Detection of Acute Myeloid Leukemia based on White Blood Cell Morphological Imaging using Naïve Bayesian Algorithm

Esti Suryani[1], Wiharto[2]*, Adi Prasetya Putra[3], Wisnu Widiarto[4]

Department of Informatics, Universitas Sebelas Maret, Surakarta, Indonesia

Abstract—The process of diagnosing AML is based on the complete blood-count analysis of the patients. As such, it involves high energy consumption, long completion times, and is rather expensive compared to conventional medical practices. One of the methods for identifying tumor cells involves the utilization of image-processing techniques based on the morphology of white blood cells (WBCs). The principal objective of this study involves the identification of AML cells—especially of the AML M1 and AML M2 types—through morphological imaging of WBCs using the Naïve Bayes' Classifier. The Image-processing methods used in this study include YCbCr color space classification, image thresholding, morphological operations, chain code representation, and the use of bounding boxes. Regardless of the processing technique used, all identification procedures, performed in this study, were based on the Naïve Bayes' Classifier. The test process was performed on 30 images of each of the AML M1 and M2 cell types. The use of the cell identification method proposed in this study demonstrated an accuracy of 73.33%. While the accuracy of cell type identification is 54.92%. Based on the results obtained in this study, it is inferred that the Naïve Bayes' Classifier method can be employed in the process of identifying dominant AML cell types amongst AML M1 and AML M2 (myeloblast, promyelocyte, myelocyte, and metamyelocyte) based on the morphology of WBCs.

Keywords—Leukemia; acute myeloid leukemia; morphology; image processing; Naïve Bayes

I. INTRODUCTION

Leukemia is a type of cancer, where the bone marrow tends to produce abnormal white blood cells (WBCs) [1]. Leukemia may be divided into four main types—acute myeloid leukemia (AML), chronic myeloid leukemia (CML), acute lymphoblastic leukemia (ALL), and chronic lymphocytic leukemia (CLL). AML is the most commonly diagnosed type of leukemia; of all cases of AML diagnosed to date, 80% have been found to occur in adults, and 15–20% cases have been found to occur in children [2].

National Cancer Institute, 2013 maintains a record of the occurrence of AML amongst adults belonging to different age groups. It is seen that in the age group of 30–34 years, a single case of leukemia is recorded per 100,000 people; between 65–69 years, these numbers increase by ten times to 10 cases per 100,000 people. This number increases still further for adults beyond 70 years of age and the trend continues to rise until the age group of 80–84 years[3]. In general, the process of diagnosing AML involves analysis of the complete blood count of the patient, wherein the pathologist counts the number of red blood cells, WBCs, platelets, and checks for the presence of abnormal WBCs[4]. However, this method is time-consuming, requires energy, and is one of the most expensive routine tests performed in clinical hematological laboratories[5].

According to the French–America–British (FAB) classification, AML is classified into eight types—M0, M1, M2, M3, M4, M5, M6, M7[6]. This classification is based on the calculation of the cell-maturity level as well as the lineage from blast cells[7]. Utilization of various image-processing techniques offers an alternate approach aiding the identification of blood cells[8]-[9]. Cell identification could be performed through process analysis of digital images of a leukemia-positive blood cell preparation, captured using a digital microscope.

Previously, research has been conducted on the identification of blood cells using image processing techniques on AML images based on the white morphology of white blood cells[6][,8], as well as classification of the types of AML[8][9]. The technique used there are four stages of image acquisition, segmentation, feature extraction, and identification. The segmentation process aims to separate white blood cells and red blood cells. The techniques used in image segmentation include colour filter, Canny Edge Detection, Ellipse Detection. While the identification process uses a Fuzzy Rule-Based System with the Sugeno Order Zero method[9]. Another study used K-means data mining to separate the nucleus into white blood cells. After separating the nucleus and white blood cells, the next step is to extract the characteristics of the shape and the characteristics of densitometry, then perform the process of classification of white blood cell types using the Naïve Bayes Classifier algorithm[8].

Segmentation of the cell nucleus can be performed through the use of the C–Y color space[10]. The C–Y color space serves to transform an RGB image of WBCs into a YCbCr image. The white-blood-cell nucleus can then be segmented based on luminance (Y) 10]. WBC segmentation with other methods has also been used, namely, Active Contour Without Edge[11]. The opening morphology operation and median filter have been used to remove noise[12]. The study presented in this paper aims to identify AML cells, especially the ones belonging to the M1 and M2 types, based on the morphology of WBCs. The proposed study involves the utilization of the C–Y color space to achieve the conversion of an RGB image into a YCbCr image[12][13]. Segmentation of the nucleus and

