Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 6 (2023)

Enhanced Leukemia Detection Based on Sequential Feature Selection with Optimized Random Nearest Neighbor Model for Microscopic Images

I.Vinurajan[1], Dr. K. P. Sanal Kumar[2], Dr. S. Anu H Nair[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

The prevalence of leukemia as a life-threatening disease emphasizes the importance of early detection for effective treatment. The medical image processing performs an important role in the diagnosis of leukemia. Blood cancer that results from abnormal or not fully formed White Blood Cells (WBCs). ALL in the WBC are the body's warriors against infected cells. The abnormal proliferation of WBC in the bone marrow can damage other cells and affect bone marrow and lymphoid tissue. The manual morphological analysis of blood cells by hematologists is time-consuming, results depend on ineffective for large datasets, expertise, and can be biased. Blood cancers are also varied and complex in shape, texture, color and severity. In addition, different staining and light changes make the identification of blood cancers more difficult. To resolve the issue, a Random Nearest Neighbor (RNN) methodology based on Machine Learning (ML) protocol was deployed. To classify the blood cancer images, we deployed such techniques. In first, we collected a blood cancer dataset at Kaggle Website. To perform image processing the RNN contains four stages they are preprocessing, segmentation, feature selection, and classification. The initial phase is preprocessing, each pixel is replaced by a weighted average of its neighbors by the use of Bilateral Filter method. This improves the image quality and makes extracting meaningful features easier. Bilateral filters protect edges while smoothing the rest of the image. This helps to preserve the important features of the leukemia images. The second stage is segmentation, in this stage the images identifies and groups pixels or regions together based on similarity criteria based on Region Growing segmentation method. It starting from a seed point, adjacent pixels are added back into the region based on similar criteria. This criterion may be based on intensity values or other features relevant to the leukemia data set. After the segmentation we select an image feature through Sequential Feature Selection (SFS) method. This method can use it to remove unnecessary features from large datasets, or to progressively add and select useful features. Finally, we classify the images through RNN method to analyze dataset images of blood examples to categorize irregular cells symptomatic of blood cancer. The RNN method uses the concept of nearest neighbors to classify leukemia images, but also uses a random selection mechanism to improve classification accuracy of 94.5%. This method addresses this limitation by introducing randomness into the neighborhood selection. This helps to ensure reliable classification even in presence of noise. The investigational consequences demonstrate that our deployed methodology beats the existing model in accuracy, sensitivity, F1 score, specificity and error rate.

***Keywords*:** leukemia, WBC, RNN, Bilateral method, SFS, region growing segmentation.

1. Introduction

Within the bloodstream, there are three main types of cells present: WBC, Red Blood Cells (RBCs), and platelets. These cells play crucial roles in maintaining the overall health and functioning of the body. To each of them is endlessly supplied in the bone marrow and discharged into the blood stream at the appropriate time. Interference with the growth of normal blood cells by the rapid development of irregular blood cells is the foremost reason of leukemia. In the blood cancer there has three major categories: ALL, lymphoma and myeloma. The bone marrow is the site of infection with a category of WBC called Acute Lymphoblastic Leukemia (ALL). "Acute" indicates rapid disease progression, potentially life-threatening without prompt treatment, categorized into L1, L2, and L3. The multiple myeloma is a teratoma of undeveloped cells known as plasma, which help clear infections. Multiple myeloma causes a low number of platelets in the blood called thrombocytopenia. This can cause bone erosion called osteolysis, which can be detected by CT scans [D. Kumar et al, (2020)]. Several characteristics of blood smears pose obstacles during morphological analysis, such as blurring, noise, blood staining changes, cell overlap, cell occlusion, and cell size changes. Blood films are investigated by hematologists, so manual examination can clearly address the issue. But, manual morphological analysis of blood cells by hematologists is time-consuming, results depend on ineffective for large datasets, expertise, and can be biased [T. Mustaqim et al (2023)].

Other challenges are related to the complexity of blood cancers, such as irregular borders and histological similarities between WBC and further blood modules, making it difficult to separate WBC as of further blood modules. Blood cancers are also varied and complex in shape, texture, color and severity. Leukemia is heterogeneous and consists of unlike categories, as well as normal cells and abnormal cells. In addition, different staining and light changes make the identification of blood cancers more difficult [T. A. M. Elhassan et al, (2022)]. To resolve the issue, a RNN method based on ML methodology was proposed. RNN method is a novel and effective method for accurate classification of leukemia images. The RNN method uses the concept of nearest neighbors to classify leukemia images, but also uses a random selection mechanism to improve classification accuracy. Traditional nearest neighbor methods rely on finding nearest neighbors based on a certain distance measure. However, these methods are sensitive to noise and outliers in the data, which can lead to inaccurate classification results. RNN methods address this limitation by introducing randomness into the neighborhood selection. This helps to ensure reliable classification even in presence of noise.

Rather than determining nearest neighbors, RNN methods select random subsets of neighborhoods based on a probability model. This random selection helps to capture the heterogeneity of the dataset and minimize the impact of outliers. Additionally, the RNN method uses a weighting mechanism that assigns different weights to selected neighborhoods based on their relevance to the query image, further improving classification accuracy. Overall, the RNN method provides a promising solution for accurately classifying leukemia images and provides a robust and efficient method for the early analysis and medication of leukemia.

1.1 The main contribution of the RNN methodology is mention below:

- The way compared with the traditional proximity method, the RNN method significantly improved the classification accuracy of leukemia images. The RNN method provides a powerful and effective solution for accurate leukemia image classification by combining stochasticity, heterogeneity, and correlation weighting.
- The random selection mechanism of the RNN method helps capture the diversity of the data set. By choosing a random subset of the neighborhood, the RNN method ensures that the classification is based on a representative sample of the dataset, resulting in more accurate results.
- Despite the random selection mechanism, RNN method is efficient, scalable and suitable for large-scale leukemia image classification tasks. This approach can handle large data sets without significant computational overhead, which makes it suitable for real-world applications.

2. Literature Survey

Maksym A. Jopek et al., (2024), carried out a Deep Learning (DL) based Multiclass Approach (MA) to examine and classify illnesses on the basis of blood platelet RNA. This represents a promising opportunity for cancer diagnosis to aid clinicians' decision-making processes in solving the black box issue. However, this method is expensive, difficult to obtain or lacks precision. Therefore, there is a need for a non-invasive, direct method to allow a wide range of blood patterns to detect and characterize cancer.

Mohammad Akter Hossain et al., (2023), carried out supervised Machine Learning (ML) methodology to precisely forecasts the chances of initial stage leukemia based on symptoms. The deployed methodology is on the basis of Decision Tree (DT) classifier that make available considerably well consequences associated to further methodologies and causes descriptive instructions that are organized to consume. The main drawback is that poor prediction accuracy, lower precision, and F1 Score.

Astha Ratley et al., (2020), carried out an Artificial Neural Network (ANN) methodology for if the information in a particular node exceeds the limit, it is sent as information to the next node with the help of connection strings Annotated collections connected to other roads. The main drawback is that by reason of the restricted illustration capability of the deployed method, conventional methodologies every so often endure from a large amount of false positive rate in order to accomplish an adequate sensitivity.

Vasundhara Acharya et al., (2023), carried out an automated computer-aided diagnosis function. It can solve key problems in leukemia detection such as normal or abnormal identification and cytoplasmic extraction without manual intervention. This approach makes it more difficult to identify the case at the time of initial identification, as the symptoms can be difficult to identify and vague.

