Blood Cancer Detection Using Improved Machine Learning Algorithm

3 authors, including:

N P Dharani
Sree Vidyanikethan Engineering College
8 PUBLICATIONS   **45** CITATIONS

Blood Cancer Detection Using Improved Machine Learning Algorithm

N. P. Dharani
Assistant Professor,
Department of Electronics and
Communication Engineering,
Mohan Babu University
Tirupati, A.P, India.
npdharani91@gmail.com

G. Sujatha
Professor,
Department of Electronics and
Communication Engineering
Sri Venkateswara College of Engineering
(SVCE), Tirupati, A.P, India.
sujatha.g@svcolleges.edu.in

R. Rani
Associate Professor,
Department of Electronics and
Communication Engineering,
VEMU Institute of Technology,
A.P, India.
ranivemuece@gmail.com

*Abstract-*The occurrence of blood cancer has been on rise over the last decade, and treatment of this disease begins as soon as possible following a correct diagnosis. There are a variety of tests and medical experts involved in the diagnostic process, which is time-consuming and expensive. In order to make an accurate prediction of its outcome, an automatic diagnosis system is necessary. This paper suggests a method for blood cancer detection using Improved Machine Learning Algorithm like Ensemble Method with the combination of Effective Fuzzy C Means (EFCM) and Iterative Morphological Process (IMP). The use of EFCM and IMP techniques helps to segment and analyse the blood image data, allowing for the identification of specific characteristics associated with blood cancers. This segmentation process enables the algorithm to focus on the relevant regions of interest, facilitating more accurate and targeted detection of cancerous cells. Moreover, a pre-processing and enhancement methods have been used to the image of blood. By utilizing Machine Learning to process images of blood cancers, accurate diagnosis is achieved, diagnosis times is reduced, and diagnostic testing is provided faster, cheaper, and with greater safety.

Key words- Blood Cancer Detection, Ensemble Model, EFCM, IMP, Improved Machine Learning, Image Enhancement.

I. INTRODUCTION

Cancer is one of the major diseases which needs to be taken care of in its early stages; otherwise, the excess cancer cells cause the most damage to the body and weaken the person. It is a priority to detect these cancer cells at an early stage to cure them simple and cause no harm to the person's life by those cells. If it is possible to find cancer cells before proceeding to further stages, then it is easy to retain many lives. Many people cannot afford to spend money to cure this cancer or to test it, so the main aim is to take this test for at low cost as possible so that everyone will be able to afford it and be able to cure it at an early stage with no harm to their lives [1]. To help us with all this, it is essential to develop a system that would help us detect cancer cell and give output accordingly. As a consequence, machine learning is an option here to predict these cells and provide an accurate yield [2].

As per the American Cancer Society, 48,610 individuals (27,880 men and 20,730 women) are estimated to develop blood cancer, and 23,720 of those individuals will die from it [3]. This condition falls within the category of hematological neoplasms which encompass diseases affecting the blood, lymphatic system and bone marrow. There can be a securedebility in the number of normal blood cells as an outcome of blood cancer cells over time [4]. There can be serious consequences in the form of anemia, bleeding, and infections owing of this condition.

In addition to spreading to lymph nodes and other organs, blood cancer cells can also cause pain or swelling. Blood cancers appears in a variety of forms. Most bones contain a soft tissue called bone marrow. Blood cells are made in a bone marrow. A white blood cell is responsible for fighting infections in the body [5]. All parts of our body receive oxygen through red blood cells. During the clotting process, platelets play an important role. White blood cells do not perform the same function as normal white blood cells. It grows faster and does not cease to grow when it should, unlike normal cells [6]. As time passes, cancerous cells in the blood crowd out healthy cells. Infections, anemia, and bleeding may result as a result of this.