*Corresponding Author

WBCs is accomplished through the use of the thresholding method followed by median filtering to make the nucleus and WBC [10]. Methods to classify cell and AML types were based on naïve Bayes' classifier algorithm. The use of this method is favored because it includes an algorithm that makes use of training data to identify similarities with test data [14][15]. Additionally, the naïve Bayes classifier method has previously been used to identify cell types from digital images, which is an added advantage[8]. The Naïve Bayes Classifier is also used to classify texture images from the Describable Textures Dataset (DTD) and Brodatz albums. The classification results show very accurate results[16]. In another study, Naïve Bayes was used to classifying tomatoes into three classes, namely raw, ripe and rotten based on the histogram characteristics, the experimental results obtained an accuracy of 76%[17]. Naïve Bayes is also used to classify the type of hepatitis from the results of blood smear images using the Fuzzy C-Means Clustering and random forest methods. The segmented image is subjected to feature extraction with SITCA (Spatio-Temporal Independent Component Analysis) which extracts every required feature. The experimental results show an accuracy of 89%[18]. Other research on image classification was also carried out using the Naïve Bayes method, based on the histogram gray feature, SIFT feature, SURF feature, and dataset dimension reduction from the image data set. The Naive Bayesian method is used to obtain the accuracy, recall, and F1 values of the image for each feature. The analysis is done by comparing the features with Naïve Bayes. The evaluation results show that the image representation using the SURF feature description can achieve better classification results[19].

II. MATERIAL AND METHOD

The data used in this study is a WBC image that identified leukemia type AML M1 and AML M2Data in the form of a WBC image that is identified AML type AML M1 and AML M2. The process of identification is performed by and, subsequently, obtained from a clinical pathologist—RSUD Dr. Moewardi—at the Clinical Pathology Installation, Surakarta, Central Java. The data comprises 30 images of each of the M1- and M2-type AML cells. Each image was obtained through observations performed using a digital microscope at 1000 times magnification of blood-cell preparations identified as AML M1 and AML M2. Sample data images, used in this study, are shown in Fig. 1.

Acute myeloblastic leukemia with little maturation (AML M1) possesses a myeloblast cell count of more than 90% and is found in the spinal cord with fine chromatin and a prominent nuclear form at locations where azurophilic granules (blue color) may be present[20} Myeloblastic cells predominate in the AML M1 cell type.

In contrast to AML M1, AML M2 is more common in children to the extent that approximately 30–45% of all AML cases in children belong to this type. In terms of cell types, AML M2 is comprised of more than myeloblast cells (more than 20% of the total cell count), found in the blood and spinal cord, and neutrophil cells (roughly 10% of total cell count) in various stages of maturity, such as Promyelocyte, Myelocyte, and Metamyelocyte. The cytoplasm of the myeloblast cells

may or may not be comprised of azurophilic grains and Auer stems[20]. Morphological features of the dominant cells in AML M1 and M2 are listed in Table I.
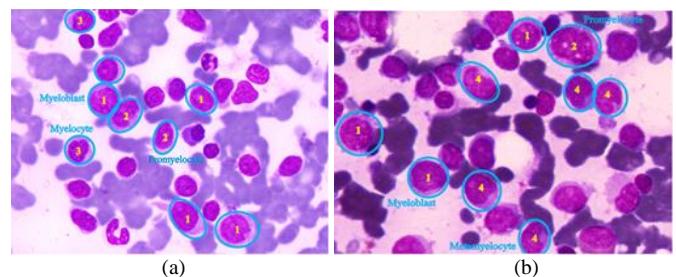
Fig. 1. Identification Type of AML M1 & AML M2 by Expert; (a) AML M1, (b) AML M2.

Information:

1) Myeloblast cells
2) Promyelocyte cells
3) Myelocyte cells
4) Metamyelocyte cells

The selection of features in the form of WBC diameter, nuclear ratio, and roundness ratio because they are the dominant features found in every granulocyte cell (Myeloblast, Promyelocyte, Myelocyte, Metamyelocyte) besides being able to see whether the cytoplasm is granular or not[21][6], as shown in Table I. However, to determine whether the cytoplasm is granular (azurophilic) or not, it is quite difficult to do because it is in the form of spots, some of which are pink, dark red, and some are purplish in the cytoplasm. it is difficult to separate the colors, so the most feasible feature was selected by measuring WBC diameter, nucleus ratio, and cell nucleus sphericity ratio in feature extraction on image data.

TABLE I. FEATURES OF DOMINANT CELLS OF AML M1 & AML M2

Cell Types	WBC Diameter (µm)	Nucleus Ratio	Nucleus Roundness
Myeloblast	15 – 20	7:1 – 5:1	Roundness
Promyelocyte	12 – 24	5:1 – 3:1	Roundness
Myelocyte	10 – 18	2:1 – 1:1	Oval
Metamyelocyte	10 – 18	1.5:1 – 1:1	Curve

Image enhancement is performed to reduce noise by converting the captured RGB image into a YCbCr image in the C–Y color space[22].

Post conversion, the YCbCr image is processed through mean and median filtering. The processes of mean and median filtering are performed to obtain a more solid image and reduce the noise that may be induced during image segmentation[23].

Image segmentation is performed to detect and segregate nuclei from WBCs by subjecting the YCbCr image to a thresholding operation based on the Y, Cr, and Cb components. The result of image segmentation is a binary image. Segmentation results are subsequently subjected to an opening operation to reduce unnecessary noise during the feature-extraction process.

The feature-extraction process relates to the quantization of image characteristics into numerical values. The methods used in this phase include those based on chain code and the bounding box algorithm. The characteristics of WBCs sought in this study are the WBC diameter, nuclear ratio, and roundness ratio [24].

