extraction of the image

The main features of the image are required to represent an object on the image. In this case, we proposed four features, which are homogeneity (F_a), entropy (F_b), energy (F_c), and contrast (F_d), as shown in the following equations:

$$F_a = \sum_{i=0} \sum_{j=0} \frac{P_{i,j}(i,j)}{1+|i-j|} \qquad (11)$$

$$F_b = -\sum_i \sum_j P_{i,j}(i,j) \times \log P_{i,j}(i,j) \qquad (12)$$

$$F_c = \sum_{i=0} \sum_{j=0} P^2(i,j) \qquad (13)$$

$$F_d = \sum_{i=0} \sum_{j=0} (i-j)^2 \times P_{i,j}(i,j) \qquad (14)$$

Each GLCM matrix is calculated to obtain the feature values of homogeneity, entropy, energy, and contrast. Therefore, we have sixty-four features. It is obtained from the result of the multiplication of sixteen offsets and four as described in Fig. 4

We have proposed to apply the features for each channel, which are the red, green, and blue channels, as shown in Fig. 4. Furthermore, we found sixty-four features (see Fig. 5) for the red, green, and blue channels, respectively. Each channel has an eight-bit scale for each pixel, 0 as the minimum and 255 ($2^{8\,bit}$-1 = 256 – 1 = 255) as the maximum pixel intensity values. It means that we have sixty-four features for the red, green, and blue. If a channel produced sixty-four features, three channels generated one hundred and ninety-two features, as shown in Fig. 4.

Furthermore, we need to compose the features into a row matrix. We have placed the green channel into the right of the right channel. Similarly, we need to put the blue channel into the right of the green channel as well. It can be shown in Fig. 6.
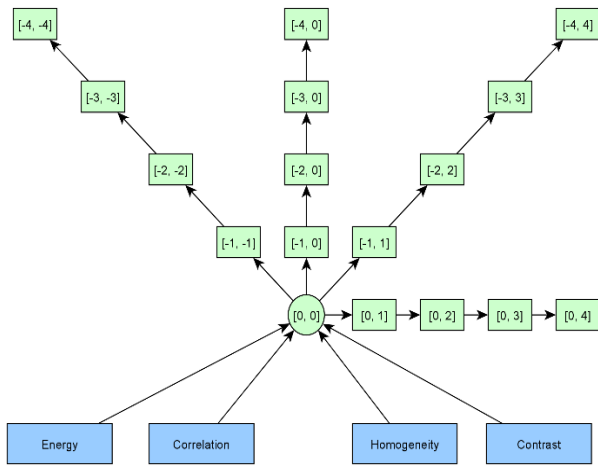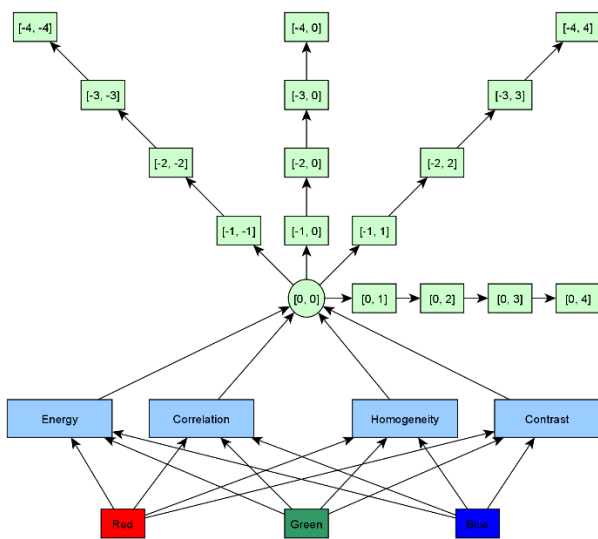
Figure 4. Features on the GLCM

