*International Journal on Recent and Innovation Trends in Computing and Communication*
*ISSN: 2321-8169 Volume: 11 Issue: 1*
***Article Received:*** *25 September 2022* ***Revised:*** *12 November 2022* ***Accepted:*** *30 December 2022*

# Automated Leukemia Disease Classification using Machine Learning on Microscopic Blood Images

**I.Vinurajan[1], Dr. K. P. SanalKumar[2], Dr. S. Anu HNair[3]**
[1]Research Scholar, Department of Computer and Information Sciences, Annamalai University, Chidambaram, India.
[2]Assistant Professor, PG Department of Computer Science, R.V Government Arts College, Chengalpattu, India.
[3]Assistant Professor, Department of CSE, Annamalai University, Chidambaram, India [Deputed to WPT, Chennai].

***Abstract-*** Leukemia is a pathology that affects teenagers and adults, which leads to several other symptoms and premature death. Computer-aided system (CAD) is used to assist specialists in the diagnosis of this disease and lower the risk of prescribing improper treatment. Microscopic analysis is an efficient strategy for performing the initial screening of patients with leukemia. This kind of test can be manually done, generating fatigue in operators. Consequently, an economical method that is robust and automatic is needed to prevent the influence of operators. Several CAD systems were designed by using computational intelligence and image processing methods. In this paper, we present an Automated Leukemia Disease Classification utilizing Machine Learning on Microscopic Blood Images (ALDC-MLMBI) method. The ALDC-MLMBI technique aims to employ ML approaches for the identification of leukemia. As a primary step, the ALDC-MLMBI technique follows median filtering (MF) based noise elimination and adaptive histogram equalization (AHE) based contrast enhancement. Besides, the segmentation process can be performed by watershed segmentation. Meanwhile, the ALDC-MLMBI technique involves a set of feature extractors namely local binary pattern (LBP), histogram of gradients (HOG), scale-invariant feature transform (SIFT), and gray level co-occurrence matrix (GLCM). Furthermore, the classification of leukemia can be made by the use random forest (RF) method. The simulation analysis of the ALDC-MLMBI system can be performed using the Kaggle image dataset. The experimental outcomes highlighted the superior performance of the ALDC-MLMBI system compared to existing classifiers.

***Keywords:*** Leukemia Disease; Machine Learning; Random Forest; Segmentation; Feature Extractor; Image Processing

## 1. Introduction

Leukemia is a kind of blood cancer that emerges in bone marrow cells and causes a vast quantity of abnormal blood cells [1]. These blood cells, commonly referred to as leukemia cells or blasts, are not full-grown. A lack of normal blood cells causes bruising, bleeding, weariness, bone discomfort, a higher risk of infection, and fever. A bone marrow biopsy or blood tests are commonly employed to make the diagnoses [2]. Researchers do not know the real cause of leukemia. It seems to be the result of a mixture of genetic and environmental factors. The delicate discernment of malignant leukocytes with low expense in the earlier stage of the disorder is an elementary issue in the realm of illness diagnostics [3]. Flow cytometry tools are in constrained supply, and the procedure available in laboratory diagnostic centers is time-consuming. Leukaemia is the prevalent form of blood cancer in all age group of people, particularly youngsters [4]. Immature growth and excessive proliferation of blood cells cause this abnormal phenomenon that may damage the immune system, red blood cells, and bone marrow. Malignant white blood cells also called lymphoblast, pass over the blood to other organs namely kidneys, spleen, liver, and brain where they spread to vital bodily tissues [5]. Based on the microscopic pictures, a hematologist in cell transplant facilities can diagnose and differentiate different types of leukemia. If the slide is properly stained, a few kinds of leukemia are easier to distinguish and identify than others, but defining leukemia needs advanced technology [6].

Recently, a great deal of research has implemented computer-aided diagnostic and machine learning (ML) approaches for laboratory image analysis, aiming to resolve the limitation of a late leukemia analysis and define its subclasses [7]. This study has investigated blood smear images for diagnosing, counting, and differentiating the cells in a variety of leukemia. ML is an eminent sector of artificial intelligence (AI), including mathematical relations and algorithms [8]. ML allows computers to be programmed without detailed understanding. The outcomes of using this method in medical data processing have made remarkable success, and they have been extraordinary in disease diagnoses [9]. The study shows that ML approaches enormously facilitate the medical decision-making process by analyzing and extracting the features. There is a pressing need for high-quality data analysis methods as the number of more advanced data was produced and a large volume of medical diagnosis tools

*International Journal on Recent and Innovation Trends in Computing and Communication*
*ISSN: 2321-8169 Volume: 11 Issue: 1*
***Article Received:*** *25 September 2022* ***Revised:*** *12 November 2022* ***Accepted:*** *30 December 2022*

increased [10]. Classical techniques cannot find data patterns or analyze an enormous volume of data.

This study presents an Automated Leukemia Disease Classification using Machine Learning on Microscopic Blood Images (ALDC-MLMBI) method. As a primary step, the ALDC-MLMBI technique follows median filtering (MF) based noise elimination and adaptive histogram equalization (AHE) based contrast enhancement. Besides, the segmentation process can be performed by watershed segmentation. Meanwhile, the ALDC-MLMBI technique involves a set of feature extractors namely histogram of gradients (HOG), local binary pattern (LBP), scale-invariant feature transform (SIFT), and gray level co-occurrence matrix (GLCM). Furthermore, the classification of leukemia can be made by the use random forest (RF) method. The simulation analysis of the ALDC-MLMBI technique can be performed using the Kaggle image dataset.