Amreen Batool et al., (2023), discussed that appropriate and exact cancer diagnosis is significant for effectual treatment that improves survival rates. To resolve the objective, a EfficientNet-B3 methodology was deployed. The model depth and complexity can lead to increased inference time and memory usage.

Shruti et al., (2024), developed a hexagonal tiled metamaterial absorber combined with ML techniques to detect blood cancer and breast cancer. Also, created a large dataset by modifying the thickness and refractive index of the sample to detect cancer cells using machine learning. Then found that using additional tree classifiers improves the detection accuracy and efficiently uses the absorbers to detect blood and breast cancers.

Kokeb Dese et al., (2021), discussed that misdiagnosis poses a major challenge in detecting blood cancers at an early stage. To resolve the issue, a Support Vector Machine (SVM) was deployed to sort four communal categories of leukemia's over a vigorous image segmentation methodology. The main disadvantage of this method is that the amount of features in one data point surpasses the amount of training data models, thus decreasing the performance.

Gemma Urbanos et al., (2021), discussed that ML methodologies for instance SVM, Random Forests (RF) and Convolutional Neural Networks (CNN) are castoff to generate estimates and make available in vivo visualization. This will enable neurosurgeons to be more precise, thus reducing damage to healthy tissue. However, to train declared ML methodologies the author selected 13 in vivo hyperspectral images from 12 different high-grade glioma patients. The foremost difficulty of the proposed methodology is that although training is fast, it takes a very long time to create predictions after training.

Emine Cengil et al., (2021), carried out a ResNet18 with SVM methodology to classify four disimilar types of WBC. Finally, the results are displayed via the identical classifier after PCA protocol. The consequences showed that the deployed method did not subsidize much to the multiple classification of WBC.

K. SRIDHAR et al., (2022), discussed that leukemia is a malignant cancer that can cause many medical complications. Professional hematologists and pathologists manually examine blood samples under a microscope for diagnosis. Technologies like image processing and pattern recognition can help these professionals. To attain the objective, an enhanced ML methodologies Lightweight Ensemble (LE) classification was deployed. Achieve

greater accuracy in image classification tasks. However, one of the disadvantages of using the proposed method is that the gradient may disappear in very deep networks, which may slow down the gradient descent process.

V. Durga Prasad Jasti et al., (2022) discussed that the cause of breast cancer is unknown, there are currently no efficient techniques for precluding or pick up the check breast cancer. Early diagnosis is the best way to detect and treat breast cancer, and early detection increases the chance of a full recovery. To attain the objective, deployed a ML methodologies for instance SVM, KNN, RF, and Naïve Bayes (NB). But the accuracy and recall was too low for classify the dataset because the time complexity is very high.

Saba Saleem et al., (2021) discussed that in order to overcome the harshness of this illness, early diagnosis of immature cell morphology is required, which ultimately leads to a reduction in the mortality rate of patients. In recent years, various segmentation and classification methods based on DL models have been proposed, but they still have some limitations. To resolve the issue, a deep network methodology was deployed. The proposed method provides more opportunities, but it is also more complicated and the development takes much longer.

Changhun Jung et al., (2022), discussed that classification of the cells observed remains a challenge, as the supply of the five cell types influences the status of the immune system. To resolve ether issue, a W-Net based CNN methodology was deployed. The deployed methodology has some drawbacks for instance the transfer learning property and the class imbalance.

Mustafa Ghaderzadeh et al., (2021), discussed that inspection of these PBS images by laboratory operators is fraught with difficulties, including diagnostic errors, as not all signs and symptoms are specified, often escorting to misdiagnosis. To resolve the issue, a fast and effective CNN protocol was deployed. This approach is difficult for doctors in the early stages, as it acquires a lot of time to classify the images in a blood smear dataset.

T. Karthick et al., (2022), carried out fast and cost-effective computer aided mechanism to classify blast cells in blood cancer via digital image processing and ML. WBC from blood smear micrographs were first extracted by the Otsu methodology, and a cell separation mechanism was used to break the overlapping cells. But this protocol is still desired to progress the diagnostic accuracy.

Sunita Chand et al., (2022), carried out a novel Deep Learning Framework (DLF) methodology to classification difficulties, further precisely, image-based classification. No feature extraction is required this time. It does not necessitate any prior training on other databases and therefore can be castoff for real-time applications in leukemia detection. Although this method significantly reduces leukemia mortality, it suffers from high false-positive rates leading to unnecessary diagnostic procedures.

Ebru Simsek et al., (2021), discussed that Machine learning is an effective way to identify leukemia subtypes where time and accurate diagnosis are of the essence. In this K-nearest neighbor, Linear Discriminant, SVM, and Ensemble classifiers ML protocols was deployed. In the node candidate detection phase, the system tries to cover all possible candidate nodes. However, it is also providing a large number of false positives, which may difficulties for diagnosis.

V. Lakshmi Thanmayi A et al., (2021), discussed that early detection and diagnosis is extremely important. Additionally, computer-aided diagnosis improves the testing process and makes the test more accurate. Therefore, SVM was used for classify among cancerous and non-cancerous cells. Although this approach is less accurate, it enables accurate and effective classification and treatment of blood cancers.

Mohammad Akter Hossain et al., (2020), carried out a supervised ML methodology that accurately predicts the likelihood of leukemia in its early stages. In the proposed method, it was dividing the data into training and test datasets and use different ML techniques like DT, RF, KNN, Linear Regression (LR), Adaboost and NB to check the accuracy. It is detected that the DT and the RF classifier beat the other methodologies.

Muhammad Zakir Ullah et al., (2021), carried out a CNN-based non-invasive methodology has been implemented to perform diagnostic tasks using medical images. The deployed mechanism contains of a CNN-based methodology that usages an attention block named Efficient Channel Attention (ECA) and VGG16 to extract high value deep aspects as of image datasets, resulting in enhanced feature illustration and improved classification consequences. This methodology has the low accuracy when the network training focus on different samples.