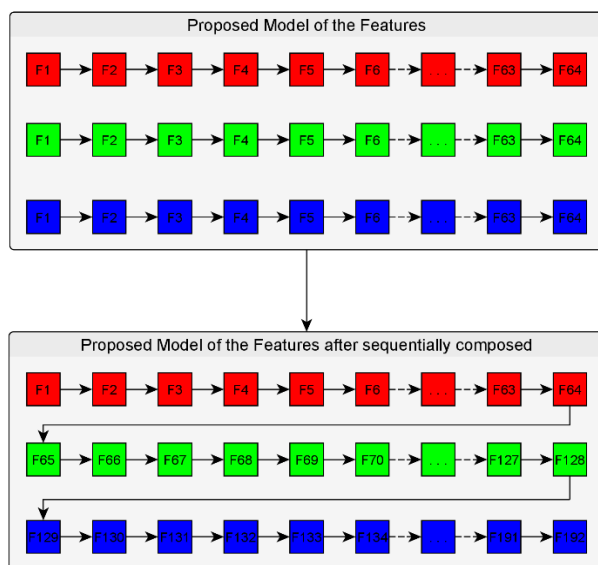Figure 5. GLCM features of the red, green, and blue channels

Figure 6. Transpose the object feature into a row feature

2.5 Image classification

Our proposed method has produced one hundred and ninety-two features for each image. We measure the feature extraction results between the training and the testing sets. In this case, we proposed canberra and chebyshev method to classify the acute lymphoblastic leukemia image (ALLI). Image classification is conducted to distinguish whether the ALLI is infected with leukemia or not. If the feature extraction results of the training and the testing sets are represented $\mathcal{F}(k)$ and $\mathcal{M}(k)$. The symbol of k represents the feature index of the training and the testing sets. In this case, the value of k has a range between 1 and 192. Modeling of the chebyshev methods can be shown in the following equation:

$$C_q = \min\left(max\left(\left|\mathcal{F}_p(k) - \mathcal{M}_q(k)\right|\right)\right) \quad (15)$$

In this equation, the value of k is $k \in 1, 2, 3, ..., NoF$. NoF represents the number of features for both the training and the testing sets. At the same time, the symbols of p and q are the index of the training and the testing image. Chebyshev method is the simple method to classify the object. Chebyshev is calculated by finding the maximum difference of the corresponding features.

Furthermore, determine the minimum value of the maximum distance from the classified image to all training sets. An index of the minimum value is the classification results of the testing set. We also apply the canberra method to classify the testing sets based on the training sets as shown in the following equation:

$$Y_q = \sum_{k=1}^{NoF} \frac{\left|\mathcal{F}_p(k) - \mathcal{M}_q(k)\right|}{\left\|\mathcal{F}_p(k)\right\| + \left|\mathcal{M}_q(k)\right|} \quad (16)$$

2.6 Accuracy calculation

The results of the calculation are used to calculate the accuracy. We applied a confusion matrix to calculate the accuracy. We need to distinguish true and prediction conditions as describe in Table 1. In this table, we have four conditions: The actual condition is positive, and the predicted result is positive. It is called a True Positive (TP). Moreover,

Table 1. Confusion matrix

Predicted \ Actual	Positive	Negative
Positive	TP	FP
Negative	FN	TN

Table 2. Ten scenarios models using canberra and chebyshev

Method	For Each Class									
	Training					Testing				
	63	64	65	66	67	67	66	65	64	63
Canberra	S1	S2	S3	S4	S5	S1	S2	S3	S4	S5
Chebyshev	S6	S7	S8	S9	S10	S6	S7	S8	S9	S10
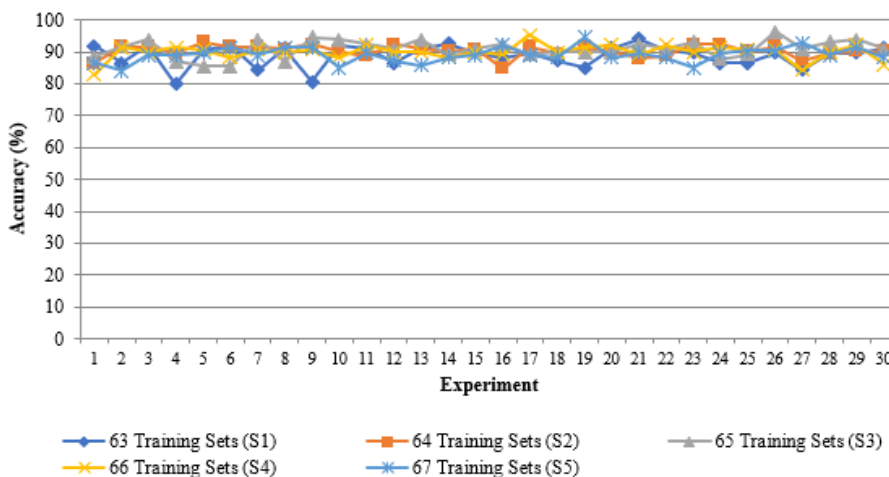
Figure 7. Classification results of the proposed method using Canberra