## 2. Literature Review

The authors [11] projected a novel classification method for blood microscopic images that differentiate between leukemia-free and affected imageries. The projected model contains 3 foremost stages namely Image Pre-processing, Classification, and Feature Extraction (FE). An optimized CNN (OCNN) was utilized for classification. OCNN was used to identify and categorize the imagery as abnormal or normal. In [12], WBCs are divided from blood smear imageries utilizing morphological models, and the segment is explored for an array of textural, statistical, and geometrical chattels. The 4 dissimilar ML methods are utilized to inspect the performance of numerous models. This model also employed EMC-SVM to categorize leukocytes. Karthick et al. [13] show an effort can be completed to project a speedy and cost-effectual CAD method for the identification of blast cells in leukemia by utilizing digital image processing (IP) and ML techniques.

Muzalevskiy and Torshin [14] projected a novel technique to the issue of identification of such cellular imageries depending upon the XGBoost algorithm, graph theory, and CNN. At first, each imagery is altered into a weighted graph utilizing a gradient of strength. Second, the amount of graph invariants is calculated by creating a group of synthetic features which is employed to train ML technique depending upon XGBoost. Karthikeyan et al. [15] project an ML technique for identifying leukemia in patients. A database containing blood smear imageries has been composed and pre-processed to remove related features. The features were employed to train and verify numerous ML classification models.

Rangini et al. [16] deliver a strong model for enhancing the consistency of ALL classification utilizing feature extraction and an ML method to distinguish between tumorous and healthy cells in imageries. To remove texture features for diagnosis, the First Order Histogram (FOH) and the GLCM techniques are used in this model. LR, NB, and ANN are applied in the categorizing method. Rupapara et al. [17] proposed a technique for blood cancer disease forecasting utilizing the supervised ML model. ADASYN resampling and Chi-squared (Chi2) models were utilized to solve excessive and higher-dimension dataset issues. ADASYN produces fake data to create the database balanced for every objective class, and Chi2 picks the finest features to train learning methods.

## 3. Proposed Methodology

In this study, we have presented a new ALDC-MLMBI method. The ALDC-MLMBI technique aims to employ ML approaches for the identification of leukemia. To accomplish that, the ALDC-MLMBI method has image preprocessing, segmentation, feature extraction, and classification processes. Fig. 1 exemplifies the complete procedure of the ALDC-MLMBI technique.

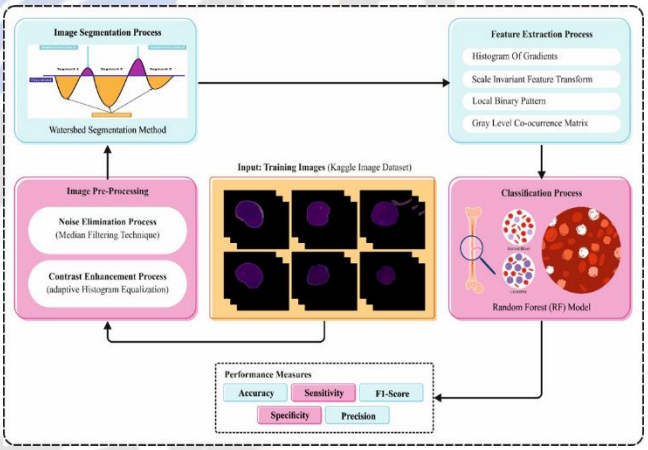


**Fig. 1.** Overall process of ALDC-MLMBI method

### 3.1. Image Preprocessing

As a primary step, the ALDC-MLMBI technique follows MF-based noise elimination and AHE-based contrast enhancement. MF is the order-statistics filtering that replaces the pixel values with the median of the gray level in the neighbourhood of that pixel [18].

$$\hat{f}(x,y) = \{g(s,t)\} \quad (1)$$

During the median computation, the original pixel values are included. MFs are comparatively traditional for specific kinds of random noise because they offer outstanding noise

*International Journal on Recent and Innovation Trends in Computing and Communication*
*ISSN: 2321-8169 Volume: 11 Issue: 1*
***Article Received:*** *25 September 2022* ***Revised:*** *12 November 2022* ***Accepted:*** *30 December 2022*

reduction abilities, with much lower blurring than linear smoothing filters.

AHE works by dividing an image into overlapping, smaller windows or tiles, and undergoes independent histogram equalization [19]. AHE adjusts its enhancement method to the local features of the tile different from classical histogram equalization that employs a unique variation to the entire image. AHE efficiently increases the details and contrasts in regions with different contrast gradients or illumination levels by evaluating the cumulative distribution function (CDF) of pixel intensity within every tile, and redistributing these intensities based on the local CDF. This adaptive method ensures that fine details are retained while avoiding the over-amplification of noise that can take place with global histogram equalization. Accordingly, AHE is very efficient in improving visual perception and improving image quality in applications namely surveillance, medical imaging, and so on.

### 3.2. Segmentation Process

Next, the segmentation process can be performed by watershed segmentation. The most commonly used technique to address object-object limitations is to use region-growing approaches like watershed [20]. On the other hand, these approaches need object markers to be effective. Using *ad hoc* rules to extract markers needs a prior understanding of a) specific image properties, b) object locations, and c) the number of objects within an image. In these cases, the parameter governing marker extraction diverges from image to image, which motivates the usage of ML methods for strong detection of object markers. The Bayesian marker removal technique used the naive Bayes classifier for generating object markers. Unfortunately, the method does not offer any limitations to confirm that the marker for all the target objects is removed or that the removed marker lies within the object boundary because the classifier is proficient on the ground truth outlining entire objects. In general, one might threshold the $P$, probability map, through the high value for threshold $\tau$ in (2). Therefore, accuracy will enhance at the recall cost, and thus pixels corresponding to object markers are removed. To make things better, we suggest training a marker detection classifier, $h_{marker}$, on ground truth adjusted by morphological erosion. Consider

$$L_{eroded} = L \ominus B \tag{2}$$

as the destruction of label image $L$ by the structural component $B$. The output of $h_{marker}$, represented as $P_{marker}$,

$$P_{marker}(i,j) = p[L_{eroded}(i,j)|f(i,j)]$$

$$= h_{marker}\big(f(i,j)\big) \tag{3}$$

Analogous to (6), $h_{marker}$ is derived. Henceforth, we represent $h_{region}$ and $P_{region}$to make the notation distinction more prominent, the classifier trained on the resulting probability map and the ground truth. The study demonstrates that the $h_{marker}$ classifier is too conservative (viz., low recall, high precision) and generates better object markers than thresholding $P_{region}$ using the highest values of $\tau$.