Blood cancer is classified into several types. Generally, blood cancers are classified according to their rate of progression and type of white blood cells they affect. Acute or chronic conditions may occur [7]. This may feel sick right away if you have acute blood cancer, as it progresses extremely quickly. There may be years before symptoms of chronic blood cancer begin to appear. In some cases, the disease may be lymphocytic, while in others it may be myeloid [8]. Lymphocytic (or lymphoblastic) blood cancer occurs within lymphocytes, an important component of white blood cells. Myelogenous blood cancer can also develop in granulocytes, red blood cells, or platelets, among other types of blood cells [9].

II. RELATED WORKS

Cancer is an incurable disease. The secret to an effective cancer treatment plan is early detection of malignancy. Techniques employed in medical image processing are crucial for disease diagnosis. Leukaemia occur as a result of white blood cells (WBCs) that are either immature or normal. WBC is typically a group that promotes a fight against potentially contagious cells in human body [10].

These cells' malfunction causes leukaemia, which is a disease that results from it [11]. The approach used in [11] has the draw back of high liability to overfitting. It used to be an exhausting and laborious method to diagnose diseases and count a number of cells in the blood. Hence, a deep learning-based approach is presented for this task in [12]. However, this technique has annotation and labelling challenges. The task of manually determining the number of cells that exist in

the blood is carried out by the use of a device known as a haemocytometer in the specialty of medicines. Nevertheless, counting requires more time, and as a result, the results are less trustworthy. A software-based answer has been offered to deal with these issues with the use of microscopic images. [13]. But, the use of the CMYK colour space for white blood cell cancer detection has certain limitations. This type of image processing makes it possible to assess a person's prospective risk of acquiring leukaemia [14].

The accurate detection of leukaemia, also known as the exact distinction of cancerous leukocytes with low expense in the initial diagnosis, is a major challenge in a field of disease diagnosis [15]. This is due to the fact that an earlier leukaemia is detected and diagnosed, more effective treatment is administered. There is a scarcity of cytometry technology, and the procedures that are accessible at laboratory health testing take a significant amount of time [16]. This is despite the fact that leukaemia is quite common.

At first, there were 116 items that were retrieved. Following the application of the criteria for inclusion and exclusion, the sample size of the study was reduced [17]. Processing leukaemia smear pictures using the ML approach may enhance accuracy, decrease the amount of time needed for diagnosis, and offer diagnostic services that are quicker, less expensive, and safer. Experts in clinical and laboratory

settings may also use machine learning techniques in diagnostic apps and tools, in addition to the diagnostic approaches that are now in use.

The aforementioned limitations are overcome by integrating the ensemble method with EFCM and IMP. Here, the suggested blood cancer detection system benefits from the combined strengths of different algorithms and methodologies. This leads to improved accuracy, reduced diagnosis time, faster and more cost-effective testing, and enhanced safety in the diagnosis of blood cancers.