1) WBC diameter The WBC diameter can be calculated based on the area of the detected WBC area and can be calculated using the following equation (1)[25]:

$$Diameter = 2\sqrt{\frac{WBC\ Area}{\pi}} \quad (1)$$

2) Nucleus ratio This is a ratio that compares the area of the nucleus with the WBC area and can be calculated using the following equation:

$$Nucleus\ Ratio = \frac{Nucleus\ Area}{WBC\ Area} \quad (2)$$

3) Roundness ratio The roundness ratio of serves to quantify the curvature of a nucleus. Its value approaching unity implies that the nucleus has a high curvature. The roundness ratio can be calculated using the following equation:

$$Roundness\ Ratio = \frac{4\pi L_{Nucleus}}{Perimeter^2} \quad (3)$$

where $L_{Nucleus}$ represents the area of the nucleus and perimeter represents the number of pixels from the edge of the nucleus.

The purpose of this step is to identify the cell and AML types. As mentioned earlier, the naïve Bayes classifier is the algorithm used in this process. Inputs for identification of the cell type include WBC diameter, nuclear ratio, and roundness ratio, while that for identification of the AML type includes number of dominant cells—myeloblast, promyelocyte, myelocyte, and metamyelocyte. The naïve Bayes' classifier method employs the Bayes' theorem, wherein laws of probability determine a possible outcome of the classification process. This method approach uses probability as a determinant of the possible outcomes of the classification process. Bayes' theorem is stated by the following formula.

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (4)$$

The Bayes classifier assumes attributes that have independent distributions. For this reason, there is the following formula:

$$P(H|X) = P(X_1|H)P(X_2|H)P(X_3|H) \dots P(X_n|H) \quad (5)$$

Where X is data with an unknown class. H is hypothesis data is a specific class. $P(H|X)$ is the probability of hypothesis H based on condition X (posteriori probability). P(H) represents Hypothesis probability H (prior probability). $P(X|H)$ Probability of X based on the conditions on the hypothesis H, and $P(X)$ is Probability of X.
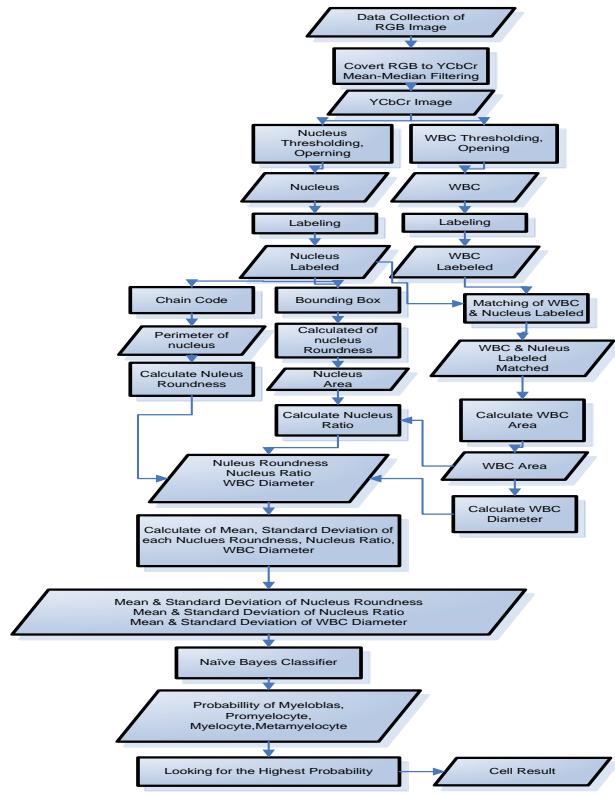
Fig. 2. The Research Methods.

The research methods of the steps can be shown in Fig. 2, from data collection, preprocessing, segmentation, characteristic extraction, and classification process with naive Bayes classifier, to find identified cell results.

III. RESULT AND DISCUSSION

A. Results

As previously mentioned, a YCbCr image is obtained as a result of image enhancement and subsequent mean and median filtering. Fig. 2 depicts the result of image enhancement performed on the RGB image depicted in Fig. 1(a). Conversion of the RGB image to the YCbCr color space is performed to obtain optimum reproduction of the red-colored image components since red is a dominant color in WBC images. The red component in the YCbCr image is the Cr component, while the RGB image is the R component. Differences in the reproduction of red-colored components in the RGB and YCbCr color spaces can be recognized as shown in Fig. 4.

The result of the nucleus segmentation performed on the YCbCr image shown in Fig. 3 is depicted in Fig. 5(a) whereas WBC segmentation is shown in Fig. 5(b).

The feature-extraction process begins with the process of labeling the segmented images of the nuclei and WBCs. This is followed by the selection process performed on the segmented nucleus and WBC images using the bounding box method. Fig. 6(a) depicts results of the selection process performed on images of the nucleus, Fig. 6(b) shows selection results for WBC images, and Fig. 6(c) depicts the result of WBC selection

with the label. The characteristics of the extraction result shown in Fig. 6(c) are listed in Table II.