if the actual condition is negative, and the predicted result is negative as well. It is called a true negative (TN). In addition, if the actual condition is negative, and the predicted result is positive, then is type 1 error is called a false positive (FP). Otherwise, if the actual condition is positive and the predicted result is negative, is type 2 error is called a false negative (FN). We can compute accuracy, sensitivity, and specificity based on Table 1 as the following equation:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (17)$$

$$Sensitivity = \frac{TP}{TP+FN} \qquad (18)$$

$$Specificity = \frac{TN}{TN+FP} \qquad (19)$$

3. Experimental results and discussions

In this paper, we employed ALL images to evaluate our proposed method. The images have been labeled by 0 for healthy and 1 for patients. We have obtained two hundred and sixty images, where one hundred and thirty images are healthy, and the rest are images of leukemia patients. Fig. 8 shows a sample of the ALL images for both the healthy and leukemia patients. We apply random sampling as the training sets to evaluate our proposed method. We carry out several experiments based on the following scenario.

Table 2 showed that we had conducted ten scenario models, where we carried out five scenarios using canberra and the rest scenarios using chebyshev. We conducted random sampling thirty times for each scenario. Therefore, we have carried out three hundred experiments (10 Scenario models (S) x 30 random samplings). In this case, we employed five scenarios using canberra, and for the rest scenario, we applied chebyshev.

3.1 Experimental results using canberra

This paper has conducted five different scenarios, where each scenario model is conducted thirty times with different index random samplings. As shown in Table 1, five scenario models are:

63 training sets, 67 testing sets (S1, 1st Scenario)
64 training sets, 66 testing sets (S2, 2nd scenario)
65 training sets, 65 testing sets (S3, 3rd scenario)
66 training sets, 64 testing sets (S4, 4th scenario)
67 training sets, 63 testing sets (S5, 5th scenario)

For example, in the first scenario, we randomly selected 63 of 130 data for the healthy person and the index of the patients; the rest of the indexes is applied to refer to the testing sets. Similarly, we also apply to the second until the last scenario. The experimental results using canberra are illustrated in Fig. 7. The results show that the accuracy of the novel method
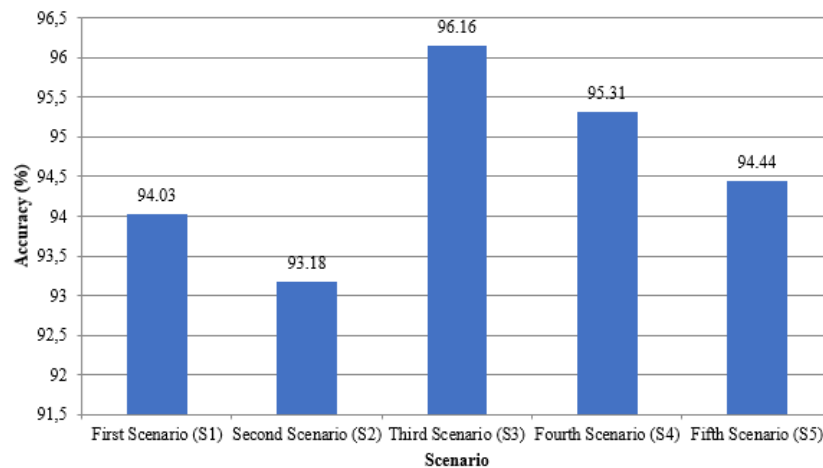
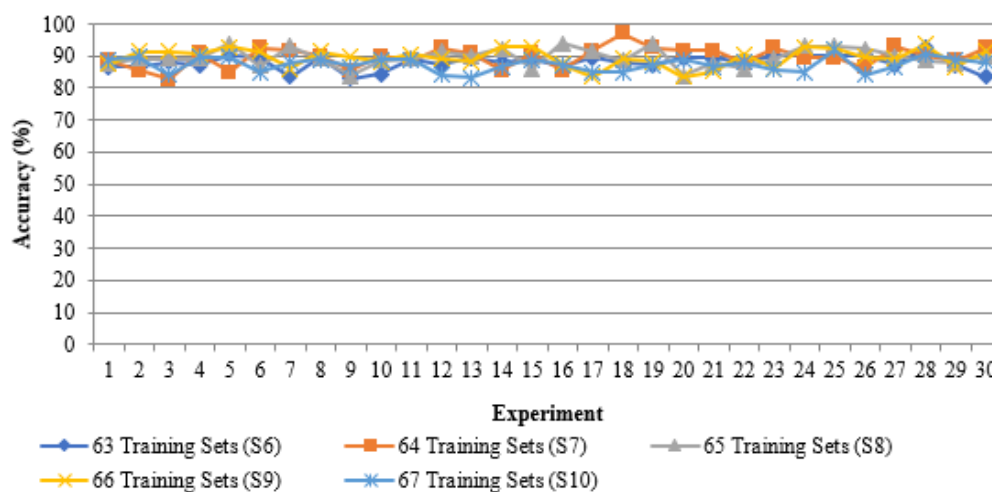Figure 8. Maximum accuracy of the proposed method using canberra