### 3.3. Feature Extraction Process

In the meantime, the ALDC-MLMBI technique involves a set of feature extractors namely HOG, LBP, SIFT, and GLCM.

#### 3.3.1. HOG Model

HOG defines the local shape and appearance of the object in an imagery utilizing the allocation of gradients [21]. The typical image vector $I(x,y)$ employing the HOG model was attained by the below-mentioned process:

Step 1—Separate the image $I(x,y)$ into equivalent blocks $(N_b \times N_b)$, whereas every block covers $(M \times M)$ regular-size cells $(8 \times 8)$ pixels.

Gradient values $(G_h, G_y)$ are calculated for every pixel utilizing an aligned $1-D$ derived filter in the vertical and horizontal directions. The subsequent masks $(S_h, S_v)$ were employed and definite by Eqs. (4) and (5):

$$S_h = [-1\ 0\ 1\ ] \tag{4}$$

$$S_v = [-1\ 0\ 1\ ] \tag{5}$$

$$G_h(x,y) = I(x,y) * S_h \tag{6}$$

$$G_v(x,y) = I(x,y) * S_v \tag{7}$$

*Step* 2—The gradient orientation $(\theta)$ and magnitude $(|G(x,y))$ of every pixel $(x,y)$ are computed utilizing Eqs. (8) and (9):

$$|G(x,y)| = \sqrt{G_h^2(x,y) + G_v^2(x,y)} \tag{8}$$

$$\theta = arctan\left(\frac{G_v(x,y)}{G_h(x,y)}\right) \tag{9}$$

Here, $G_v$ and $G_h$ signify the vertical and horizontal at pixel $(x,y)$, correspondingly.

*Step* 3—The orientation histogram depending upon the gradient in every cell was computed by measuring unknown gradients at every pixel in 9 channels coordination.

*International Journal on Recent and Innovation Trends in Computing and Communication*
*ISSN: 2321-8169 Volume: 11 Issue: 1*
***Article Received:*** *25 September 2022* ***Revised:*** *12 November 2022* ***Accepted:*** *30 December 2022*

*Step* 4—Every cell is standardized utilizing a histogram in their identified blocks. We utilize L2-norm in this paper for block normalization (BN); the factor of normalization has been computed by utilizing the below-given calculation:

$$Hist_n = \frac{Hist}{\sqrt{\|Hist\|_2^2 + \varepsilon}} \quad (10)$$

Whereas, $Hist$ denotes the vector of non-normalized covering every histogram in a block; $\epsilon$ refers to a regularization term; $\|Hist\|_2$ specifies the L2 type of descriptor vector.

*Step* 5—Every block was developed by linking the histogram vector of every cell in the block. The HOG has been created by linking the feature vectors of every block for an assumed image.

### 3.3.2. LBP Model

LBP is nothing but a gray-scale and rotation-invariant texture classifier [22]. The most significant features of LBP are easiness and efficacy. Also, it is utilized in several applications like image processing and pattern detection, remote sensing, visual inspection, face and facial expression recognition, motion, and gender prediction. To describe the implementation of the LBP operator. It sets a label to every image pixel by thresholding a 3x3 neighborhood with the value of a center pixel, whereas if the value of the center pixel is larger than the value of the neighbor, then write $zero$. Or else, write $one$. Next, the outcomes of 0 and 1 were measured as binary numbers. Through analysis of the value of clockwise, the binary outcomes will be acquired. In the last code, the highest pixel of left in the neighborhood was tangled as the significant bit. This delivers a decimal label value of the 8-bit stable with intensity pixel among the neighbors and center pixel. LBP outcome signifies the vector feature. Generally, the LBP of an assumed pixel location $(x_c, y_c)$ is computed as below:

$$LBP(x_c, y_c) = \sum_{n=0}^{n} s[l_n - l_c)2^n \quad (11)$$

Whereas $l_n$ refers to the gray value of 8 surrounding pixels, $l_c$ parallels to the gray value of the center pixel $(x_c, y_c)$, and $s(k)$ was definite below:

$$s(xk) = \{1, \quad k \geq 0\ 0, \quad k < 0 \quad (12)$$

An instance of the LBP code calculation from $3 \times 3$ of the central pixel. Every pixel, initiating from the corner of the upper-left (137) is equated with the central pixel (82) and produces 1 because its value is higher than a center pixel, otherwise, if the pixel is lesser than the value of the center, $(79 < 82)$ is equivalent to 0. The outcome is an 8-bit binary code (11100001) which was changed to a decimal value (225).

### 3.3.3. SIFT Model

SIFT descriptor aids in the alteration of image data into synchronizes that are invariant to measure qualified to local features [23]. SIFT offers descriptors that are useful in object recognition and image matching. The main steps for making SIFT descriptors are given below:

1. Scale-space extrema recognition
2. Keypoint localization
3. Orientation assignment
4. Keypoint descriptor

SIFT is a significant model for image retrieval. SIFT has significant properties which are mentioned below and will be more useful for image retrieval.