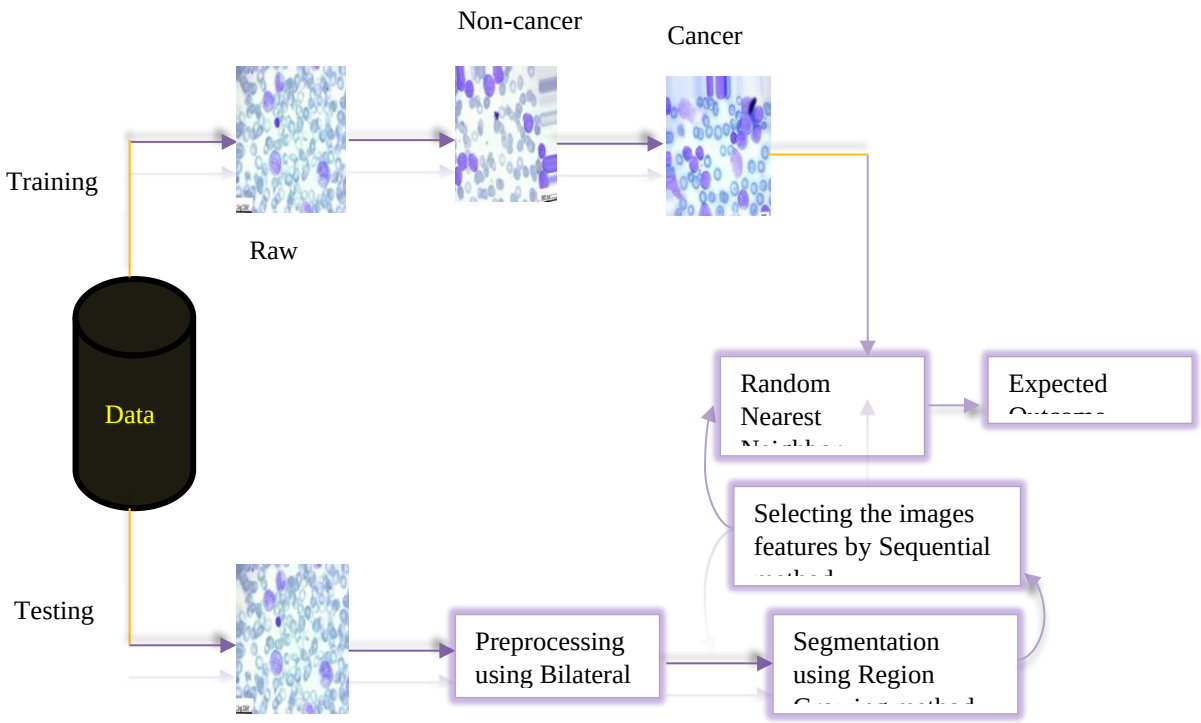
S. Rajeswari et al., (2022), the author aimed to develop a method for detecting leukemia using ML methodologies for instance image processing and transfer learning, and to determine whether the leukemia is AML, ALL, CML or CLL. To attain the objective, a Xception convolutional methodology was deployed. In addition to this, ML-based models are trained over a small number of samples, which creates generalization problems.

Umair Saeed et al., (2024), discussed that the traditional diagnosis of this disease is made by medical pathologists by analyzing microscopic images of WBC. However, the process relies on manual monitoring and occasionally yields inaccurate results. To resolve the issue, a DeepLeukNet methodology was deployed. Lack of interpretability is a serious concern in the proposed approach where understanding the reasoning behind the diagnosis is crucial.

3. Implementation of the Proposed Method

This section describes the integrated approach of the deployed methodology. The figure 1 shows the details of the deployed approach. In this novel, we accumulate a blood cancer dataset from Kaggle website.

Figure 1. The proposed diagram for RNN method

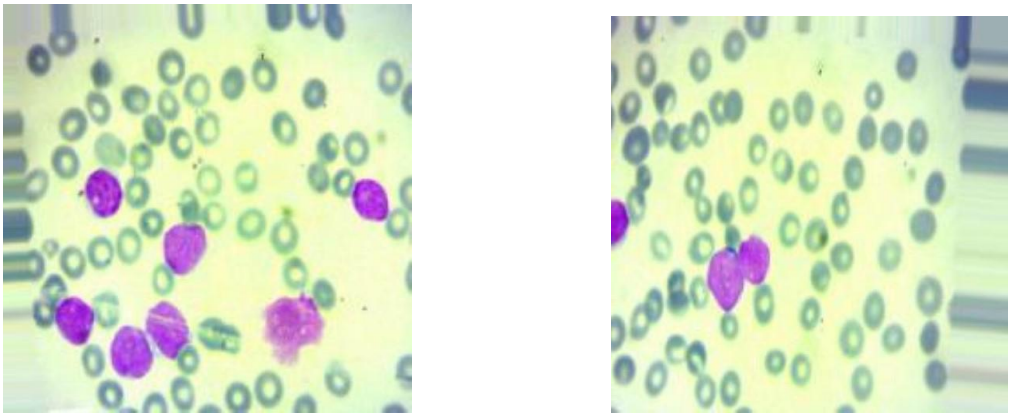


As depicted in figure 1, the dataset will be gathered and preprocessed. Subsequently, each pixel will be analyzed with a weighted average of its neighboring pixels using the Bilateral Filter technique. This improves the image quality and makes extracting meaningful features easier. After pre-processing, the images are segment and groups pixels or regions together based on similarity criteria based on Region Growing segmentation method. It starting from a seed point, adjacent pixels are added back into the region based on similar criteria. Then we select

the features of dataset images by the use of SFS method. In this way, it can be used to remove redundant features from large data sets or to select useful features by gradually adding them. Finally, the proposed method RNN is used to classify the images between ALL and lymphoma based on non-cancerous and cancerous.

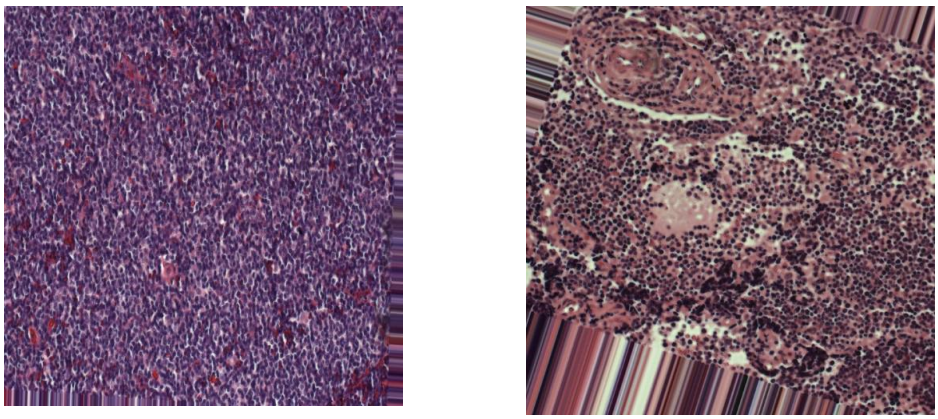
3.1 Dataset Collection

The available dataset is collected on Kaggle website and the dataset name is blood cancer dataset. It is a publicly available dataset of micro blood samples. The database highlights ALL, lymphoma, and theoretically dangerous blood cancer types. In our dataset, experienced oncologists named ALL and lymphoma. This dataset consists of 38,528 single-cell images (64x64 pixels) collected from peripheral blood smears from patients with identified blood cancers. All micro blood example images are in .jpg format with a 24-bit color depth. Again, this dataset contains two subfolders within the ALL and Lymphoma folders (e.g. test and train). To be more precise, ALL consists of 11679 images to train the method and 6218 images to test the method. Similarly, it consists of 15000 images used to train the lymphoma method and then 5631 images used to test the method.



(a) test **(b) train**

Figure 2. Acute lymphoblastic leukemia test and train model image dataset



(a) test **(b) train**

Figure 3. Lymphoma test and train model image dataset

As shown in Figures 2 and 3, using Kaggle's blood cancer dataset, this dataset can be trained and tested to classify blood cancers and accurately distinguish healthy cells from cancer cells.

3.2 Bilateral Filter

In this section, each pixel can perform edge detection and smoothing of leukemia image regions by evaluating the weighted average of its neighboring pixels. A process that combines threshold filtering can predict leukemia by analyzing the average of neighboring pixel values using a bilateral filtering method. During image processing the pixel values of small neighborhoods can be normalized by merging the interdependent leukemia disease image into a smooth region. A bilateral filter method can be utilized to analyze the average of the relative variances in pixel values induced by noise in the Leukemia disease image obtained from the dataset. A normalization constant can be analyzed using a threshold-based bilateral filter to predict the result of a low-pass domain filter on leukemia images. The dark pixel value can be assessed by averaging the central and neighboring bright pixel values using a bidirectional filter in leukemia image processing.