Figure 9. Classification results of the proposed method using chebyshev

has ranged between 80% and 97%. Furthermore, standard deviation for each scenario is 0.03, 0.02, 0.02, 0.02, and 0.02. It means that our proposed method has produced accuracy with a small difference. The experimental results also show that 65 training sets of the 3rd scenario have produced better accuracy than the others, as the accuracy yield has higher fluctuation. In the third scenario, our proposed method has generated a standard deviation, which is 0.02. It means that the difference in accuracy produced is very small—the smaller the standard deviation value that is raised, the better the system's performance.

Based on the experimental investigation, results show that classification error is caused by segmentation error. The segmentation results of the blood image objects have an impact on the classification process.

Poor image quality has caused segmentation errors. Therefore, the results of feature extraction produced are inaccurate. These results will have an

impact on the classification process. Fig. 8 shows the maximum accuracy of the proposed method using canberra.

3.2 Experimental results using chebyshev

We also classified our feature extraction results using chebyshev. Similar to canberra, the second model has five scenarios: the sixth and tenth scenarios (S6, S7, S8, S9, and S10). We have conducted thirty experiments for each scenario, where the system randomly determines the training sets' indexes. The 6th scenario has produced accuracy between 88.06% and 92.54%, whereas the standard deviation generated is 0.02 of the thirty experiments. The 7th scenario has produced accuracy between 89.70% and 96.97%. Our proposed method has generated a standard deviation based on thirty experiments, which is 0.03. The 8th, 9th, and last scenarios have produced accuracy between minimum accuracy 89.64%, 89.48%, and 87.33 respectively,
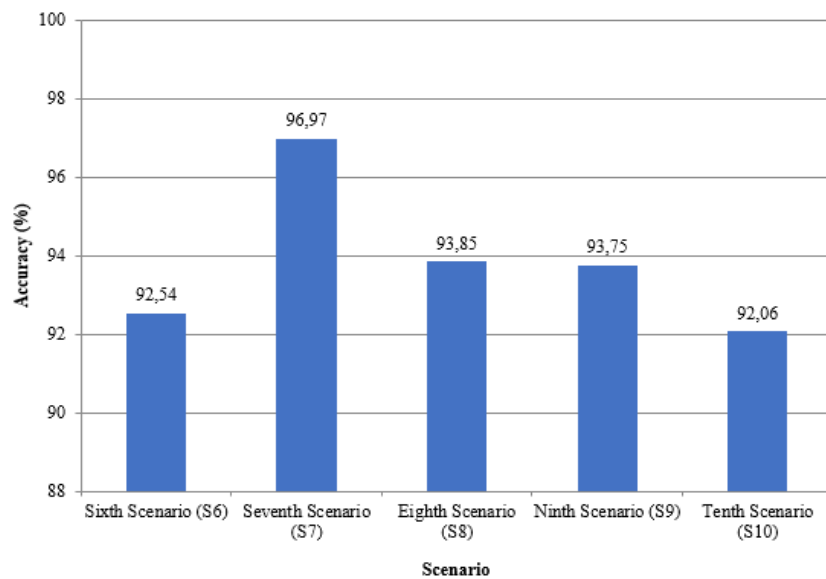
Figure 10. Maximum accuracy of the proposed method using chebyshev