The attributes of feature extraction—WBC diameter, nuclear ratio, and roundness ratio—are used as test data. Based on the above feature extraction result, AML image characteristics (cell 6 in Fig. 6(c)) have WBC diameter, 18.161 µm; nucleus ratio, 0.647; and roundness ratio, 0.706. The percentage similarity of the cell types obtained through calculations based on the naïve Bayes' classifier algorithm is as quoted, Myeloblast = 40.1%, Promyelocyte = 39.1%, Myelocyte = 7.6%, Metamyelocyte = 13.2%.

Cell 6 in Fig. 6(c) is identified by the system as a myeloblast cell, because it has the highest percentage of 40.1%. A comparison of the results of cell type identification performed by a laboratory expert and that corresponding to Fig. 6(c) performed by the proposed system is presented in Table III.

As can be seen in Table III, there exist two cells—3 and 8—for which the result obtained using the proposed method did not match the identifications made by the expert. Cell 3 with characteristics of WBC diameter, 13.035 µm; nuclear ratio, 0.768; and roundness ratio of 0.815 is identified as a myeloblast cell because it has the following similarity percentages, Myeloblast = 61.2%, Myelocyte = 25.4%, Promyelocyte = 8.6%, Metamyelocyte = 4.8%. The WBC diameter of cell 3 demonstrates similarity percentages of 54.5%, 14.8%, 22.7%, and 8% with myeloblast, promyelocyte, myelocyte, and metamyelocyte cells, respectively. Thus, the proposed method identifies cell 3 as a myeloblast cell.

Fig. 3. Image Enhancement Result.
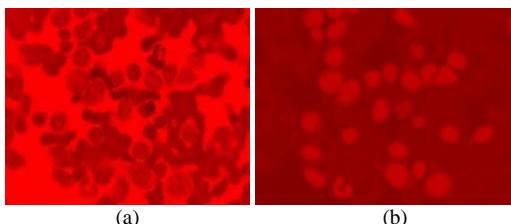
Fig. 4. The Difference of the Red Components in RGB Image and YCbCr Image(a) The Red Components (R) in RGB Image (b) The Red Components (Cr) in YCbCr Image.
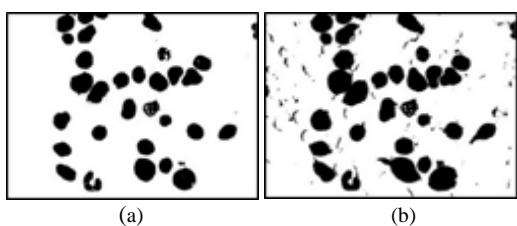
Fig. 5. Segmentation Result (a) Nucleus Segmentation, (b) WBC Segmentation.
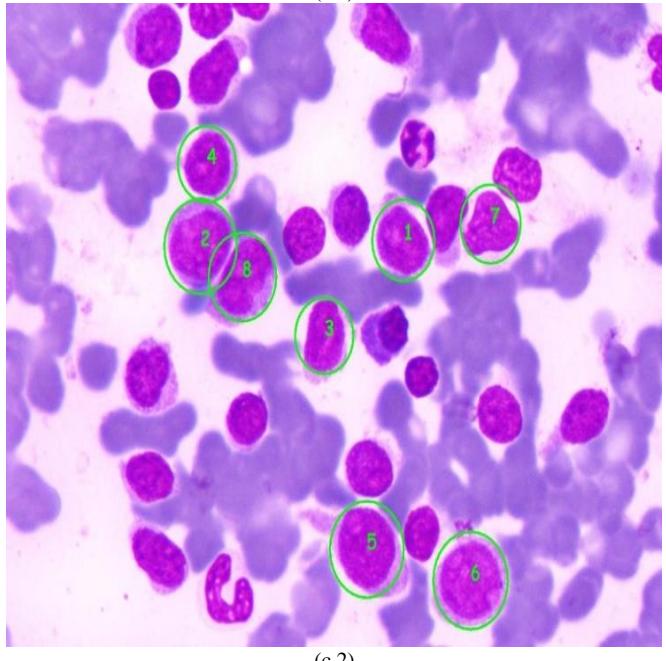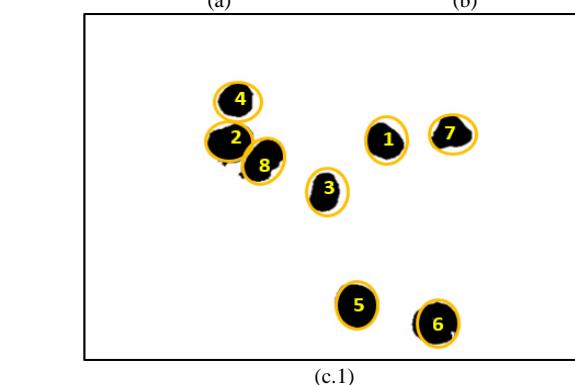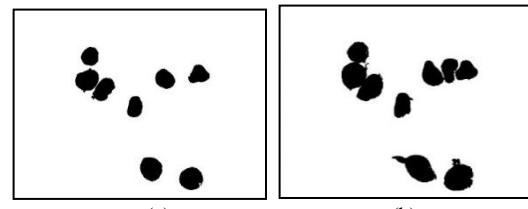
Fig. 6. Selection Result of Nucleus and WBC (a) Selection Result of Nucleus, (b) Selection Result of WBC (c.1.c2) WBC Labelling.