Equation 1 demonstrates that the output can be approximated by applying a low-pass domain filter to predict disease in the Leukemia image. Moreover, the normalization constant can be determined by considering the image function's range to calculate the vector difference during the threshold filtering procedure. Let's assume ξ −near point, L-neighbourhood centre, h- input and output multiband, $u(\xi, L)$-measure the geometric, $F(L)$ −produced filter images, F-range of image function, u-closeness function, w_v −constant, $w\big(F(\xi), F(L)\big)$ −measure the photometric, w-unbiased function.

$$\frac{h(w)=w_v^{-1}(L)\int_{-\infty}^{\infty}\square\int_{-\infty}^{\infty}\square e(\xi)u(\xi,L)v_\xi}{w_v(L)=\int_{-\infty}^{\infty}\square\int_{-\infty}^{\infty}\square u(\xi,L)v_\xi\ h(L)w_z^{-1}(L)\int_{-\infty}^{\infty}\square\int_{-\infty}^{\infty}\square F(\xi)L\big(F(\xi),F(L)\big)v_\xi\ W_w(L)=\int_{-\infty}^{\infty}\square w\big(F(\xi),F(F)\big)v_\xi} \quad (1)$$

Estimate the Euclidean distance parameter of the Gaussian function by analyzing geometric and image metric normalizations as shown in equations 2 and 3. Where σ_v −geometric spread, $w(L)$ −normalization weight, o-distance

$$\frac{h(L)=w^{-1}(u)\int_{-\infty}^{\infty}\square\int_{-\infty}^{\infty}\square F(\xi)u(\xi,L)s\big(F(\xi),F(L)\big)v_\xi}{w(L)=\int_{-\infty}^{\infty}\square\int_{-\infty}^{\infty}\square u(\xi,L)s\big(F(\xi),F(u)\big)o_\xi} \quad (2)$$

$$\frac{u(\xi,L)=e^{\frac{1}{2}\left(\frac{v(\in,L)}{\sigma_v}\right)^2}}{v(\xi,L)=v(\xi-L)=\|\xi-L\|} \quad (3)$$

The similarity function evaluation between Euclidean distances in Leukemia image processing is denoted by equations 4 and 5. Let's assume $s(\xi, L)$ −gaussian function, L and ξ -Euclidean distance, n −function.

$$s(\xi, L) = e^{\frac{1}{2}\left(\frac{\delta(F(\xi).F(L))}{\sigma^W}\right)^2} \quad (4)$$

$$\delta(\varphi, n) = \|\varphi - n\| \quad (5)$$

As described in Equations 6 and 7, the leukemia image regions can be smoothed by detecting edges using a bilateral filter through gray-level pixel value. Let's assume φ, F −intensity value, $y(\varphi)$ −frequency distribution grey level, φ −grey value, $w(\varphi, F)$-mapping kernel, $y(F)$ −histogram, F-normalized unit area, w −density function.

$$h = \int_0^{\infty}\square\varphi w(\varphi, F)v_\varphi \quad (6)$$

$$w(\varphi, F) = \frac{s(\varphi,F)y(\varphi)}{\int_0^{\infty}\square h(\varphi,F)y(\varphi)}v_\varphi \quad (7)$$

Leukemia disease images from the dataset can be a valuable resource for evaluating the characteristics of Leukemia disease. By applying a bilateral filtering technique to leukemia images, the magnitude of the pixel values can be effectively analyzed and predicted.

3.3 Region Growing Technique (RGT)

The region-growing segmentation method can be employed to detect pixels or regions by utilizing similarity criteria. Adjacent pixels or pixel collections with similar features in a leukemia image segment are

analyzed using region-growing techniques to form larger regions. The region-growing technique can estimate the segmentation features of pixels within each group. Additionally, the similarity between regional growth image segments and leukemia image pixel values for each neighborhood in a group can be represented on metric scales. In the process of mapping the gray-scale pixel values in the leukemia image, the segmentation features between the gray-scale of the pixel value can be analyzed. The RGT technique allows for the computation of Euclidean distance in color space within leukemia images, facilitating the identification and analysis of specific regions based on their leukemia image color feature.

Equation 8 demonstrates the calculation of the difference between the gray level of a pixel in a specified region and the image pixel, which is then added to the threshold pixel value. Let's assume z-grey level, z_P −grey level region, and w-threshold.

$$W = |Z - Z_P| \leq w \tag{8}$$

Calculate the Euclidean distance required to connect the corresponding pixel in the HSI color values image segment, as outlined in Equation 9. Furthermore, the homogeneity function can be used to combine the pixel area of leukemia color images. Let's assume h-hue, s-saturation, i-intensity values.

$$\sum \square |h - h_P|^2 + |s - s_P|^2 + |i - i_P|^2 \tag{9}$$

To normalize the range of RGB color values, estimate the HSI color space region of the pixels utilizing equations 10. The grey value pixel measurements of two or more leukemia regions are computed using equations 11 and 12. Let's assume RGB-red, green, blue color value, u-pixel, $F(u)$ −second order neighbourhood, i_a −labled region, e-mean, $\delta(u, i_a)$ −different measure of pixel region, $\delta(x)$ −minimum region.

$$\frac{i=\frac{1}{3}(R+G+B) \qquad i\in[0,1]}{s=1-\frac{3}{(R+G+B)}[min\ (R+G+B)] \qquad s\in[0,1]\ h=Cos^{-1}\left\{\frac{1/2[R-G]+(R-B)}{[(R-G)^2+(R-B)(G-B)]^{\frac{1}{2}}}\right\}} \tag{10}$$

$$\delta(u, I_a) = \left|Z(u) - e_{v\in I_a(u)}[Z(v)]\right| \tag{11}$$

$$a(u) = \{a|F(u) \cap i_a \neq \emptyset \wedge \delta(u)\} \tag{12}$$

Equation 13 calculates the highest similarity between neighboring pixels by normalizing the overall standard deviation. Let's assume u-pixel value, L, s and a-mean values, σ_F −nornmalized standard deviation, $\sigma_{M_{ax}}$ −maximum of standard deviation.

$$\frac{\sigma_F=\frac{\sigma}{\sigma_{M_{ax}}}}{\sigma_u=\sqrt{\frac{1}{9}\sum_{a=1}^{9}\square\left(u_a-\underline{u}\right)^2}} \tag{13}$$

Calculate the maximum relative Euclidean distance range between neighboring pixels within the growing region, as depicted in Equation 14. Let's assume O_a −euclidean distance, P-region.

$$O_a = \frac{\sqrt{(v-v_a)^2+\left(L_j-L_{ja}\right)^2+\left(L_q-L_{qa}\right)^2}}{\sqrt{v^2+L_j^2+L_q^2}} \tag{14}$$

A region-based image segmentation analysis can be conducted to measure the gray values of two or more leukemia regions and standardize the overall deviation. Furthermore, the developmental areas of leukemia images can be examined by determining the maximum Euclidean distance between neighboring pixels.