III. PROPOSED METHODOLOGY

Through a procedure of region growing, the boundary of the tumor is determined as well as growth rate of the tumor. Growth of a region is the process of grouping pixels together. The process of calculating neighbouring pixels is known as region growing. A region is grown by dividing it into smaller areas and then dividing them into larger areas. In order to assign pixels for growing the region, every pixel should have the same set of properties. This paper proposes an approach for blood cancer detection using Effective Fuzzy C Means (EFCM) and Iterative Morphological Process (IMP) as shown in Fig. 1.
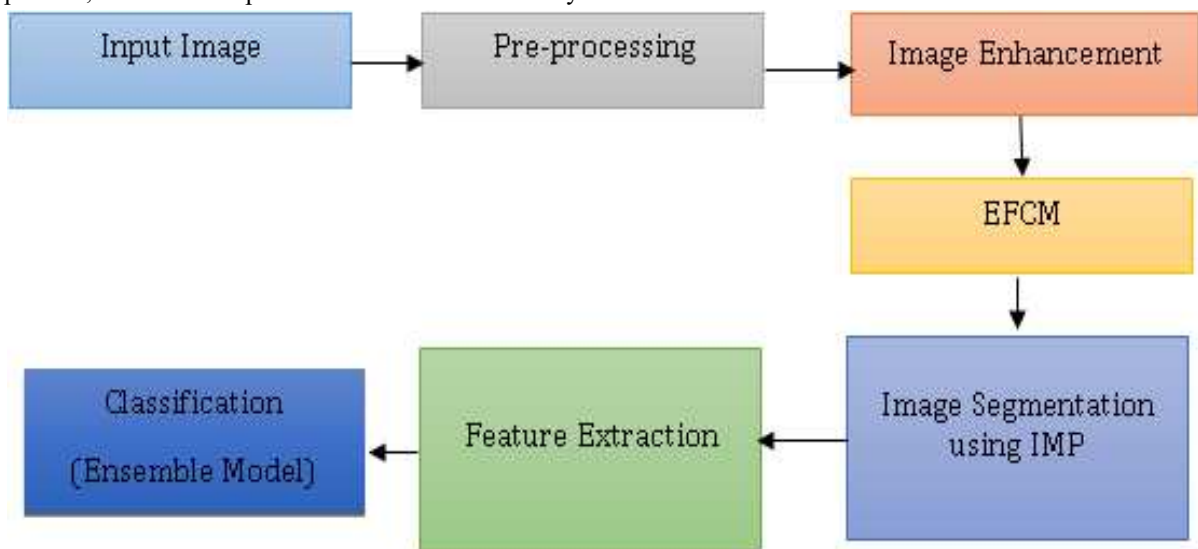

Fig. 1. Overall Workflow of our Proposed System

In order to achieve this goal, a pre-processing and enhancement methods have been used to the image of blood. Using the Machine Learning Ensemble method to process Imagery of blood cancers may improve diagnosis accuracy, shorten diagnostic times, and facilitate quicker, more cost-effective, and more safe diagnostic procedures.

A. Pre-processing

Images that have been processed to the highest level of abstraction are commonly called "pre-processing", whose inputs and outputs are usually images of intensity. Pre-processing is intended to enhance the image data by removing undesirable distortions and enhancing certain parts of the image that are necessary for the following processing step. The purpose of this procedure is to achieve this goal.

B. EFCM

Enhanced fuzzy c-mean algorithms are among the commonly used algorithms to segment images by grouping pixels with similar values into cluster regions. It is appropriate to use fuzzy clustering for the segmentation of medical images. The EFCM algorithm is compared to the fuzzy variation of k-means algorithm. A clustering algorithm that assigns a degree of belonging to each cluster based on membership. In this algorithm, a weighted sum of squared errors within group objective function is minimized in order to obtain an optimal c partition. It is widely used for segmentation of images and recognition of patterns.

Algorithm 1: Our EFCM algorithm

Step 1: Each item in Data has a frequency H.

Step 2: Vectorize I = max_Data: min_Data

Step 3: At least two random centroides should be chosen.

Step 4: Calculate the membership matrix as follows:

$$u_{IJ} = \frac{1}{\sum_{k=1}^{C} \left[\frac{I_I - C_J}{I_I - C_K}\right]^{\frac{2}{M-1}}} \qquad (1)$$

Step 5: For cluster centers, follow these steps:

$$c = \frac{\sum_{I=1}^{N} u^M * h * i}{\sum_{I=1}^{N} u^M * h} \qquad (2)$$

It is important to note that the EFCM uses frequency values rather than actual values. For example, that there will be more than 256 values in gray images, and this will increase processing time however, the enhanced algorithm will process only 256 items, at worst. There is no requirement for the proposed algorithm to be based on the full data of the original image, but rather on data representing how frequently each item occurs within the original image. There can be a maximum of 256 frequencies.

