

Detection and Classification of Blood Cancer using an Auto Encoder and a Regularised Extreme Learning Machine

Dr.Sharath M N¹, Dr. Rajesh T M², K.R Surabhi³, Sanjay M⁴, Nikshepa T⁵ and Sowmya S R⁶

^{1,4}Assistant Professor, Department of Computer Science and Engineering, Rajeev Institute of Technology, Hassan, India
Email: sharathmn64@rithassan.ac.in, sanjaym2612@rithassan.ac.in

²Associate Professor, Department of Computer Science and Engineering, Dayananda Sagar University, Bengaluru, India
Email: rajesh-cse@dsu.edu.in

^{3,5,6}Assistant Professor, Department of Information Science and Engineering Dayananda Sagar Academy of Technology and Management, Bengaluru, India
Email: kr.surabhi@dsatm.edu.in , nikshepa_t@dsatm.edu.in , sowmya_sr@dsatm.edu.in

Abstract—Leukemia is by far the most common type of blood cancer in humans. The length of time it takes to recover from an illness is determined by how quickly it is recognized and treated. Leukemia develops when the bone marrow generates an abnormally large amount of white blood cells. Because images of human blood samples are studied under the microscope, hematologists use techniques such as microscopic color imaging, image segmentation, grouping, and classification to diagnose patients as quickly and precisely as possible. Microscopy can identify blood cancer in visible and immature white blood cells in a variety of methods. Early and correct diagnosis of leukemia is critical for clinicians to properly treat patients. Following the submission of the input image, Image processing in image.

Index Terms— Convolutional Neural Network (CNN), Grey Level Co-occurrence Matrix (GLCM), Recurrent Neural Network (RNN).

I. INTRODUCTION

The basic cause of blood cancer in certain patients is that aberrant blood cells multiply at an excessive rate, impeding the creation of healthy blood cells. Lymphoma, leukaemia, and myeloma are the three most common kinds of blood cancer. White blood cell cancer, also known as acute lymphocytic leukaemia (ALL), begins in the bone marrow. As one of the major causes of mortality in the United States and around the world, cancer has a tremendous cultural impact. Leukaemia is a cancer that affects white blood cells produced in the bone marrow. The abnormal white blood cells proliferate in an uncontrollable monoclonal fashion. Lymphocytes, which are part of the immune system and combat infection, are where lymphomas start. The majority of lymphoma instances fall into one of two categories: Hodgkin lymphoma or non-Hodgkin lymphoma. Follicular lymphomas and B-cell lymphomas progress at various rates and respond well to different therapies. Reed- Sternberg cells are utilised to distinguish Hodgkin lymphoma from other lymphomas. Reed-Sternberg cells, which form when normal B cells degenerate, are indicative of Hodgkin lymphoma. Reed-Sternberg cells are large and have many nuclei. Despite extensive awareness of its frequency, the causes of non-Hodgkin

lymphoma are unknown. [1][2].

Lymphoma can be treated at a hospital. This emphasises the critical importance of detecting and treating acute lymphoblastic leukaemia as soon as feasible. A morphological examination can aid in the diagnosis of acute lymphoblastic leukaemia.

The presence of a considerable number of malignant cells in the bone marrow is conclusive evidence of acute lymphoblastic leukaemia (B-lymphoblast cells). It is critical to distinguish malignant cells from healthy cells when diagnosing acute lymphoblastic leukaemia (B-lymphoid precursors). Cancer cells, on the other hand, are nearly indistinguishable from healthy cells under a microscope. Blood counts are often used to diagnose various types of blood cancer [3]. Counting can be done manually by a trained individual (98% accuracy) or automatically by a machine (95% accuracy).

Automatic counting is more precise, although it is still not perfect. Both methods have pros and cons. AI-driven technologies have lately been used to diagnose a variety of ailments. Deep learning is challenging to detect and classify white blood cells in microscope images without using pre-processing or traditional segmentation [4]. There is a shortage of data, data noise, and data from numerous sources among these. Shallow machine learning was commonly employed by academics to extract a features map from a digital image, which was subsequently fed into prediction algorithms. Acute lymphoblastic leukaemia detection is a sensitive problem because it impacts people's lives and health. We also discovered that the L1, L2, and L3 ALL subtypes have received comparatively little attention. These subtypes are famously difficult to detect and classify due to their increased intra- and inter-class diversity and similarity. An accurate diagnosis is critical since your subtype will assist your doctor decide on a treatment plan. Unlike acute lymphoblastic leukaemia and acute myeloid leukaemia, myeloma is a plasma cell cancer that suppresses antibody synthesis. As a result, the immune system takes a significant hit. Building automated systems that can improve cancer detection efficiency and help detect the disease at an earlier stage is critical because human visual examination and diagnosis is time-consuming, labour-intensive, and less accurate than automated techniques.