3.4 Sequential Feature Selection Algorithm

The optimal feature can be chosen using the SFS method to eliminate unnecessary features from the blood cancer dataset. Sequential feature selection algorithms containing row and column features can be utilized

to analyze leukemia imaging data matrices. By utilizing a column vector containing leukemia data values and class labels, the SFS method enables the selection and determination of image features. Logistic vectors can also play a role in the SFS method for selecting feature subsets to predict leukemia image data. Furthermore, correlation values can be used to assess linear relationships and directionality, measuring the intensity between leukemia image features and class labels. To achieve high classification accuracy at each node of the search tree, maintaining a heap that prioritizes sorting unvisited features based on cost-sensitive image feature selection accuracy is essential. Through iterative exploration of analysis branches and selection of the largest feature variable, the SFS algorithm can effectively predict optimal leukemia imaging.

1. Algorithm (SFS):

Input: segmented the pixel region growing O_a

Output: Selected subset feature H_n

Start

Step 1. Calculate the minimum subset of predefined features$H_n \leftarrow K$

Step 2. Calculates variables that assess the correlation direction and linear correlation.

$$L = \frac{f\sum_{a,b}^{} u_a u_b - \left(\sum_{a}^{} u_a\right)\left(\sum_{b}^{} u_b\right)}{\sqrt{f\left(\sum_{a}^{} u_a^2\right) - \left(\sum_{b}^{} u_b^2\right)}\sqrt{f\left(\sum_{b}^{} u_b^2\right) - \left(\sum_{a}^{} u_a^2\right)}} \quad (15)$$

Step 3. Compute the metric between a series of selected features and class labels.

$$h_x = 2\left[\frac{A_Z(v)}{S(u)+S(v)}\right] \quad (16)$$

$$h_x(ab) = c_z|h_x(aL) - h_x(bL)| \quad (17)$$

Step 4. Calculate the weights to measure the similarity between features

$$R = i * L + (1 - i) * h_x \quad (18)$$

Step 5. While Calculate the maximum number of features subsetEf_N do

Step 6. Evaluate the best fit value $\leftarrow J_{Ny} = 0$

Step 7. Compute the leukemia image subset feature$\leftarrow h_n$

For each feature $\leftarrow N_R$ but not K do

$K' = K \cap n$

If Compute the leukemia image feature selection $M(K') > J_{Ny}$ then

$$\frac{J_{Ny} = M(K')}{h_N = N}$$

End if

End for each

Step 8. Evaluate the selected subset of features$\leftarrow h_N$

Return h_N

Stop

At every node of the search tree, the image features can be organized for selection to improve accuracy, and the relationship between features can be assessed by measuring correlation direction and linear correlation. Additionally, the SFS algorithm can accurately predict the optimal features for leukemia imaging. Let's assume

L-Correlation, u-data matrix variable, v-column vector, a_Z −information gain, $s(u), s(v)$ −entropy variable, a and b-feature, h_X −measure information value, R-measure, Ef_N −maximum number of feature, N_R −available leukemia feature image, K-subset feature, $M(K')$ −evaluation of selected subset feature.

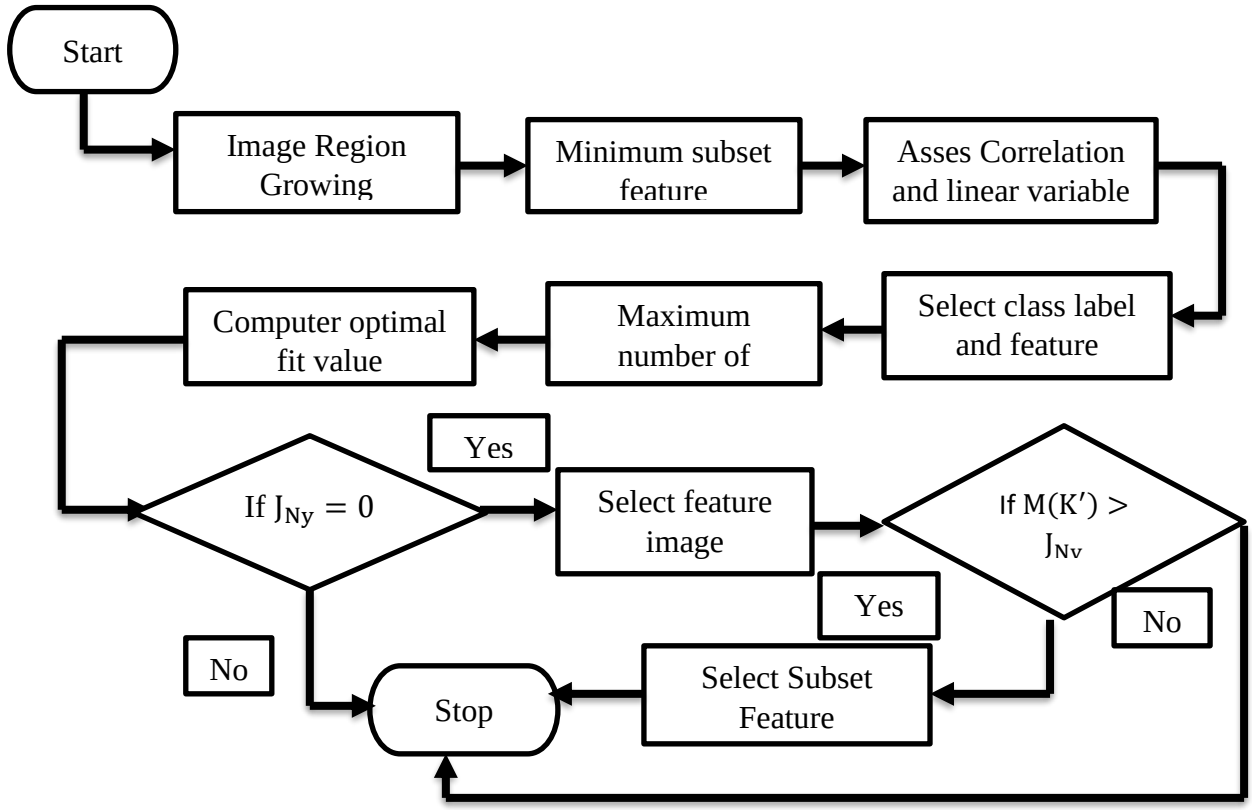


Figure 4. Feature Selection Flowchart Diagram based on SFS

As shown in figure 4, the sequential feature selection process can be depicted in a flowchart diagram. Furthermore, the leukemia image features are determined by analyzing the program vector values according to the class labels, the features representing the rows points and the data matrix columns can be selected.

3.5 Random Nearest Neighbor (RNN)

In this section, accuracy can be improved by utilizing image processing with a random forest method to classify cancerous or non-cancerous cases. The leukemia feature image can be randomly selected from the attributes of the nodes in the subset by implementing a random selection of attributes at each node in the decision tree. Each decision tree can be differentiated from the others and the RNN technique can be used to optimize the diversity of the system including maximum information gain rate and minimum Gini index. The Euclidean distance measure of the linear distance between two data points can be analyzed using the nearest neighbor method. In leukemia image processing, the square root of the distance between two data points is evaluated. The nearest-neighbor algorithm is then applied to assess the weight of the inverse distance for classifying the data as cancer or non-cancer.