whereas maximum accuracy is 93.85% for the 8th scenario, 93.75% the 9th scenario, and 92.06% for the last scenario. The last three scenarios resulted in the same standard deviation, which is 0.02.

All of the experimental results can be shown in Fig. 9. Different color and point models indicate the different scenarios. Based on Fig. 10, we can show that the 7th scenario has produced the best results, where the 7th scenario has resulted in the highest values of the minimum and the maximum accuracy. However, there is a difference in the standard deviation to the others, which is 0.01. However, the difference in accuracy results of the 7th scenario is small as well. The maximum accuracy of the second model results can be shown in Fig. 10. The highest accuracy has been obtained by the 7th scenario, which is 96.97%. The investigation results show that only 4 of 132 images have error classification. 3 of the 4 error classifications are false positives. It means that three negatives of ALL have been detected as positive of ALL. The rest of the miss-classification is a false negative. Our proposed method has miss-classification 1 of 132 patients, where it should be positive of the ALL. Still, our proposed method has detected the negative of ALL. It means that our proposed method has produced false positive and negative errors less than 0.75 % and 2.2%.

We have investigated the experimental results in detail. The results show that the segmentation error has caused the real blood of the ALL image cannot be captured so that the GLCM built does not match the original object. The fatal impact of the segmentation error is the error of the feature extraction results as well. Finally, missing of the feature extraction caused the error classification.

3.3 Comparing to other methods

We also compared our novel method to other methods, which are SVM-LBP [24], SVM shape [24], naïve bayes deterministic (NBD) [25], color correlation (CC) [30], fuzzy based leukemia detection (Fuzzy-BLD) [27], hausdorff SVM-based leukemia detection (H-SVM-LD) [28], and automated differential learning vector quantization (LVQ) [29], multi-level perceptron (MLP) [29], as well as SVM-shape [29]. The results show that our proposed method has outperformed the others. The comparison results can be seen in Fig. 11. It means that our proposed method can be accepted to be used as ALL image classification.

Furthermore, our proposed method has produced better accuracy than the others. We enhanced the segmentation process to obtain a better image so that the images are easy to be segmented. However, feature extraction results are influenced by the segmentation results. Additionally, we have proposed multi distances to represent the leukemia object of the image for each channel. Our proposed method has produced one hundred and ninety-two features.

Moreover, our proposed method can maintain the originality of the main object during the feature extraction process. Therefore, the feature extraction results can distinguish an image, whether the image is leukemia or not. Meanwhile, other methods prioritize developing the measurement method, while
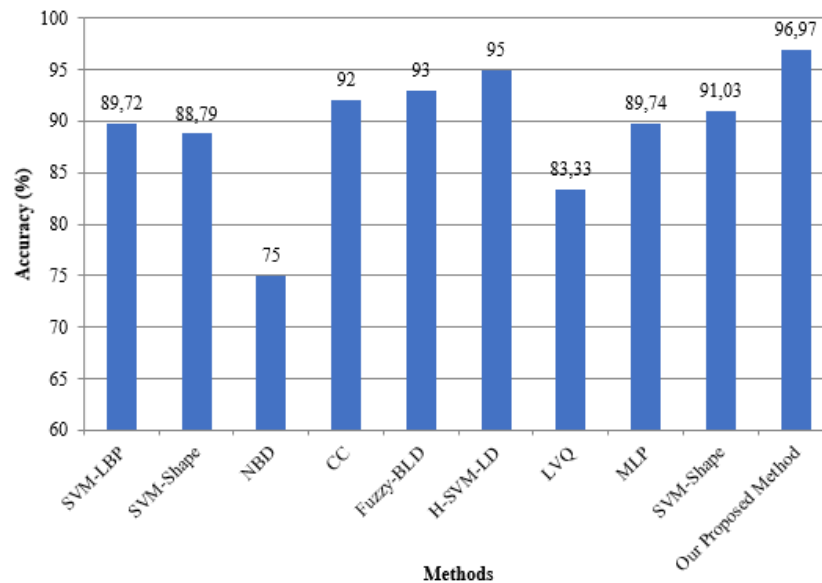
Figure 11. Comparison results between our proposed method to other methods