1. SIFT is extremely unique and is useful in matching features of dual imageries to recover parallel images from huge databases.
2. SIFT makes a huge amount of descriptors including image over complete sort of measures and positions which are major for recovering visually parallel imageries.

### 3.3.4. GLCM Model

GLCM is a technique that contains information regarding the pixels' locations which contain similar values of gray level. The main intention of this model is to remove the texture feature. GLCM is not common in application but it has been employed in numerous areas for texture analysis that describes texture, amount of offered numerical scales of texture, $3Dco$-occurrence matrices were employed in uses of CBIR, special multi-dimensional $co$-occurrence matrices were used for object matching and recognition, while the multi-dimensional texture analysis is used in grouping models. The texture features in the statistical analysis are computed from the statistics spread of severe observation mixtures in exact places with every other in the imagery. The statistics were classified into 3 like second-order, first-order, and higher-order statistics as per the number of intensity points. Also, GLCM was measured as a matrix and its columns and rows were equivalent to the number of gray levels in the imagery. The matrix element $P(i, j|d, \theta)$ is the

*International Journal on Recent and Innovation Trends in Computing and Communication*
*ISSN: 2321-8169 Volume: 11 Issue: 1*
***Article Received:*** *25 September 2022* ***Revised:*** *12 November 2022* ***Accepted:*** *30 December 2022*

frequency of proportional with dual pixels like intensity $i$ and intensity $j$ divided by distance $d$ and stated by the exact angle $(\theta)$.

The GLCM algorithm is mentioned below:

1. Calculate the complete pair of pixels whereas the 1st pixel will be set to I value and the equivalent pair of a pixel is moved from the 1st pixel by $d$ which has the $j$ value.
2. The computation is added in the $j$ column and $i$ row of the matrix. $P_d[i,j]$.
3. It is value to be stated that $[i,j]$ is not symmetric if the number of pair pixels that have the gray levels $[i,j]$ does not parallel to the number of pair pixels having the gray levels $[i,j]$.
4. Every entry will be separated by the total amount of pairs of pixels that standardize the $[i,j]$ elements.
5. The normalization of GLCM $N[i,j]$ is definite by:

$$N[i,j] = \frac{P_d[i,j]}{\sum_i \sum_j P_d[i,j]} \quad (13)$$

GLCM has 3 parameters such as: The 1st one is a range utilized measuring an input imagery into gray level which has the value of least and highest grayscale in the picture as bound. The 2nd parameter is the number of gray levels that define the size of GLCM. The 3rd parameter states that the distance between the interest and its neighbor depends upon the value of the angle.

### 3.4. Classification using RF Model

Finally, the classification of leukemia can be made by the use RF model. The ensemble of numerous uncorrelated DT and capturing architectures builds the RF algorithm [24]. Now, a particular quantity of data will be randomly removed by employing the bootstrap sampling technique. Out-of-bag (OOB) predictors have been built employing bootstrap training sets. RF could be utilized to execute classification tasks. The categories with the maximum votes developed the model-prediction output class that can be produced through each separate tree in the RF. Since, DT will be excessively complex to their training data, which provides numerous samples, produce a massive type of DT. For including additional uncorrelation, all DTs employ individual features to classify tasks. These features randomly offer higher predictable capability for RF. Therefore, it is demonstrated to be more excellent than a single DT, as all of them will be not impacted by errors. Fig. 2 illustrates the infrastructure of the RF classifier.

Every tree in RF must be produced by arbitrarily choosing the row and column interpretations in the database. By employing a single DT, it is challenging to appropriately forecast the output. Alternatively, combined outcomes from integrating various DTs that accurately generate the overall output. In the improved architecture, a reduction in the Gini Index value. The formula for computing the Gini Index will be offered in Eq. (14). Now $i$ describe the existing sample, $P_i$ refers to the possibility of the $i^{th}$ instance and $n$ signifies the total amount of instances in the database.

$$Gini\ Index = \sum_{i=1}^{n} (P_i)(1 - (P_i)) = 1 - \sum_{i=1}^{n} (P_i)^2 \quad (14)$$

The RF's main advantage is resistance to variances and biases. In classification issues, the RF definite variable can be forecasted as the output. Additionally, the RF unit's ability to create evaluation through a random instance of predictable variables, which produces it particularly appropriate for processing datasets with tremendously higher prediction variables. With the help of inspiring system variety, this random FS finally increases classification effectiveness. The training subcategory and the feature subsets for all systems could be arbitrarily sampled to make the RF. The last prediction will be integrated into the maximum vote. It creates the RF well-executed and strong in various applications.

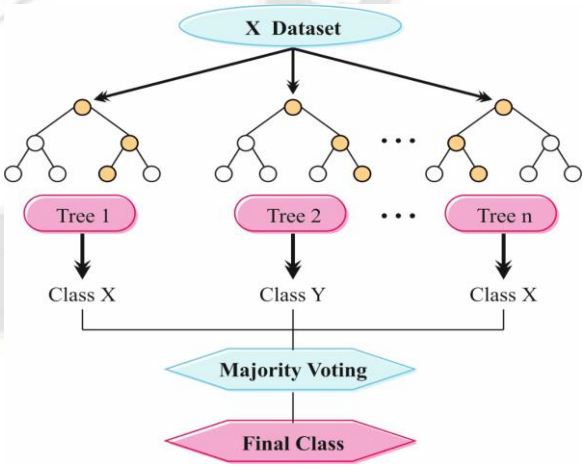


**Fig. 2.** Structure of RF classifier

### 4. Performance Validation

In this study, the performance analysis of the ALDC-MLMBI technique is tested utilizing a leukemia classification dataset from Kaggle [25]. The dataset includes 3875 samples with 1927 normal images and 1948 ALL classes as defined in

*International Journal on Recent and Innovation Trends in Computing and Communication*
*ISSN: 2321-8169 Volume: 11 Issue: 1*
***Article Received:*** *25 September 2022* ***Revised:*** *12 November 2022* ***Accepted:*** *30 December 2022*

Table 1. Fig. 3 represents the sample of ALL and normal images. Fig. 4 depicts the visualization of pre-processed images.