TABLE II.   RESULTS OF CELL-FEATURE EXTRACTION

Label	Wbc Diameter (µm)	Nucleus Ratio	Nucleus Roundness
1	14.161	0.7428	0.866
2	17.306	0.676	0.720
3	13.035	0.768	0.815
4	13.552	0.718	0.869
5	18.212	0.640	0.855
6	18.161	0.647	0.706
7	12.987	0.833	0.771
8	16.569	0.645	0.621

TABLE III. COMPARISON OF CELLS IDENTIFICATION PERFORMED BY AN EXPERT AND THE PROPOSED SYSTEM

Label	Cell Identification by System	Cell Identification by Expert	Explanation
1	Myeloblast	Myeloblast	Matched
2	Myeloblast	Myeloblast	Matched
3	Myeloblast	Promyelocyte	Not matched
4	Myeloblast	Myeloblast	Matched
5	Myeloblast	Myeloblast	Matched
6	Myeloblast	Myeloblast	Matched
7	Myeloblast	Not Identified	Not Identified
8	Myeloblast	Promyelocyte	Not matched

The nucleus ratio of cell 3 demonstrates similarity percentages of 62.3%, 13.85%, 12.35%, and 11.5% with myeloblast, promyelocyte, myelocyte, and metamyelocyte cells, respectively. The nucleic feature of cell 3 demonstrates similarity ratios of 55.7%, 24.1%, 14.9%, and 5.3% with myeloblast, promyelocyte, myelocyte, and metamyelocyte cells, respectively. So, the nucleus ratio and roundness ratio of cell 3 has similarities to myeloblast characteristics.

Cell 8 with characteristics of WBC diameter, 16.569 μm; nuclear ratio, 0.645; and roundness ratio of 0.621 is identified as a myeloblast cell because it has the following similarity percentages, Myeloblast = 46.4%, Myelocyte = 11.3%, Promyelocyte = 23.3%, Metamyelocyte = 19%.

The WBC diameter of cell 8 demonstrates similarity percentages of 59.2%, 18%, 14.4%, and 8.4% with myeloblast, promyelocyte, myelocyte, and metamyelocyte cells, respectively. As such, the feature of WBC diameter for cell 8 has a strong resemblance to myeloblast cells. The nucleus ratio corresponding to cell 8 demonstrates similarity percentages of 44.5%, 31.3%, 14.1%, and 10.1% with myeloblast, promyelocyte, myelocyte, and metamyelocyte cells, respectively. At the same time, the nucleic feature of 8 demonstrates similarity percentages of 49.3%, 21.7%, 8.3%, and 20.7% with myeloblast, promyelocyte, myelocyte, and metamyelocyte cells, respectively. Thus, cell 8 has been assumed to possess a strong resemblance to myeloblast cells. The differences in the characteristics of WBC diameter, nuclear ratio, and roundness ratio between myeloblast cells and neutrophil cells (promyelocyte cells, myelocyte cells, and metamyelocyte cells) are almost indistinguishable in the above cases since each feature has an uncertain range of values. For example, cell 3 is identified by the expert as a promyelocyte cell, whereas cell 4, which possesses characteristic values similar to those of cell 3, is identified by the specialist as a myeloblast cell.

Results of the cell type feature extraction comprise 264 cell type data obtained from 50 AML images. Based on identifications made by the laboratory expert, these 264 cells comprised 143 myeloblasts, 62 promyelocytes, 33 myelocyte, and 26 metamyelocyte cells.

Table IV presents a comparison of these results with those obtained from cell type identification performed by the proposed system.

TABLE IV. RESULTS OF CELL TYPE IDENTIFICATION PERFORMED BY EXPERTS AND THE PROPOSED SYSTEM

Identification By Expert	Identification By System			
	Myeloblast (A)	Promyelocyte (B)	Myelocyte (C)	Metamyelocyte (D)
Myeloblast (A)	125	13	0	5
Promyelocyte (B)	41	20	0	1
Myelocyte (C)	31	2	0	0
Metamyelocyte (D)	20	6	0	0

An example of testing myeloblast cells using a confusion matrix based on Table IV. TP (True Positive): The number of myeloblast cells detected by the system and experts as myeloblast cells is 125 cells. FN (False Positive): The number of myeloblast cells identified by the expert but detected by the system as other cells (promyelocyte, myelocyte, and metamyelocyte) is $13 + 5 = 18$ cells. FP (False Positive): The number of other cells (promyelocyte, myelocyte, and metamyelocyte) identified by the expert but detected by the system as myeloblast cells is $41 + 31 + 20 = 92$ cells. TN (True Negative): The number of other cells (promyelocyte, myelocyte, and metamyelocyte) identified by the expert and the system as other cells (promyelocyte, myelocyte, and metamyelocyte) is $20 + 0 + 1 + 2 + 0 + 0 + 6 + 0 + 0 = 29$ cells.

$$Precision = \frac{TP}{TP+FP} = \frac{125}{125+92} = 0.576$$

$$Sensitivity = \frac{TP}{TP+FN} = \frac{125}{125+18} = 0.8741$$

$$Spesificity = \frac{TN}{TN+FP} = \frac{29}{29+92} = 0.2396$$

The analysis of each cell is done by using a confusion matrix with precision, sensitivity, and specificity of each type of cell test. The results of confusion matrix testing can be seen in Table V. As seen in Table V, the precision of the myeloblast cells is 57.6% implying that the number of myeloblast cells was correctly identified as 57.6% of the total cell count. Sensitivity is the level of success achieved by the system in rediscovering information. As seen, the sensitivity of myeloblast cells is 87.41%, which means that 87.41% of the actual number of myeloblast cells was correctly identified by the system.