feature extraction and segmentation are not priorities such as naïve bayes deterministic (NBD) [25], color correlation (CC) [30], fuzzy based leukemiadetection (Fuzzy-BLD) [27], local binary pattern, hausdorff SVM-based leukemia detection (H-SVM-LD) [28], and automated differential learning vector quantization (LVQ) [29] as well.

4. Conclusion

We have conducted several experiments to evaluate our novel method. Based on the experimental, we have proved that the best performance of our proposed method has classified the ALL image up to 96.97%, where false positive and negative error less than 0.75 % and 2.2%. We have investigated our experimental results, where classification error can occur if the object's segmentation process is inaccurate. However, the segmentation process also plays an important role in classification.

Our proposed method has been compared to other methods, such as SVM-LBP, SVM-shape, NBD, CC, fuzzy-BLD, Hausdorff SVM-based leukemia detection (H-SVM-LD), and automated differential LVQ, MLP, and SVM-shape as well. The results show that our proposed method has produced better performance than the others. The future work will be desk research and compiling our literature reviews and our two years research results to be a literature review paper about leukemia detection using a computer science approach.

Conflicts of interest

The authors declare no conflict of interest

Author contributions

The authors' contributions are as follows: "Conceptualization, Arif Muntasa; methodology, Arif Muntasa; software, Arif Muntasa; validation, Arif Muntasa; formal analysis, Arif Muntasa; investigation, Arif Muntasa; literature reviews Muhammad Yusuf; resources, Arif Muntasa; data curation, Arif Muntasa; writing—original draft preparation, Arif Muntasa and Muhammad Yusuf; writing—review and editing, Arif Muntasa and Muhammad Yusuf; revising, Arif Muntasa and Muhammad Yusuf; corresponding, Muhammad Yusuf; Proofreading, Muhammad Yusuf; visualization, Arif Muntasa; supervision, Arif Muntasa; project administration, Arif Muntasa and Muhammad Yusuf; funding acquisition, Arif Muntasa and Muhammad Yusuf".

Acknowledgments

References

[1] S. Mishra, B. Majhi, P. K. Sa, and L. Sharma, "Gray level co-occurrence matrix and random forest based acute lymphoblastic leukemia detection", *Biomed. Signal Process. Control*, Vol. 33, pp. 272–280, 2017, doi: 10.1016/j.bspc.2016.11.021.

[2] J. Rawat, A. Singh, H. S. Bhadauria, and J. Virmani, "Computer Aided Diagnostic System for Detection of Leukemia Using Microscopic Images", *Procedia Computer Science*, 2015, Vol. 70, pp. 748–756, doi: 10.1016/j.procs.2015.10.113.

[3] A. Muntasa and M. Yusuf, "Three Channels for Gray Level Co-occurrence Matrix (GLCM) to detect Acute Lymphoblastic Leukemia (ALL) Images", 2020, doi: 10.1145/3425577.3425583.

[4] T. Bhagya, K. Anand, D. S. Kanchana, and A. A. S. Remya, "Analysis of image segmentation algorithms for the effective detection of leukemic cells", In: *Proc. of Int. Conf. Trends Electron. Informatics, ICOEI 2019*, No. Icoei, pp. 1232–1236, 2019, doi: 10.1109/ICOEI.2019.8862696.