C. Segmentation using IMP

An essential characteristic of morphology is erosion, in which pixels are removed peripheries are remove from a feature, or enlargement, in which pixels are added to the periphery of the feature. A pixel whose background neighbor shares an edge or corner will be removed by classical erosion. The classic dilation, on the other hand, adds any pixel in the feature that is touched by a pixel in the input image. The size of features or structures is altered by both of these operations, and as a result, they are used together most frequently. A closing is an erosion that is followed by a dilation, whereas an opening is an erosion that is followed by a dilation. As a result of these operations, irregular borders can be smoothed, and small lines or noises can be filled in or removed.

Iterative morphological operations are illustrated in this proposed model for binary images. To add or remove more pixels, the tutorial allows you to apply each function multiple times. During an opening that is iterated M times, there are M steps of erosion, then M steps of dilation, and in the reverse order.

Algorithm 2: Iterative Morphological Processing

function M = morp (I, K)

```
I = im2double (I);
I = im2double (I);
for n = 1: M
for i = 1: size (F, 1)
        [Distance, CN] = min (DAL (i, 1: K));
DAL (i, K+2) = Distance;
end for
for i = 1: M
ifsum_CENTS (:) = 0
for M= 1: size (image)
CENTS (iterative) = F (random (size (iterative)), :);
end for
end for
end
```

D. Feature Extraction

It is common to find that in machine learning, matching patterns, and the analysis of images, features are extracted from a set of measurements. In order to construct meaningful and non-redundant derived values, derived values have to be constructed. In some cases, this also leads to improved human understanding for future studies and generalisation steps. There is a connection between feature reduction and feature extraction.

E. Classification

To classify an image, a number of visual properties must first be numerically analysed, and the findings must then be arranged into various groups. During the testing step that follows, these functionality divisions are used in order to categorise picture characteristics. In the process of analysing blood cells, differential counts of RBCs and WBCs both have important roles to play in the diagnosis of a wide range of diseases. There are several segmentation and classification strategies that are used for the purpose of distinguishing between the numerous subtypes of WBCs and those of RBCs.

In machine learning, ensemble learning refers to combining predictions from multiple models to enhance the performance as shown in fig. 1.

As a general principle, ensemble learning is based on the concept of "the wisdom of the crowd". The accuracy of predictions can be improved by combining multiple models.
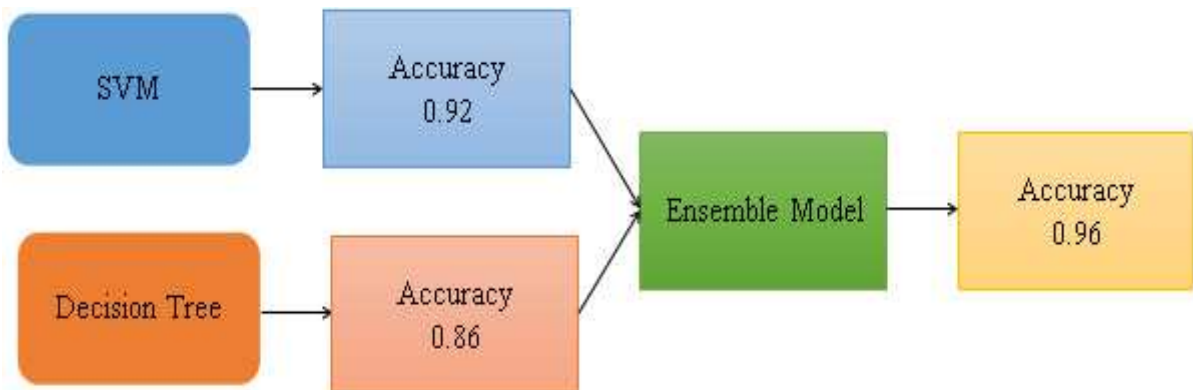

Fig. 2. Ensemble Model with Accuracy Value

It is a combination of multiple learning algorithms or models combining multiple algorithm or model combinations to produce an optimal prediction model as shown in fig. 2. In comparison with the base learners taken alone, the model produced has a higher performance. A number of other examples of ensemble learning can be found in the selection of the most significant characteristics, data fusion, etc.