TABLE V. CONFUSION MATRIX TEST RESULT

Cell Type	Precision	Sensitivity	Specificity
Myeloblast	57.6 %	87.41 %	23.96 %
Promyelocyte	48.78 %	32.25 %	90.00 %
Myelocyte	0	0	100.00 %
Metamyelocyte	0	0	97.47 %
Overall Cells	54.92 %	54.92 %	85.14 %

Each cell type in the proposed study was subjected to a different mean test to examine the differences in the characteristics of WBC diameter, nucleus ratio, and roundness ratio between each cell type. The testing was performed using the T-test to know the significant value (p-value) of each cell.

An error rate of 5% was set meaning that if the p-value exceeds 0.05, it implies that the character of the compared cell is significantly different. The mean difference of the test results corresponding to each cell is listed in Table VI.

The WBC diameter of myeloblast cells, in theory, measures between 15–20 μm, and that of promyelocyte cells measures between 12–24 μm. Myeloblast and promyelocyte cells have similar typical WBC diameters between 15–20 μm. However, this similarity is not identified by the mean difference test result, which states that WBC diameters of myeloblast and promyelocyte cells are significantly different. Similarly, the nuclear ratio of the myeloblast cell, in theory, ranges from 7:1–5:1, while that of promyelocyte cells lies in the range of 5:1–3:1. This is in line with the mean difference test result that the nuclear ratios of the two cell types differ significantly.

The mean test result for the roundness ratio demonstrates that there exists no significant difference between the two cell types, which corresponds to the theoretical basis that myeloblast and promyelocyte cells have a still-filled nucleus.

K-fold cross-validation of the cell type data was performed to determine the accuracy of naive Bayes' classifier algorithm applied to new data. The test was performed using the Leave One Out Cross Validation (LOO-CV) method on the naive Bayes' classifier algorithm using the experimental values k = 2 to k = 10. The test results demonstrate that the algorithm possessed the highest accuracy at k = 4 with an accuracy of 54.92%. The complete cross-validation test results are presented in Table VII.

Out of the 60 AML images that were processed using the proposed technique, 44 were identified correctly. A summary of the results of AML identification can be seen in Table VIII.

The obtained accuracy results of the AML type identification are as follows:

$$Accurate \ of \ System = Acc = \frac{The \ amount \ of \ data \ is \ correct}{Total \ data \ tested}$$

$$= \frac{44}{60} \times 100\% = 73.33\%$$

B. Discussion

Many studies on the diagnosis of leukemia based on image processing have been carried out, including research on ALL (Acute Myeloid Leukemia) classification using the Naïve Bayes Classifier method on WBC segmentation results with k-NN (k-Nearest Neighbor), with a calcification accuracy of 75% [8]. Another study was conducted to classify AML M0 and AML M1 on the results of WBC segments with RGB to YCbCr conversion using the k-NN classification method, obtaining an accuracy of 59.87% [12]. Then the classification of AML M2 and AML M3 diseases with the Momentum backpropagation method from the results of WBC segmentation with the Watershed Distance Transform method obtained an accuracy of 94.285% [26]. While the classification on AML M1. M2 and M3 using the Backpropagation Momentum Method from the results of WBC segmentation with the ACWE method obtained an average precision of 84.754%, Sensitivity 75.88%, specificity 95.090%, and accuracy 94.285% [11]. In the same year, cell classification was carried out in AML M4, AML M5, and AML M7 with the

Support Vector Machine method on the results of WBC segmentation with K-NN and Watershed Distance Transform. The cell types were myeloblast, promyelocyte, granulocyte, monoblast, promonocyte, monocyte, megakaryoblast. and supporting cells with accuracy of 98.67%, 98.01%, 84.05% 99.67%, 95.35%, 89.70%, 99.34% and 98.01%, respectively [27]. The classification of AML M0 and AML M1 diseases has also been carried out using the Naïve Bayes Classification method from the results of WBC segmentation with Multi-Otsu Thresholding compared to Static Thresholding, respectively, the accuracy is 83.81% and 75.35% [28]. Then a comparison of the accuracy of the segmentation results on WBC in AML MI obtained an accuracy of 90.67% with the ACWE segmentation method, the results are higher than the accuracy obtained with the Sheet Region Growing and Otsu Thresholding segmentation [29].