**Table 1** Details of dataset

| **Classes** | **No. of Samples** |
|---|---|
| NORMAL | 1927 |
| ALL | 1948 |
| **Total Samples** | **3875** |

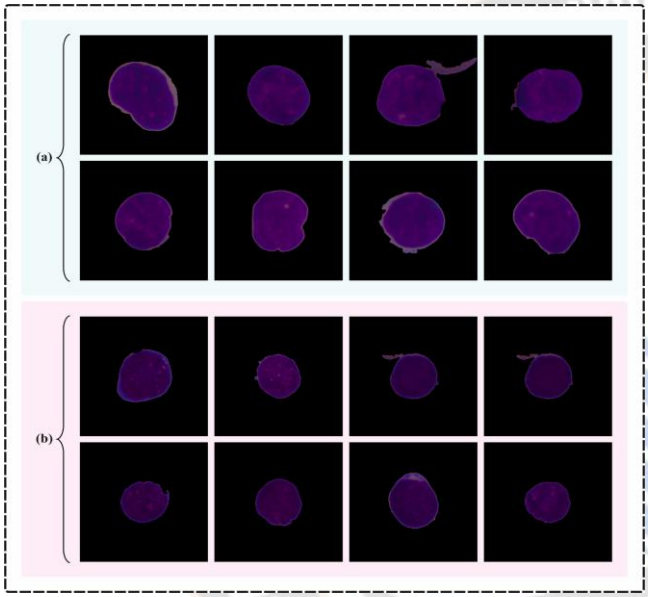


**Fig. 3.** Sample Images a) ALL b) Normal

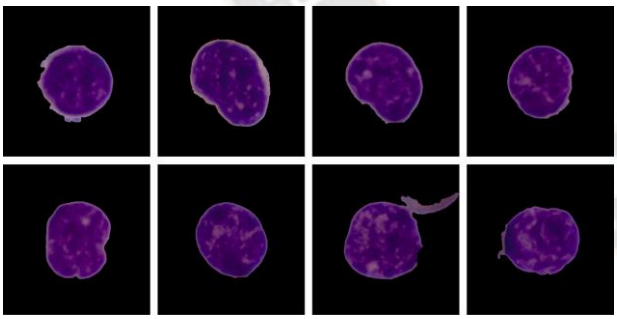

**Fig. 4.** Pre-processed Images

Fig. 5 establishes the classifier outcomes of the ALDC-MLMBI system under the training and testing data. Fig. 5a displays the confusion matrices delivered by the ALDC-MLMBI system. The figure signified that the ALDC-MLMBI system has recognized and categorized all 2 class labels exactly. Similarly, Fig. 5b determines the PR study of the ALDC-MLMBI system. The figure identified that the ALDC-MLMBI system has attained the highest performance of PR below all classes. Lastly, Fig. 5c reveals the ROC study of the ALDC-MLMBI technique. The outcome shows that the ALDC-MLMBI model has resulted in skilled results with the highest ROC values under dissimilar classes.

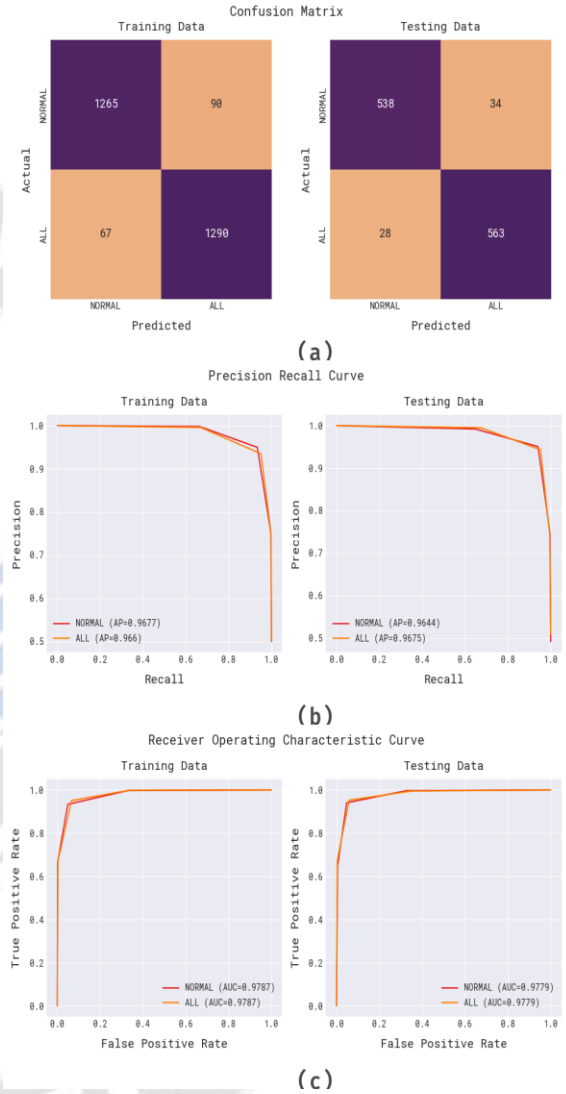


**Fig. 5.** Training and Testing Data a) Confusion Matrix b) PR Curve c) ROC Curve

Table 2 represents the complete leukemia detection results of the ALDC-MLMBI system under training and testing data. In Fig. 6, the detailed leukemia detection outcomes of the ALDC-MLMBI system are provided. The results stated that the ALDC-MLMBI model properly identified the normal and ALL classes. With normal class, the ALDC-MLMBI technique offers $an\ accu_y$ of 94.21%, $prec_n$ of 94.97%, $sens_y$ of 93.36%, $spec_y$ of 95.06%, and $F1_{score}$ of 94.16%. Also, with ALL classes, the ALDC-MLMBI system provides an $accu_y$ of 94.21%, $prec_n$ of 93.48%, $sens_y$ of 95.06%, $spec_y$ of 93.36%, and $F1_{score}$ of 94.26%.