[5] S. Shinde, N. Sharma, P. Bansod, M. Singh, and C. K. Singh Tekam, "Automated nucleus segmentation of leukemia blast cells : CColor spaces study", In: *Proc. of 2nd Int. Conf. Data, Eng. Appl. IDEA 2020*, 2020, doi: 10.1109/IDEA49133.2020.9170721.

[6] M. V. Rege, M. B. Abdulkareem, S. Gaikwad, and B. W. Gawli, "Automatic Leukemia Identification System Using Otsu Image segmentation and MSER Approach for Microscopic Smear Image Database", In: *Proc. of Int. Conf. Inven. Commun. Comput. Technol. ICICCT 2018*, No. Icicct, pp. 267–272, 2018, doi: 10.1109/ICICCT.2018.8473101.

[7] K. S. Nayaki, J. Denny, M. M. Rubeena, and J. K. Denny, "Cloud based Acute Lymphoblastic Leukemia Detection Using Deep Convolutional Neural Networks", In: *Proc. of 2nd Int. Conf. Inven. Res. Comput. Appl. ICIRCA 2020*, pp. 530–536, 2020, doi: 10.1109/ICIRCA48905.2020.9183249.

[8] N. H. Harun, J. A. Bakar, Z. A. Wahab, M. K. Osman, and H. Harun, "Color Image Enhancement of Acute Leukemia Cells in Blood Microscopic Image for Leukemia Detection Sample", In: *Proc. of ISCAIE 2020 - IEEE 10th Symp. Comput. Appl. Ind. Electron.*, No. 3, pp. 24–29, 2020, doi: 10.1109/ISCAIE47305.2020.9108810.

[9] S. Shafique, S. Tehsin, S. Anas, and F. Masud, "Computer-assisted Acute Lymphoblastic Leukemia detection and diagnosis", In: *Proc. of 2019 2nd Int. Conf. Commun. Comput. Digit. Syst. C-CODE 2019*, pp. 184–189, 2019, doi: 10.1109/C-CODE.2019.8680972.

[10] M. Sharma, S. Kanwal, A. Bhan, and A. Goyal, "Computer Based Diagnosis of Leukemia in Blood Samples Using Improved Region Based Deformable Models", In: *Proc. of 2nd Int. Conf. Trends Electron. Informatics, ICOEI 2018*, No. Icoei, pp. 1437–1441, 2018, doi: 10.1109/ICOEI.2018.8553737.

[11] P. Jagadev and H. Virani, "Detection of Leukemia and its Types using Image Processing and Machine Learning", pp. 522–526, 2017.

[12] C. Raje and J. Rangole, "Detection of Leukemia in microscopic images using image processing", In: *Proc. of Int. Conf. Commun. Signal Process. ICCSP 2014 - Proc.* pp. 255–259, 2014, doi: 10.1109/ICCSP.2014.6949840.

[13] S. Khobragade, D. D. Mor, and C. Y. Patil, "Detection of leukemia in microscopic white blood cell images", In: *Proc. of IEEE Int. Conf. Inf. Process. ICIP 2015*, pp. 435–440, 2016, doi: 10.1109/INFOP.2015.7489422.

[14] T. Dharani and S. Hariprasath, "Diagnosis of Leukemia and its types Using Digital Image Processing Techniques", In: *Proc. of 3rd Int. Conf. Commun. Electron. Syst. ICCES 2018*, No. Icces, pp. 275–279, 2018, doi: 10.1109/CESYS.2018.8724075.

[15] R. Sigit, M. M. Bachtiar, and M. I. Fikri, "Identification Of Leukemia Diseases Based On Microscopic Human Blood Cells Using Image Processing", In: *Proc. of 2018 Int. Conf. Appl. Eng. ICAE 2018*, pp. 1–5, 2018, doi: 10.1109/INCAE.2018.8579387.

[16] W. Daelemans, "Machine Learning Approaches", pp. 285–304, 1999, doi: 10.1007/978-94-015-9273-4_17.

[17] S. Mandal, V. Daivajna, and V. Rajagopalan, "Machine learning based system for automatic detection of leukemia cancer cell", In: *Proc. of 2019 IEEE 16th India Counc. Int. Conf. INDICON 2019 - Symp. Proc.*, pp. 1–4, 2019, doi: 10.1109/INDICON47234.2019.9029034.