Another study was also conducted regarding the classification of leukocyte cells in AML with the random forest method from the WBC segmentation process with Multi-Otsu Thresholding, which obtained 93.45% classification accuracy and 65% precision [30]. Then the classification is carried out on the case for ALL classification using the Naïve Bayes Classification method from the results of WBC segmentation with thresholding, 80% accuracy is obtained [31]. From the results of several studies that have been carried out for the classification of AML from the segmentation results of WBC cells in Acute Myeloid Leukemia, it is true that for this study the precision, sensitivity, specificity, and accuracy are relatively small compared to other similar studies that have been carried out.

TABLE VI.   MEAN DIFFERENCE TEST OF EACH CELL TYPES

The Type of Cells That Are Compared	P-Value		
	WBC Diameter	Nucleus Ratio	Roundness Ratio
Myeloblast – Promyelocyte	0.000	0.000	0.105
Myeloblast – Myelocyte	0.008	0.169	0.922
Myeloblast – Metamyelocyte	0.092	0.515	0.001
Promyelocyte – Myelocyte	0.000	0.052	0.188
Promyelocyte – Metamyelocyte	0.468	0.014	0.010
Myelocyte – Metamyelocyte	0.002	0.542	0.002

TABLE VII.   IDENTIFICATION RESULT OF CELL TYPES WITH K-FOLD CROSS VALIDATION TEST

K	2	3	4	5	6	7	8	9	10
Correct	144	142	145	144	141	140	144	143	143
Acc	54.55%	53.79%	54.92%	54.55%	53.41%	53.03%	54.55%	54.17%	54.17%

TABLE VIII.   IDENTIFICATION RESULT OF AML

Type of Aml	Identification By System	
	AML M1	AML M2
AML M1	29	1
AML M2	15	15

IV. Conclusion

From the above analysis and supporting discussion, it can be inferred that the naïve Bayes' classifier algorithm could be used in the identification of dominant cell AML types—myeloblast, promyelocyte, myelocyte, and metamyelocyte—based on WBC morphological imaging. The AML type identification performed in this study using the naive Bayes classifier algorithm demonstrated a system accuracy of 73.33% in a sample space comprising 60 AML images. Additionally, cell type identification accuracy of 54.92% was achieved in a sample space comprising 264 cell type data. The above accuracy was achieved with a precision of 54.92%, sensitivity of 54.92%, and specificity of 85.14%. As part of a future endeavor, the authors suggest normalization of the contrast in the image before the segmentation process to reduce noise generated in the segmentation process.

Acknowledgment

References

[1] Kumar et al., "Automatic Detection of White Blood Cancer From Bone Marrow Microscopic Images Using Convolutional Neural Networks," IEEE Access, vol. 8, pp. 142521–142531, 2020, doi: 10.1109/ACCESS.2020.3012292.

[2] A. Vakiti and P. Mewawalla, "Acute Myeloid Leukemia," 2021.

[3] American Cancer Society. (2013). Leukemia: Acute Myeloid (Myelogenous) Overview. American Cancer Society.

[4] College of American Pathologists. (2011). Blood Cancer Acute Myeloid Leukemia. College of American Pathologists.

[5] B. Houwen, "The Differential Cell Count," Laboratory Hematology, vol. 7, pp. 89–100, 2001.

[6] A. Bell and S. Sallah, The Morphology of Human Blood Cells: Seventh Edition, Seventh. English: Bookbaby, 2020.

[7] Mitchell, Kumar, Abbas, and Fausto, Robbins amp Cotran Buku Saku Dasar Patologis Penyakit Edisi 7 Mitchell. Elseiver Inc, 2009.

[8] S. Selvaraj and B. Kanakaraj, "Naïve Bayesian classifier for Acute Lymphocytic Leukemia detection," ARPN Journal of Engineering and Applied Sciences, vol. 10, no. 16, pp. 6888–6892, 2015.

[9] E. Suryani, W. Wiharto, and N. Polvonov, "Identification and Counting White Blood Cells and Red Blood Cells using Image Processing Case Study of Leukemia," International Journal of Computer Science & Network Solutions, vol. 2, no. 6, pp. 35–49, 2015.

[10] R. Mohammed, O. Nomir, and I. Khalifa, "Segmentation of Acute Lymphoblastic Leukemia Using C-Y Color Space," ijacsa, vol. 5, no. 11, 2014, doi: 10.14569/IJACSA.2014.051117.

[11] A. Harjoko, T. Ratnaningsih, E. Suryani, Wiharto, S. Palgunadi, and N. P. T. Prakisya, "Classification of acute myeloid leukemia subtypes M1, M2 and M3 using active contour without edge segmentation and momentum backpropagation artificial neural network," in MATEC Web of Conferences, Yogyakarta, 2018, vol. 154, p. 01041. doi: 10.1051/matecconf/201815401041.

[12] E. Suryani, Wiharto, S. Palgunadi, and Y. R. Putra, "Cells identification of acute myeloid leukemia AML M0 and AML M1 using K-nearest neighbour based on morphological images," Proceedings of 2017 International Conference on Data and Software Engineering, ICoDSE 2017, vol. 2018-Janua, pp. 1–6, 2018, doi: 10.1109/ICODSE.2017.8285851.

[13] W. Wiharto, E. Suryani, and Y. R. Putra, "Classification of blast cell type on acute myeloid leukemia (AML) based on image morphology of white blood cells," TELKOMNIKA (Telecommunication Computing Electronics and Control), vol. 17, no. 2, p. 645, 2018, doi: 10.12928/telkomnika.v17i2.8666.