*International Journal on Recent and Innovation Trends in Computing and Communication*
*ISSN: 2321-8169 Volume: 11 Issue: 1*
***Article Received:*** *25 September 2022* ***Revised:*** *12 November 2022* ***Accepted:*** *30 December 2022*

**Table 2** Leukemia detection of ALDC-MLMBI technique under training and testing data

| Data Split | Measures | NORMAL | ALL | Overall |
|---|---|---|---|---|
| Training | Accuracy | 94.21 | 94.21 | 94.21 |
| | Precision | 94.97 | 93.48 | 94.22 |
| | Sensitivity | 93.36 | 95.06 | 94.21 |
| | Specificity | 95.06 | 93.36 | 94.21 |
| | F1-Score | 94.16 | 94.26 | 94.21 |
| Testing | Accuracy | 94.67 | 94.67 | 94.67 |
| | Precision | 95.05 | 94.30 | 94.68 |
| | Sensitivity | 94.06 | 95.26 | 94.66 |
| | Specificity | 95.26 | 94.06 | 94.66 |
| | F1-Score | 94.55 | 94.78 | 94.67 |

In Fig. 7, the detailed leukemia detection outcomes of the ALDC-MLMBI approach are provided. The outcomes stated that the ALDC-MLMBI system correctly recognized the normal and ALL classes. With normal class, the ALDC-MLMBI method offers $an\,accu_y$ of 94.67%, $prec_n$ of 95.05%, $sens_y$ of 94.06%, $spec_y$ of 95.26%, and $F1_{score}$ of 94.55%. Also, with ALL classes, the ALDC-MLMBI system offers an $accu_y$ of 94.67%, $prec_n$ of 94.30%, $sens_y$ of 95.26%, $spec_y$ of 94.06%, and $F1_{score}$ of 94.78%.

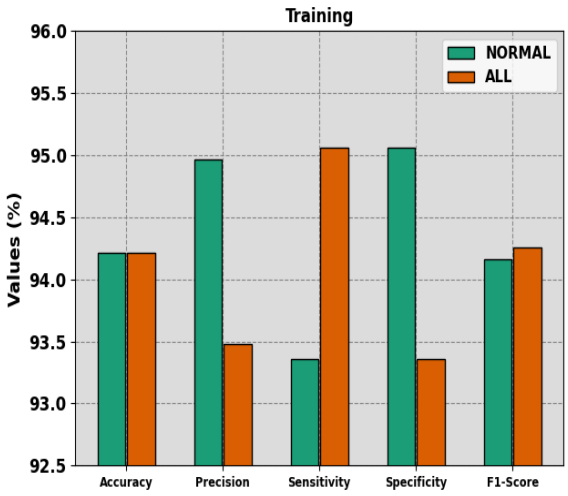


**Fig. 6.** Leukemia detection of ALDC-MLMBI technique under training data

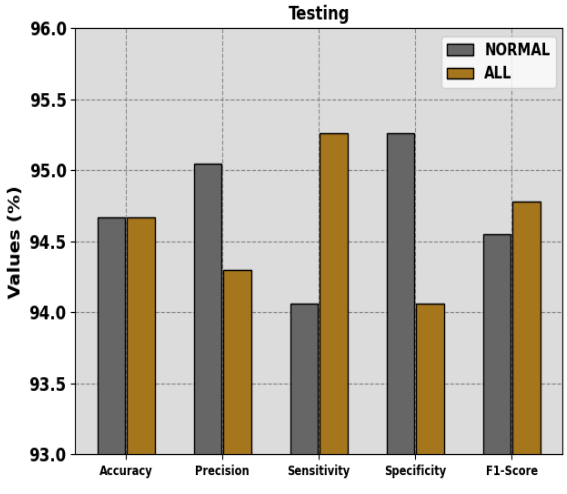


**Fig. 7.** Leukemia detection of ALDC-MLMBI technique under testing data

In Table 3, a brief comparison study of the ALDC-MLMBI approach with recent DL techniques is given [26, 27]. In Fig. 8, the comparative results of the ALDC-MLMBI system in terms of $sens_y$ and $spec_y$ are given. The outcomes highlighted that the ALDC-MLMBI technique gains better performance. Based on $sens_y$, the ALDC-MLMBI technique exhibits a higher $sens_y$ of 94.66% while the CNN, VT, CNN-ECA, GAO, and SVM-cell energy feature models portray lower $sens_y$ of 92.43%, 92.44%, 91.74%, 91.82%, and 91.32%, correspondingly. Additionally, based on $spec_y$, the ALDC-MLMBI approach exhibits a greater $spec_y$ of 94.66% whereas the CNN, VT, CNN-ECA, GAO, and SVM-cell energy feature methods represent lower $spec_y$ of 89.18%, 93.24%, 89.56%, 91.67%, and 88.99%, respectively.