Algorithm: Creating Ensemble Model

```
defensemble_dataset (members, inputX):
          ensemble_X = 0
          for model in members:
                    prediction = model.predict (X, v=0)
                    ifensemble_X is 0:
                              ensemble_X            =
prediction_model
                    else:
                              ensemble_X = ensemble
_output ((X, prediction))
                    returnensemble_X
```

Algorithm: Classification and Prediction using Ensemble Model

```
defensemble_prediction (members, model, X):
          # using dataset of ensemble
          ensemble_X = ensemble_dataset (members, X)
          # make a prediction
          prediction = model.predict (ensemble_X)
          return prediction
```

From the above method, concluding Stacking models can provide more efficient results than single models.

IV. RESULTS AND DISCUSSIONS

A technique for segmenting brain tumours is preferred to deliver improved final outcomes in the analysis of blood cell images. The effectiveness of the suggested blood cancer detection process using improved machine learning algorithm is evaluated in python software using Blood cell image dataset. Initially, the dataset is collected and pre-processed, following which the relevant features are extracted. Then the ensemble method is implemented and the EFCM clusters the extracted features of the blood cell images. Subsequently, the segmentation and classification results are refined using IMP. Finally, the ensemble model is trained and its performance is evaluated. As stated in the following description, accuracy means that an outcomes are accurate.

$$Accuracy = \frac{(Tp+TN)}{(TP+TN+FN+FP)} \times 100 \qquad (3)$$

A mean square error (MSE) is a metric for signal or image fidelity. A signal reliability measure, also known as a fidelity measure, evaluates the degree of consistency or similarity between two images using a mathematical formula. If MSE is intended, we assume that one image is perfectly distinct while the other image has been altered in some way and is explained as follows:

$$MSE = \frac{1}{M} \times N\varepsilon\big(f(x,y) - F^r f(x,y)\big)^2 \qquad (4)$$

Peak Signal-to-Noise Ratio data is available from this database. It measures an overall ability to reassemble a processed image. The formula for this can be found below.

$$PSNR = \frac{20log10(2N-1)}{MSE} \qquad (5)$$

A signal to noise ratio has been found to be favourable when a MSE decreases and the PSNR is greater. The values of the two pixels are then added together to calculate the boundary displacement error. An output stages listed below are essential for finding blood malignancies. The input image is provided in fig. 3, and it is pre-processed and the resultant images are given in fig.4. Subsequently, after EFCM processing, the obtained output is given in fig.5. Following which the image is segmented using IMP and the obtained results are provided in fig.6. Finally, the classification results of the images are presented in fig.7.
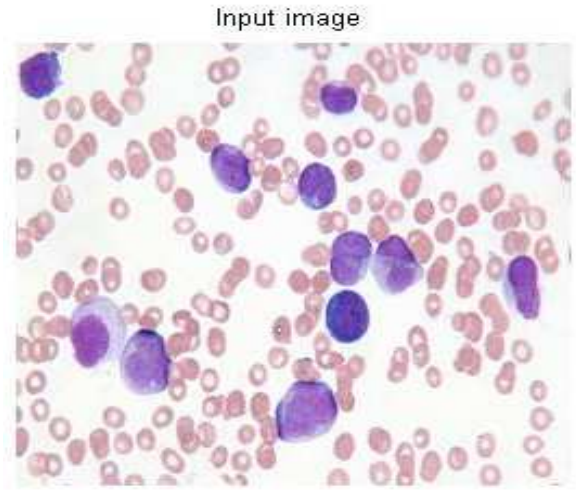