[14] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques., Third Edition. USA: Morgan Kaufmann is an imprint of Elsevier, 2012.

[15] H. Hamidi and A. Daraei, "Analysis of Pre-processing and Post-processing Methods and Using Data Mining to Diagnose Heart Diseases," IJE, vol. 29, no. 7, pp. 921–930, Jul. 2016, doi: 10.5829/idosi.ije.2016.29.07a.06.

[16] A. M. Mansour, "Texture Classification Using Naive Bayes Classifier," IJCSNS International Journal of Computer Science and Network Security, VOL.18 No.1, January 2018.

[17] A. Kusuma, D. R. I. M. Setiadi, and M. D. M. Putra, "Tomato Maturity Classification using Naive Bayes Algorithm and Histogram Feature Extraction," Journal of Applied Intelligent System (e-ISSN : 2502-9401 | p-ISSN : 2503-0493), Vol. 3 No. 1, August 2018, pp. 39 – 48

[18] V. Vanitha and D. Akila, "Image Segmentation and classification Hepatitis viral infection in human blood smear with a hybrid algorithm combining Naive Bayes Classifier," Turkish Journal of Computer and Mathematics Education, Vol.12 No.11 (2021), 5873-5881.

[19] K. Fang, "Naive Bayes Image Classification Based on Multiple Features," Computer Software and Media Applications (2019), Volume 2, doi: 10.24294/csma.v2i1.1171.

[20] A. Kaur and B. V. Kranthi, "Comparison between YCbCr Color Space and CIELab Color Space for Skin Color Segmentation," International Journal of Applied Information Systems, vol. 3, no. 4, pp. 30–33, 2012.

[21] College of American Pathologists. (2016). Hematology and Clinical Microscopy Glossary. College of American Pathologists.

[22] A. Bacarea, "Diagnosis of Acute Myeloid Leukaemia," in Myeloid Leukemia - Clinical Diagnosis and Treatment, S. Koschmieder, Ed. InTech, 2012. doi: 10.5772/26177.

[23] M. A. Fattahzadeh and A. Saghaei, "A Statistical Method for Sequential Images–based Process Monitoring," IJE, vol. 33, no. 7, pp. 1285–1292, Jul. 2020, doi: 10.5829/ije.2020.33.07a.15.

[24] M. Ghaderzadeh, F. Asadi, A. Hosseini, D. Bashash, H. Abolghasemi, and A. Roshanpour, "Machine Learning in Detection and Classification of Leukemia Using Smear Blood Images: A Systematic Review," Scientific Programming, vol. 2021, pp. 1–14, Jun. 2021, doi: 10.1155/2021/9933481.

[25] E. S. A. Baquero, "Methodology for Automatic Classification of Atypical Lymphoid Cells from Peripheral Blood Cell Images," Disertation, Universitat Politècnica de Catalunya • BarcelonaTech, Barcelona, 2015.

[26] E. Suryani, Wiharto, S. Palgunadi, and N. P. T. Prakisya, "Classification of Acute Myelogenous Leukemia (AML M2 and AML M3) using Momentum Back Propagation from Watershed Distance Transform Segmented Images," International Conference on Computing and Applied Informatics 2016, IOP Conf. Series: Journal of Physics: Conf. Series 801 (2017) 012044, IOP Publishing, doi:10.1088/1742-6596/801/1/012044.

[27] A. Setiawan 1, A. Harjoko2, T. Ratnaningsih, E. Suryani, Wiharto, S., and Palgunadi, "Classification of Cell Types In Acute Myeloid Leukemia (AML) of M4, M5 and M7 Subtypes With Support Vector Machine Classifier," 2018 International Conference on Information and Communications Technology (ICOIACT), 978-1-5386-0954-5/18/$31.00 ©2018 IEEE.

[28] E Suryani, E I Asmari, and B Harjito, "Image Segmentation of Acute Myeloid Leukemia Using Multi Otsu Thresholding," ICERIA 2020, Journal of Physics: Conference Series 1803 (2021) 012016, IOP Publishing, doi:10.1088/1742-6596/1803/1/012016.

[29] N. P. T. Prakisya and A. Setiawan, "A Comparative Study of Digital Image Segmentation Algorithms for Acute Myeloid Leukemia M1 White Blood Cells Images," Indonesian Journal of Informatics Education (IJIE), ISSN: 2549-0389.

[30] S. Dasariraju, M. Huo, and S. McCalla, "Detection and Classification of Immature Leukocytes for Diagnosis of Acute Myeloid Leukemia Using Random Forest Algorithm," Bioengineering 2020, 7, 120; doi:10.3390/bioengineering7040120.

[31] S. Shidada, B. Hariyanto, "Identifikasi Acute Lymphoblastic Leukemia pada Citra Mikroskopis Menggunakan Algoritma Naïve Bayes," Journal of Science and Technology, https://journal.trunojoyo.ac.id/rekayasa, Rekayasa, 2021; 14(1): 78-83, ISSN: 0216-9495 (Print), ISSN: 2502-5325 (Online).