In Fig. 9, the comparative outcomes of the ALDC-MLMBI method in terms of $accu_y$, $prec_n$, and $F1_{score}$ are given. The outcomes highlighted that the ALDC-MLMBI method gains superior performance. Based on $accu_y$, the ALDC-MLMBI system parade greater $accu_y$ of 94.67% whereas the CNN, VT, CNN-ECA, GAO, and SVM-cell energy feature techniques represent lesser $accu_y$ of 88.25%, 88.20%, 91.10%, 93.84%, and 94.00%, correspondingly. Moreover, based on $prec_n$, the ALDC-MLMBI methodology reveals a greater $prec_n$ of 94.68% while the CNN, VT, CNN-ECA, GAO, and SVM-cell energy feature models describe lower $prec_n$ of 88.72%, 91.70%, 90.85%, 89.01%, and 92.20%, correspondingly. Finally, based on $F1_{score}$, the ALDC-MLMBI model exhibits a higher $F1_{score}$ of 94.67% while the CNN, VT, CNN-ECA, GAO, and SVM-cell energy feature methods represent a lower $F1_{score}$ of 91.77%, 87.76%, 92.49%, 88.22%, and 93.88%, correspondingly.

*International Journal on Recent and Innovation Trends in Computing and Communication*
*ISSN: 2321-8169 Volume: 11 Issue: 1*
***Article Received:*** *25 September 2022* ***Revised:*** *12 November 2022* ***Accepted:*** *30 December 2022*

**Table 3** Comparative analysis of ALDC-MLMBI technique with recent DL models

| **Methods** | $Accu_y$ | $Prec_n$ | $Sens_y$ | $Spec_y$ | $F1_{Score}$ |
|---|---|---|---|---|---|
| CNN Model | 88.25 | 88.72 | 92.43 | 89.18 | 91.77 |
| Vision Transformer | 88.20 | 91.70 | 92.44 | 93.24 | 87.76 |
| CNN-ECA Module | 91.10 | 90.85 | 91.74 | 89.56 | 92.49 |
| GAO-Based Method | 93.84 | 89.01 | 91.82 | 91.67 | 88.22 |
| SVM-Cell Energy Feature | 94.00 | 92.20 | 91.32 | 88.99 | 93.88 |
| ALDC-MLMBI | 94.67 | 94.68 | 94.66 | 94.66 | 94.67 |

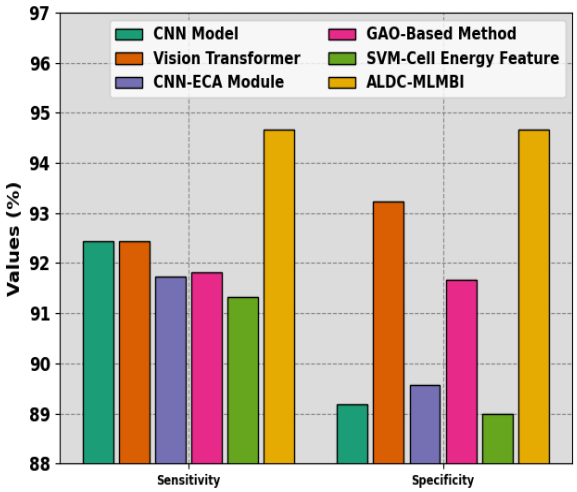


**Fig. 8.** $Sens_y$ and $spec_y$ analysis of ALDC-MLMBI technique with recent DL models

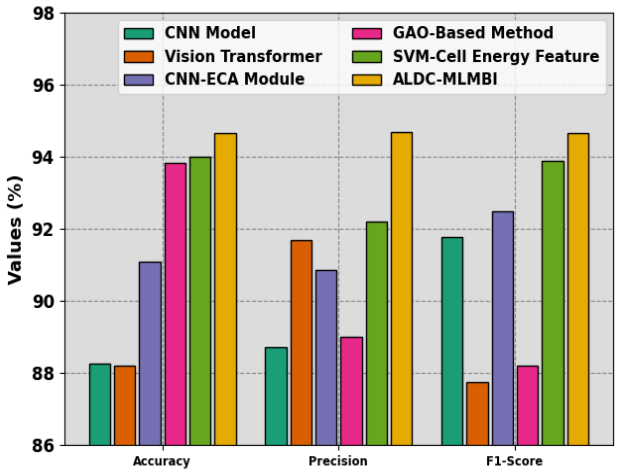


**Fig. 9.** Comparative analysis of ALDC-MLMBI technique with recent DL models

Thus, the ALDC-MLMBI method can be applied for an effectual leukemia detection process.

## 5. Conclusion

In this paper, we have projected a novel ALDC-MLMBI method. The ALDC-MLMBI technique aims to employ ML approaches for the identification of leukemia. To accomplish that, the ALDC-MLMBI method has image preprocessing, segmentation, feature extraction, and classification processes. As a primary step, the ALDC-MLMBI technique follows MF-based noise elimination and AHE-based contrast enhancement. Besides, the segmentation procedure can be performed by watershed segmentation. Meanwhile, the ALDC-MLMBI system involves a set of feature extractors namely HOG, LBP, SIFT, and GLCM. Furthermore, the classification of leukemia can be made by the use RF model. The simulation analysis of the ALDC-MLMBI technique can be performed using the Kaggle image dataset. The experimentation results highlighted the superior performance of the ALDC-MLMBI system compared to existing classifiers.

## References

[1] Bodzas, A., Kodytek, P. and Zidek, J., 2020. Automated detection of acute lymphoblastic leukemia from microscopic images based on human visual perception. *Frontiers in Bioengineering and Biotechnology*, *8*, p.1005.

[2] Umamaheswari, D. and Geetha, S., 2018, May. Review on image segmentation techniques incorporated with machine learning in the scrutinization of leukemic microscopic stained blood smear images. In *International Conference on ISMAC in Computational Vision and Bio-Engineering* (pp. 1773-1791). Springer, Cham.

[3] Marzahl, C., Aubreville, M., Voigt, J. and Maier, A., 2019. Classification of leukemic b-lymphoblast cells from blood smear microscopic images with an attention-based deep learning method and advanced augmentation techniques. In *ISBI 2019 C-NMC challenge: classification in cancer cell imaging* (pp. 13-22). Springer, Singapore.