Equation (19) demonstrates that the random forest algorithm is capable of estimating the information gain and Gini index by dividing the sample based on attributes in node assignment optimization. Let's assume i-attribute, O-sample set, Z_a −gini index, Z_i −gain, E-entropy, O^y −branch of node, y-value.

$$O = h_N\{Z_i(O,i) = M(O) - \sum_{Y=1}^{y} \frac{|O^y|}{O} M(O^y)\, Z_a(O,i) \sum_{Y=1}^{y} \frac{|O^y|}{O} Z_a(O^y)\, M(O^y) = -\sum_{Q=1}^{|v|} R_Q R_Q \;\; Z_a(O) = \sum_{Q=1}^{|v|} \sum_{Q' \neq Q} R_Q {R_Q}^1 = 1 - \sum_{Q=1}^{|v|} {R_Q}^2 \quad (19)$$

As shown in Equation 20, calculate the weighting coefficients of joint node area and feature area in the adaptive parameter selection process. Let's assume h_w −splitting, α, β −coefficient weight, s-minimal value.

$$s = \frac{min}{\alpha,\beta \in P} N\{O,i\} = \alpha Z_a(O,i) - \beta Z_i(O,i) \quad (20)$$

$$h_w = \{\alpha + \beta = 1\ 0 \leq \alpha, \beta \leq 1 \quad (21)$$

Error rate and accuracy are utilized to evaluate the effectiveness of RNN techniques. Furthermore, the classification error rate for the sample set is calculated as depicted in equations 22 and 23. Where N-feature vector, M-error rate, i-accuracy rate, e-measure.

$$M(N,O) = \frac{1}{e} \sum_{a=1}^{e} I\, I(N(u_a) \neq I_a) \quad (22)$$

$$i(N,O) - \frac{1}{e} \sum_{a=1}^{e} I\, I(N(u_a) = I_a) = 1 - M(N,O) \quad (23)$$

In Equation 24, the Euclidean distance measure for an N-dimensional vector plane between two data points can be calculated using the RNN technique. Let's assume $M_O -$ Euclidean distance, u_a, v_a −dimensional vector plane, u and v- two data points.

$$M_O(u_a, v_a) = \sqrt{\sum_{a=1}^{f} (u_a, v_a)^2} \quad (24)$$

The weight variable in the distance calculation is presented in equation 25 as the reciprocal of the distance. Additionally, assess the distance measurement function in the weighted nearest-neighbour algorithm.

$$T(u_a, v_a) = \frac{1}{o(u_a, v_a)} \quad (25)$$

$$\hat{O}(u_a, v_a) = \frac{\sum_{a=1}^{f} T(u_a, v_a) * O(u_a, v_a)}{\sum_{a=1}^{f} T(u_a, v_a)} \quad (26)$$

Analyzing leukemia images in the diagnosis of lymphoma and acute lymphoblastic leukemia (ALL) can accurately predict the classification of cancerous or non-cancerous cases. By randomly selecting properties from each node in the decision tree, the leukemia features image can be classified based on the subset node.

4. Result and Discussion

This section can be utilized to categorize blood cancer or non-cancer using image processing with an RNN model tailored for testing and training on a blood cancer dataset. Moreover, the performance evaluation of sensitivity, precision, recall, F1 score, and low error rate can be assessed using the blood cancer dataset. In addition, the proposed RNN technique can be used to classify leukemia cancer types such as ALL and lymduring image processing. Similarly, the website https://www.kaggle.com/datasets/mohammadamireshraghi/blood-cell-cancer-all-4class can evaluate leukemia diagnoses and differentiate them as cancerous or non-cancerous. The accuracy of these blood cancer image datasets will be compared and characterized with the proposed RNN method and other methods derived from previous techniques like ANN, SVM, and ResNet.

Table 1. Simulation Parameter

Simulation	Value
Dataset Name	Blood Cancer dataset
No of Dataset	38528
Language	Python

Tool	Jupyter
ALL Train Data	11679
ALL Test Data	6218
lymphoma Train Data	15000
lymphoma Test Data	5631

Leukemia images obtained from the blood cancer dataset were used to predict leukemia image accuracy for ALL and lymphoma classification using the Python-based Jupyter tool, as shown in table 1.

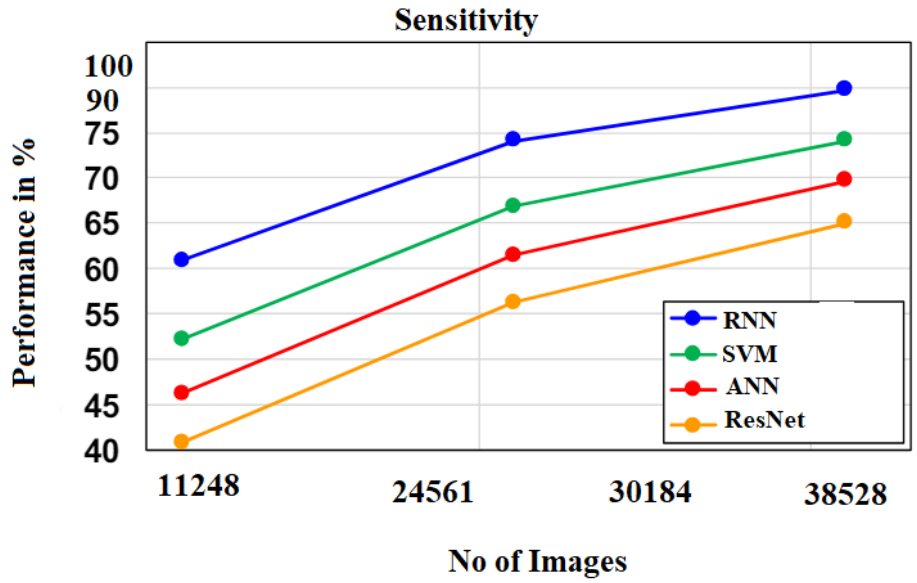


Figure 5. Analysis of Sensitivity

As depicted in figure 5, sensitivity analysis can forecast the precision of categorizing leukemia images from the blood cancer dataset into ALL and lymphoma. Additionally, the precision of sensitivity analysis to 79.35% when predicting blood cancer or non-cancer utilizing the suggested RNN method. Furthermore, the precision of the suggested method has enhanced to 63.23%, 67.8%, and 70.12% in comparison to SVM, ANN, and ResNet methods extracted from existing literature.

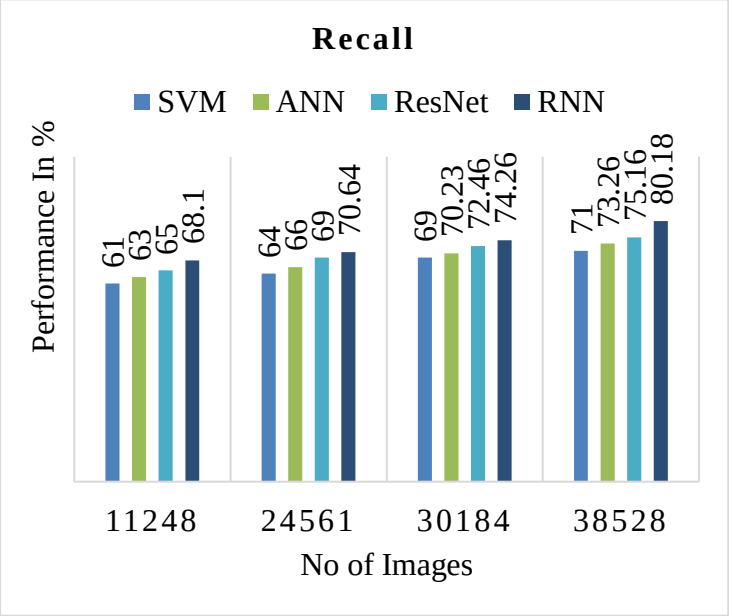


Figure 6. Analysis of Recall

In figure 6, the recall analysis demonstrates the ability to predict the accuracy of classifying leukemia images as ALL and lymphoma from the blood cancer dataset. Utilizing the proposed RNN method for predicting blood cancer or non-cancer resulted in a recall analysis accuracy of 80.18%. Furthermore, the proposed method exhibited enhanced accuracy of 68.1%, 70.64, and 74.26% compared to RESNET, SVM, and ANN methods mentioned in previous literature.