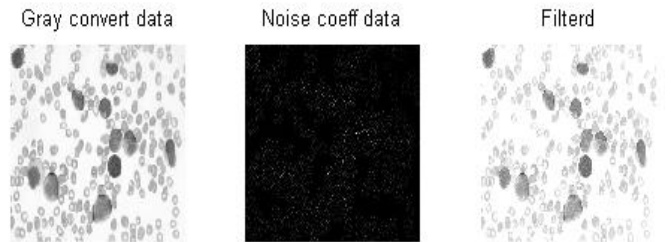Fig. 3. Input Image

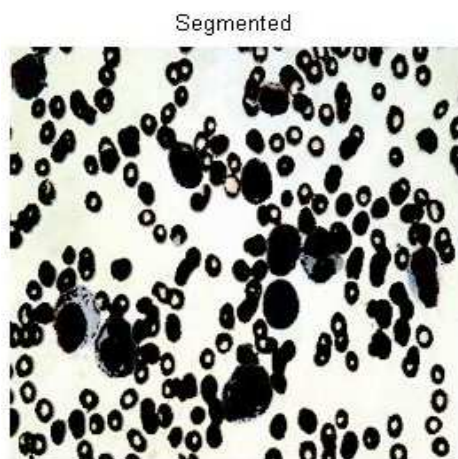Fig. 4. Data Pre-processing

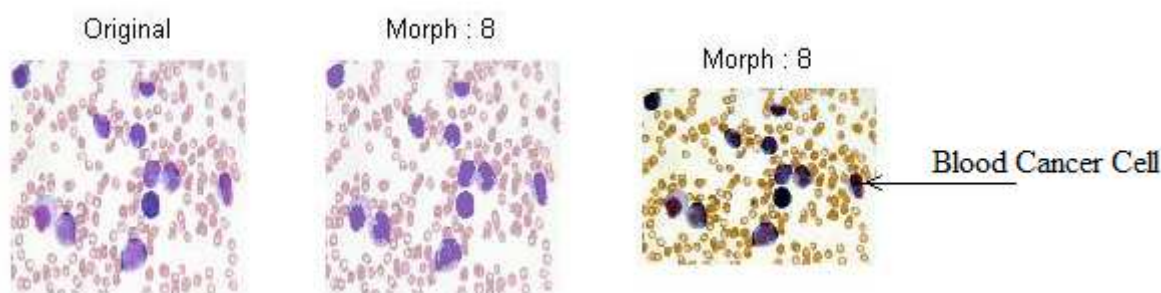Fig. 5. EFCM Processing

Fig. 6. Image Segmentation using IMP

Fig. 7. Classification of Detected Blood Cancer Cell

Table I. COMPARISON OF VARIOUS SEGMENTATION TECHNIQUES FOR ACCURACY

Algorithms	MSE	PSNR	Accuracy
SVM	23.2	20.2	91.5
Decision Tree	22.5	19.3	85.3
Proposed Ensemble Model	16.1	12.5	96.2

An accuracy of the ensemble model increases by 96.2 percent when compared to two algorithms, according to the findings of the present study, which are shown in Table 1.

V. CONCLUSION

A segmentation of a perfect aberrant area in blood cell pictures is an essential part of medical imaging applications. The accuracy of blood cell medical imaging is improved by segmenting them using a number of algorithms. These algorithms' effectiveness is assessed. As well as a results of experiments with Improved Machine Learning Algorithm like Ensemble Method by the combination of the new EFCM with IMP. The ground truth image is analyzed to compare the results with those of the new EFCM with IMP. Thereby, it is seen that the employment of the suggested machine learning algorithms for processing blood images, ensures that the diagnostic process is more accurate, reducing the chances of misdiagnosis. It also highlights the potential for faster and more cost-effective diagnosis, as well as enhanced safety for patients undergoing diagnostic testing. Overall, the integration of an improved machine learning algorithm, along with advanced image processing techniques, holds promise for enhancing the accuracy and efficiency of blood cancer detection. By leveraging the power of machine learning, accurate diagnosis is achieved, leading to timely treatment and improved patient outcomes.