[4] Anilkumar, K.K., Manoj, V.J. and Sagi, T.M., 2020. A survey on image segmentation of blood and bone marrow smear images with emphasis on automated detection of Leukemia. *Biocybernetics and Biomedical Engineering*, *40*(4), pp.1406-1420.

[5] Jha, K.K. and Dutta, H.S., 2019. Mutual information based hybrid model and deep learning for acute lymphocytic leukemia detection in single cell blood

smear images. *Computer methods and programs in biomedicine*, *179*, p.104987.

[6] Atteia, G., Alhussan, A.A. and Samee, N.A., 2022. BO-ALLCNN: Bayesian-Based Optimized CNN for Acute Lymphoblastic Leukemia Detection in Microscopic Blood Smear Images. *Sensors*, *22*(15), p.5520.

[7] Tusar, M., Khan, T.H. and Anik, R.K., 2022. Automated Detection of Acute Lymphoblastic Leukemia Subtypes from Microscopic Blood Smear Images using Deep Neural Networks. *arXiv preprint arXiv:2208.08992*.

[8] Anilkumar, K.K., Manoj, V.J. and Sagi, T.M., 2021. Automated detection of leukemia by pretrained deep neural networks and transfer learning: a comparison. *Medical Engineering & Physics*, *98*, pp.8-19.

[9] Dese, K., Raj, H., Ayana, G., Yemane, T., Adissu, W., Krishnamoorthy, J. and Kwa, T., 2021. Accurate Machine-Learning-Based classification of Leukemia from Blood Smear Images. *Clinical Lymphoma Myeloma and Leukemia*, *21*(11), pp.e903-e914.

[10] Rastogi, P., Khanna, K. and Singh, V., 2022. LeuFeatx: Deep learning–based feature extractor for the diagnosis of acute leukemia from microscopic images of peripheral blood smear. *Computers in Biology and Medicine*, *142*, p.105236.

[11] Talaat, F.M. and Gamel, S.A., 2024. Machine learning in detection and classification of leukemia using C-NMC_Leukemia. *Multimedia Tools and Applications*, *83*(3), pp.8063-8076.

[12] More, P. and Sugandhi, R., 2023. Automated and enhanced leucocyte detection and classification for leukemia detection using a multi-class SVM classifier. *Engineering Proceedings*, *37*(1), p.36.

[13] Karthick, T., Ramprasath, M. and Sangeetha, M., 2022, April. Classification of Blast Cells in Leukemia Using Digital Image Processing and Machine Learning. In *Proceedings of International Conference on Deep Learning, Computing and Intelligence: ICDCI 2021* (pp. 1-18). Singapore: Springer Nature Singapore.

[14] Muzalevskiy, D. and Torshin, I., 2024. Image Classification of Leukemic Cells Using Invariants of Triangle-Free Graphs as Synthetic Features. *IEEE Access*.

[15] Karthikeyan, S., Moses, M.L., Ramya, P., Thrisha, E., Kalarani, K. and Susheel, R.N., 2023, April. Machine Learning based Algorithmic approach for Detection and Classification of Leukemia. In *2023 2nd International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN)* (pp. 1-6). IEEE.

[16] Rangini, M., Pundir, S., Soudagar, M.E.M., Vekariya, D., Patil, H. and Rajkumar, S., 2024, January. Texture-based Feature Extraction and Machine Learning Model for the Detection of Acute Lymphoblastic Leukemia. In *2024 5th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI)* (pp. 505-511). IEEE.

[17] Rupapara, V., Rustam, F., Aljedaani, W., Shahzad, H.F., Lee, E. and Ashraf, I., 2022. Blood cancer prediction using leukemia microarray gene data and hybrid logistic vector trees model. *Scientific reports*, *12*(1), p.1000.

[18] Gupta, G., 2011. Algorithm for image processing using improved median filter and comparison of mean, median and improved median filter. *International Journal of Soft Computing and Engineering (IJSCE)*, *1*(5), pp.304-311.

[19] Sarkar, M. and Mandal, A., 2023. SLAAHE: Selective Apex Adaptive Histogram Equalization. *Franklin Open*, *3*, p.100023.

[20] Levner, I. and Zhang, H., 2007. Classification-driven watershed segmentation. *IEEE Transactions on Image Processing*, *16*(5), pp.1437-1445.

[21] Neggaz, I. and Fizazi, H., 2022. An Intelligent handcrafted feature selection using Archimedes optimization algorithm for facial analysis. *Soft Computing*, *26*(19), pp.10435-10464.

[22] Albera, S., 2016. Vehicle logo recognition using image processing methods. *Atilim University*.

[23] Srivastava, P., Khare, M. and Khare, A., 2017, June. Content-based image retrieval using scale invariant feature transform and gray level co-occurrence matrix. In *Second International Workshop on Pattern Recognition* (Vol. 10443, pp. 151-156). SPIE.

[24] Raj, A., Shetty, S.D. and Rahul, C.S., 2024. An efficient indoor localization for smartphone users: Hybrid metaheuristic optimization methodology. *Alexandria Engineering Journal*, *87*, pp.63-76.

[25] https://www.kaggle.com/datasets/andrewmvd/leukemia-classification

[26] Bukhari, M., Yasmin, S., Sammad, S., El-Latif, A. and Ahmed, A., 2022. A deep learning framework for leukemia cancer detection in microscopic blood samples using squeeze and excitation learning. *Mathematical Problems in Engineering*, *2022*.

[27] Batool, A. and Byun, Y.C., 2023. Lightweight EfficientNetB3 model based on depthwise separable convolutions for enhancing classification of leukemia white blood cell images. *IEEE Access*.