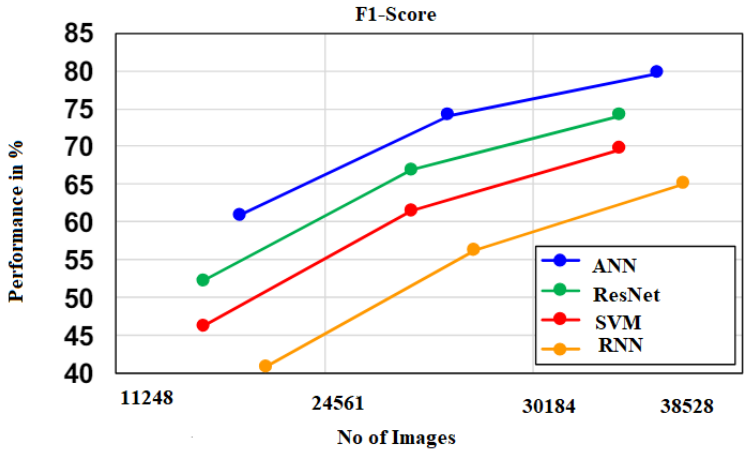


Figure 7. Analysis of F1-Score

Figure 7 indicates the F1-score analysis, demonstrating its effectiveness in predicting the accuracy of classifying leukemia images as ALL and lymphoma from the blood cancer dataset. Furthermore, the proposed method exceeds ResNet, ANN, and SVM methods (69.28%, 73.56%, and (82.63%) mentioned in previous approaches. Through the utilization of the proposed RNN method, the F1-score analysis achieved an accuracy of 79.16% in predicting leukemia or non-cancer.

Figure 8. Analysis of Low Error Rete

Figure 8 depicts the analysis of low error rates, demonstrating its effectiveness in predicting the accuracy of classifying leukemia images into ALL and lymphoma within the blood cancer dataset. Furthermore, compared to other methods documented in the literature, such as ResNet-27.56%, ANN-29.4%, and SVM-31.28%, this method exhibits lower error rate when used for classifying blood cancer data. Similarly, the proposed RNN approach yielded a low error rate analysis estimate of 24.7% for predicting leukemia or non-cancer.

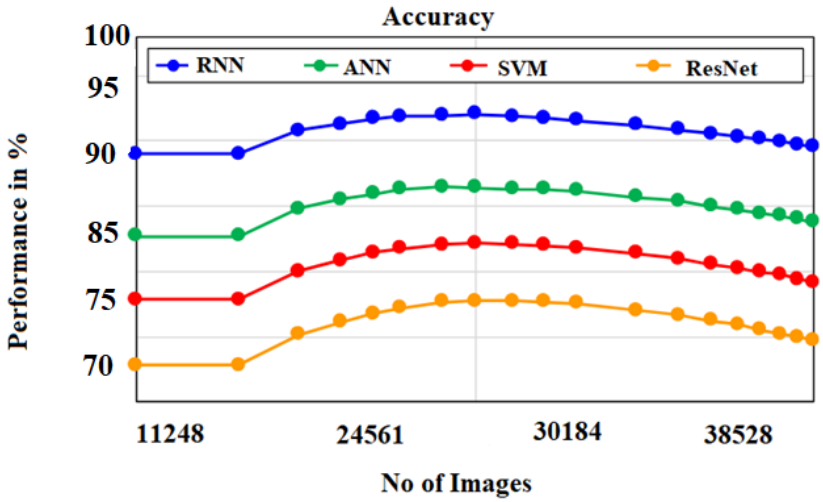


Figure 9. Analysis of Classification Accuracy

The evaluation of classification accuracy is outlined in figure 9 to assess the efficacy of the proposed methodology in predicting the classification of leukemia images as ALL and lymphoma using the blood cancer dataset. In comparison to various methodologies indicated in existing literature, such as ANN with an accuracy of 74.69%, ResNet with 79.35%, and SVM with 86.7%, the proposed method demonstrates comparatively lower accuracy in classifying leukemia data. Moreover, the proposed RNN approach achieved a notably high accuracy rate of 94.7% in the analysis of classification accuracy for predicting leukemia or non-cancer conditions.

5. Conclusion

In our research paper, a machine learning-based RNN method is proposed for classification of blood cancer images. Similarly, this study examines the blood cancer dataset to assess the narrative and comprehensiveness of preprocessing models. Use preprocessing models to improve image quality and facilitate meaningful feature selection based on bilateral filters. The pre-processed image is then identified and grouped based on the similarity criteria of the region extension segmentation method. After pre-processing and segmentation, we select the features of the dataset images based on the continuous feature selection method. In this way, it can be used to remove redundant features from large data sets or to select useful features by gradually adding them. Based on the analysis, the proposed RNN method can analyze blood cancer dataset images of blood samples to recognize abnormal cells exhibiting signs of blood cancer. This method improves the ability to distinguish between ALL and lymphoma based on non-cancerous and cancerous. Furthermore, the RNN method performed well and achieved an accuracy of 94.7%, sensitivity of 79.35%, F1 score of 82.63%, recall of 80.18%, and low error rate of 24.7% using the blood cancer dataset. Furthermore, based on an exhaustive assessment of the proposed method's accuracy rate, the RNN achieved higher accuracy compared with previous methods. The RNN method provides a promising solution for accurately classifying leukemia images and provides a robust and efficient method for the initial diagnosis and medication of leukemia.

References

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

[2] T. Mustaqim, C. Fatichah and N. Suciati, "Deep Learning for the Detection of Acute Lymphoblastic Leukemia Subtypes on Microscopic Images: A Systematic Literature Review," in IEEE Access, vol. 11, pp. 16108-16127, 2023, doi: 10.1109/ACCESS.2023.3245128.

[3] T. A. M. Elhassan, M. S. M. Rahim, T. T. Swee, S. Z. M. Hashim and M. Aljurf, "Feature Extraction of White Blood Cells Using CMYK-Moment Localization and Deep Learning in Acute Myeloid Leukemia Blood Smear Microscopic Images," in IEEE Access, vol. 10, pp. 16577-16591, 2022, doi: 10.1109/ACCESS.2022.3149637.

[4] M. A. Jopek et al., "Deep Learning-Based, Multiclass Approach to Cancer Classification on Liquid Biopsy Data," in IEEE Journal of Translational Engineering in Health and Medicine, vol. 12, pp. 306-313, 2024, doi: 10.1109/JTEHM.2024.3360865.

[5] M. A. Hossain, A. K. M. M. Islam, S. Islam, S. Shatabda and A. Ahmed, "Symptom Based Explainable Artificial Intelligence Model for Leukemia Detection," in IEEE Access, vol. 10, pp. 57283-57298, 2022, doi: 10.1109/ACCESS.2022.3176274.