[18] K. K. Jha and H. S. Dutta, "Mutual Information based hybrid model and deep learning for Acute Lymphocytic Leukemia detection in single cell blood smear images", *Comput. Methods Programs Biomed.*, Vol. 179, p. 104987, 2019, doi: 10.1016/j.cmpb.2019.104987.

[19] R. J. A. Cabrera, C. A. P. Legaspi, E. J. G. Papa, R. D. Samonte, and D. D. Acula, "HeMatic: An automated leukemia detector with separation of

overlapping blood cells through Image Processing and Genetic Algorithm", In: *Proc. of 2017 IEEE Int. Conf. Appl. Syst. Innov. Appl. Syst. Innov. Mod. Technol. ICASI 2017*, pp. 985–987, 2017, doi: 10.1109/ICASI.2017.7988618.

[20] A. Belhekar, Y. Bhelkar, K. Gagare, K. Rajeswari, R. Bedse, and M. Karthikeyan, "Analytics", 2019.

[21] M. Ysia, "*Nor Hazlyna Harun, *MY Mashor, *NR.Mokhtar, *Aimi Salihah, A.N, #Rosline Hassan, *R.A.A. Raof, *MK Osman", No. Isspa, pp. 749–752, 2010.

[22] M. N. Q. Bhuiyan, S. K. Rahut, R. A. Tanvir, and S. Ripon, "Automatic acute lymphoblastic leukemia detection and comparative analysis from images", In: *Proc. of 2019 6th Int. Conf. Control. Decis. Inf. Technol. CoDIT 2019*, pp. 1144–1149, 2019, doi: 10.1109/CoDIT.2019.8820299.

[23] S. Mohapatra and D. Patra, "Automated cell nucleus segmentation and acute leukemia detection in blood microscopic images", In: *Proc. of Int. Conf. Syst. Med. Biol. ICSMB 2010 - Proc.*, No. December, pp. 49–54, 2010, doi: 10.1109/ICSMB.2010.5735344.

[24] V. Singhal and P. Singh, "Local Binary Pattern for automatic detection of Acute Lymphoblastic Leukemia", In: *Proc. of 2014 20th Natl. Conf. Commun. NCC 2014*, No. February 2014, pp. 3–8, 2014, doi: 10.1109/NCC.2014.6811261.

[25] S. Selvaraj and B. Kanakaraj, "Naïve Bayesian classifier for Acute Lymphocytic Leukemia detection", *ARPN J. Eng. Appl. Sci.*, Vol. 10, No. 16, pp. 6888–6892, 2015.

[26] V. Singhal and P. Singh, "Correlation based feature selection for diagnosis of acute lymphoblastic leukemia", In: *Proc. of ACM Int. Conf. Proceeding Ser.*, Vol. 10-13-Augu, pp. 5–9, 2015, doi: 10.1145/2791405.2791423.

[27] S. Mohapatra, S. S. Samanta, D. Patra, and S. Satpathi, "Fuzzy based blood image segmentation for automated leukemia detection", 2011, doi: 10.1109/ICDECOM.2011.5738491.

[28] S. Mohapatra and D. Patra, "Automated leukemia detection using hausdorff dimension in blood microscopic images", In: *Proc. of Int. Conf. "Emerging Trends Robot. Commun. Technol. INTERACT-2010*, pp. 64–68, 2010, doi: 10.1109/INTERACT.2010.5706196.

[29] G. Ongun, U. Halici, K. Leblebicioglu, V. Atalay, M. Beksac, and S. Beksac, "Feature extraction and classification of blood cells for an automated differential blood count system", In: *Proc. of Int. Jt. Conf. Neural Networks*, Vol. 4, No. 2, pp. 2461–2466, 2001, doi: 10.1109/ijcnn.2001.938753.

[30] F. Scotti, "Robust segmentation and measurements techniques of white cells in blood microscope images", In: *Conf. Rec. - IEEE Instrum. Meas. Technol. Conf.*, No. April, pp. 43–48, 2006, doi: 10.1109/IMTC.2006.235499.