REFERENCES

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

[2] S. Meher, B. Nayak and P. K. Das, "A lightweight deep learning system for automatic detection of blood cancer", Measurement, vol. 191, pp. 110762, 2022.

[3] E. Lee, W. Aljedaani, F. Rustam, V. Rupapara, H. F. Shahzad, and I. Ashraf, "Blood cancer prediction using leukemia microarray gene data and hybrid logistic vector trees model", Scientific reports, vol. 12, no. 1, pp. 1000, 2022.

[4] B. Alotaibi, M. Alotaibi, and M. E. Karar, "Intelligent medical iot-enabled automated microscopic image diagnosis of acute blood cancers," Sensors, vol. 22, no. 6, pp. 2348, 2022.

[5] V. A. Diya, S. Meher, R. Panda, P. K. Das, and A. Abraham, "A systematic review on recent advancements in deep and machine learning based detection and classification of acute lymphoblastic leukemia", IEEE Access, 2022.

[6] R. Ahmad, M. Awais, N. Kausar, and T. Akram, "White Blood Cells Classification Using Entropy-Controlled Deep Features Optimization," Diagnostics, vol. 13, no. 3, pp. 352, 2023.

[7] S. Singh, and Y. K. Prajapati, "Novel Bottom-Side Polished PCF-Based Plasmonic Biosensor for Early Detection of Hazardous Cancerous Cells," IEEE Transactions on NanoBioscience, 2023.

[8] U. Farahdina, A. S. Muliawati, V. Z. Zulfa, M. Firdhaus, I. Aziz, H. Suprihatin, D. Darsono, N. Nasori, and A. Rubiyanto, "Electrochemical and Optical Analysis of Various Compositions of Au and Ag Layers for Blood Cancer Prognosis," Coatings, vol. 13, no. 1, pp. 186, 2023.

[9] R. Raina, N. K. Gondhi, D. Singh, M. Kaur, and H. N. Lee, "A Systematic Review on Acute Leukemia Detection Using Deep Learning Techniques", Archives of Computational Methods in Engineering, pp. 1-20, 2022.

[10] Y. Wang, L. He, H. Xie, Y. Lu, J. Zhao, C. Li and Y. F. Li, "Non-targeted metallomics through synchrotron radiation X-ray fluorescence with machine learning for cancer screening using blood samples", Talanta, vol. 245, pp. 123486, 2022.

[11] P. M. Ameer and M. R. Reena, "A content-based image retrieval system for the diagnosis of lymphoma using blood micrographs: An incorporation of deep learning with a traditional learning approach", Computers in Biology and Medicine, vol. 145, pp. 105463, 2022.

[12] K. K. Anilkumar, V. J. Manoj, and T. M. Sagi, "Automated detection of b cell and t cell acute lymphoblastic leukaemia using deep learning", Irbm, vol. 43, no. 5, pp. 405-413, 2022.

[13] M. Muthumanjula, and R. Bhoopalan, "Detection of White Blood Cell Cancer using Deep Learning using Cmyk-Moment Localisation for Information Retrieval", Journal of IoT in Social, Mobile, Analytics, and Cloud, vol. 4, no. 1, pp. 54-72, 2022.

[14] M. E. Billah, and F. Javed, "Bayesian convolutional neural network-based models for diagnosis of blood cancer", Applied Artificial Intelligence, vol. 36, no. 1, pp. 2011688, 2022.

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

[16] G. Atteia, A. A. Alhussan, and N. A. Samee, "BO-ALLCNN: Bayesian-Based Optimized CNN for Acute Lymphoblastic Leukemia Detection in Microscopic Blood Smear Images," Sensors, vol. 22, no. 15, pp. 5520, 2022.

[17] M. S. M. Rahim, T. T. Swee, T. A. M. Elhassan, 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", IEEE Access, vol. 10, pp. 16577-16591, 2022.