[6] A. Ratley, J. Minj and P. Patre, "Leukemia Disease Detection and Classification Using Machine Learning Approaches: A Review," 2020 First International Conference on Power, Control and Computing Technologies (ICPC2T), Raipur, India, 2020, pp. 161-165, doi: 10.1109/ICPC2T48082.2020.9071471.

[7] V. Acharya, V. Ravi, T. D. Pham and C. Chakraborty, "Peripheral Blood Smear Analysis Using Automated Computer-Aided Diagnosis System to Identify Acute Myeloid Leukemia," in IEEE Transactions on Engineering Management, vol. 70, no. 8, pp. 2760-2773, Aug. 2023, doi: 10.1109/TEM.2021.3103549.

[8] A. Batool and Y. -C. Byun, "Lightweight EfficientNetB3 Model Based on Depthwise Separable Convolutions for Enhancing Classification of Leukemia White Blood Cell Images," in IEEE Access, vol. 11, pp. 37203-37215, 2023, doi: 10.1109/ACCESS.2023.3266511.

[9] Shruti, S. Pahadsingh and B. Appasani, "Hexagonal Tiled Terahertz Metamaterial Absorber for Cancer Detection Incorporating Machine Learning," in IEEE Sensors Journal, doi: 10.1109/JSEN.2024.3384972.

[10] Kokeb Dese, Hakkins Raj, Gelan Ayana, Tilahun Yemane, Wondimagegn Adissu, Janarthanan Krishnamoorthy, Timothy Kwa, Accurate Machine-Learning-Based classification of Leukemia from Blood

Smear Images, Clinical Lymphoma Myeloma and Leukemia, Volume 21, Issue 11, 2021, Pages e903-e914, ISSN 2152-2650, https://doi.org/10.1016/j.clml.2021.06.025.

[11] Urbanos, Gemma, Alberto Martín, Guillermo Vázquez, Marta Villanueva, Manuel Villa, Luis Jimenez-Roldan, Miguel Chavarrías, Alfonso Lagares, Eduardo Juárez, and César Sanz. 2021. "Supervised Machine Learning Methods and Hyperspectral Imaging Techniques Jointly Applied for Brain Cancer Classification" Sensors 21, no. 11: 3827. https://doi.org/10.3390/s21113827

[12] Cengil E, Çınar A, Yıldırım M. A hybrid approach for efficient multi-classification of white blood cells based on transfer learning techniques and traditional machine learning methods. Concurrency Computat Pract Exper. 2022; 34(6):e6756. doi:10.1002/cpe.6756

[13] K. Sridhar, Ajay Reddy Aeruva, Renjith PN, Asmita Dixit, Atif Jamshed, Ravi Rastogi, "Enhanced Machine learning algorithms Lightweight Ensemble Classification of Normal versus Leukemic Cel", Journal of Pharmaceutical Negative Results, pp. 496–505, Nov. 2022, doi: 10.47750/pnr.2022.13.S09.056.

[14] V. Durga Prasad Jasti, Abu Sarwar Zamani, K. Arumugam, Mohd Naved, Harikumar Pallathadka, F. Sammy, Abhishek Raghuvanshi, Karthikeyan Kaliyaperumal, "Computational Technique Based on Machine Learning and Image Processing for Medical Image Analysis of Breast Cancer Diagnosis", Security and Communication Networks, vol. 2022, Article ID 1918379, 7 pages, 2022. https://doi.org/10.1155/2022/1918379

[15] Saleem, S., Amin, J., Sharif, M. et al. A deep network designed for segmentation and classification of leukemia using fusion of the transfer learning models. Complex Intell. Syst. 8, 3105–3120 (2022). https://doi.org/10.1007/s40747-021-00473-z

[16] Jung, C., Abuhamad, M., Mohaisen, D. et al. WBC image classification and generative models based on convolutional neural network. BMC Med Imaging 22, 94 (2022). https://doi.org/10.1186/s12880-022-00818-1

[17] Ghaderzadeh M, Aria M, Hosseini A, Asadi F, Bashash D, Abolghasemi H. A fast and efficient CNN model for B-ALL diagnosis and its subtypes classification using peripheral blood smear images. Int J Intell Syst. 2022; 37: 5113-5133. doi:10.1002/int.22753

[18] Karthick, T., Ramprasath, M., Sangeetha, M. (2022). Classification of Blast Cells in Leukemia Using Digital Image Processing and Machine Learning. In: Manogaran, G., Shanthini, A., Vadivu, G. (eds) Proceedings of International Conference on Deep Learning, Computing and Intelligence. Advances in Intelligent Systems and Computing, vol 1396. Springer, Singapore. https://doi.org/10.1007/978-981-16-5652-1_1

[19] Chand, S., Vishwakarma, V.P. A novel Deep Learning Framework (DLF) for classification of Acute Lymphoblastic Leukemia. Multimed Tools Appl 81, 37243–37262 (2022). https://doi.org/10.1007/s11042-022-13543-2

[20] Simsek, E., Badem, H., Okumus, I.T. (2022). Leukemia Sub-Type Classification by Using Machine Learning Techniques on Gene Expression. In: Yang, XS., Sherratt, S., Dey, N., Joshi, A. (eds) Proceedings of Sixth International Congress on Information and Communication Technology. Lecture Notes in Networks and Systems, vol 217. Springer, Singapore. https://doi.org/10.1007/978-981-16-2102-4_56

[21] Thanmayi A, V.L., Reddy, S.D., Kochuvila, S. (2021). Detection of Leukemia Using K-Means Clustering and Machine Learning. In: Kumar, N., Vinodhini, M., Venkatesha Prasad, R.R. (eds) Ubiquitous Communications and Network Computing. UBICNET 2021. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 383. Springer, Cham. https://doi.org/10.1007/978-3-030-79276-3_15

[22] Hossain, M.A. et al. (2021). An Effective Leukemia Prediction Technique Using Supervised Machine Learning Classification Algorithm. In: Kaiser, M.S., Bandyopadhyay, A., Mahmud, M., Ray, K. (eds) Proceedings of International Conference on Trends in Computational and Cognitive Engineering. Advances in Intelligent Systems and Computing, vol 1309. Springer, Singapore. https://doi.org/10.1007/978-981-33-4673-4_19

[23] Zakir Ullah, Muhammad, Yuanjie Zheng, Jingqi Song, Sehrish Aslam, Chenxi Xu, Gogo Dauda Kiazolu, and Liping Wang. 2021. "An Attention-Based Convolutional Neural Network for Acute Lymphoblastic Leukemia Classification" Applied Sciences 11, no. 22: 10662. https://doi.org/10.3390/app112210662

[24] S. Rajeswari, C. S. Vasanth, C. Bhavana and K. S. S. Chowdary, "Detection and Classification of Various Types of Leukemia Using Image Processing, Transfer Learning and Ensemble Averaging Techniques," 2022 2nd Asian Conference on Innovation in Technology (ASIANCON), Ravet, India, 2022, pp. 1-6, doi: 10.1109/ASIANCON55314.2022.9909377.

[25] Saeed, U., Kumar, K., Khuhro, M.A. et al. DeepLeukNet—A CNN based microscopy adaptation model for acute lymphoblastic leukemia classification. Multimed Tools Appl 83, 21019–21043 (2024). https://doi.org/10.1007/s11042-023-16191-2