II. LITERATURE REVIEW

[5] Described a simple approach for automatically segmenting and detecting nucleated cells in blood smear images in order to identify Adult Acute Myeloid Leukaemia (AML). They employed the local binary pattern approach and a comparison of Hausdorff dimensions to categorise the snapshot. After taking 80 microscopic images of blood samples, the approach was demonstrated to be accurate 98.79 percent of the time. [6] A unique approach for recognising AML and its subtypes in blood micrographs was developed. Using segmentation, many characteristics were recovered from the nuclei of white blood cells in multi-cell smear images. To extract diseases and locations of interest, medical images must be segmented and classified. When segmenting an image, there are usually two stages involved. The first describes diseased tissue, while the second describes specific anatomical features or study regions. The next step in the analysis is to accurately segregate objects of "interest" from a noisy and convoluted image [7].

Because of their excellent accuracy and reliability, CNNs have recently acquired prominence as a classification tool [8]. [9] Described a method for diagnosing leukaemia using blood slides. K-means clustering, histogram equalisation, and the Zack technique were used to isolate WBCs. Using SVM analysis, the model was determined to be right 92.87% of the time. This study intends to construct an autonomous model capable of diagnosing leukaemia in blood micro pictures in order to speed up the process. See [10] quick Relevance Vector Machine (RVM) for rapid WBC image segmentation. Fast RVM is a time-saving alternative to traditional RVM. [11] Relies on a dataset made accessible by VIT Mumbai, which includes 200 photographs of patients diagnosed with leukaemia. The photos are pre-processed to reduce undesirable light and noise before being analysed for the presence of glaucoma using a support vector machine (SVM) classifier, which analyses colour and morphological information retrieved from the images.

[12] Individual cells were successfully separated using a gradient scale, morphological operation, thresholding, and division modification. The results of an analysis of the effects of around fifty images on the suggested method and its segmentation performance are presented in this paper. [13] To detect malaria parasites in photographs of Giemsa-stained blood cells, a colour histogram, flat texture, and gradient area granulomatis features, as well as a decision tree classifier, were developed. In [14], we identify a machine learning technique for automating the recognition and tally of three hematologic subgroups. The suggested method counts numerous blood cell types in a reliable manner. The authors of recommendation paper [15] propose that blood cell classification be automated using transfer learning. To achieve accurate characterization findings, we train the model with bone marrow imaging data using convolutional neural networks, a powerful dividing learning

approach. Deep learning and convolutional neural networks (CNNs). The method takes a bone marrow image as input, sorts the data, and then classifies the findings as "normal" or "subtype"[16].

Using a convolutional neural network model that is a hybrid of VGGNet and MobileNet. It is feasible to extract high-level features from input photographs by combining deep learning with other information expansion algorithms. [17] conducted an in-depth examination of how to identify white blood cells (WBCs) using microscopic images from haematological reports in order to diagnosis potentially lethal disorders like as leukaemia, blood cancer, and AIDS. Many challenges were identified, such as the complexity of the classification, the time required to do it, the unusual morphology of WBCs, and other concerns. [18] Provided a short explanation of the many distinct types of leukaemia and the multiple techniques used in the field of machine learning in their overview of leukaemia and image segmentation methodologies for detecting leukaemia in microscopic pictures.

III. PROPOSED SYSTEM

Obtaining high-quality photographs was a top objective during the system's development. Because artefacts look differently in small photos, image preparation is required. As a result, the initial component of this section will concentrate on the technicalities of acquiring and manipulating images for blood cell segmentation. The essay finishes with a summary of the feature extraction and classification method used.

A. Pre-processing

The goal of the image pre-processing stage was to clean up the blood image and make it ready for the segmentation stage [23]. The following procedures were used during the preprocessing stage.

1. Image Resizing: To reduce processing time, a 400×400 version of the original RGB image (2600×1950 pixels) was used.
2. Enhancement of Image: This really aids in the creation of an image with homogeneous illumination. To fulfil this mission, the image intensity adjustment option was used to boost the default intensity of 0.1 to 0.8. The Gaussian filter was then used to reduce the impacts of camera noise and erroneous pixel values.
3. Image De-Blurring: The image has been processed using a Wiener filter, which effectively de-blurs and eliminates such noises.

B. Image Segmentation

To partition the blood sample images, the clustering algorithm is applied. We used the Adaptive k-means clustering algorithm to do this. This approach can be used to process both colour and grey scale photos. The following are some of the issues with the k-means clustering on which it is based. First, estimating the K value or the number of clusters ahead of time may result in erroneous findings (i.e., clusters with no data). 2) It's slow and has bad time-scaling. 3) It may reveal poor local optimums, and fourth, if different divisions are employed initially, the outcome for the same image may be inconsistent.

C. Feature Extraction

The technique of defining a large amount of additional data into a feature set with reduced dimension is known as feature extraction. Correct feature selection becomes a vital stage since it enhances classifier performance significantly [24]. The need for very accurate detection of distinct types of blast cells drives the development of blood cell characteristics. Feature extraction converts input data into features. Because feature selection has a significant impact on classifier performance, it is an important process step.

1. Colour features: In this analysis, we mapped RGB values to HSV values. The standard deviation of their hue values has been calculated.

The procedure for measuring Hausdorff Dimensions using the box counting approach is presented as an algorithm in fig 1.

D. Training the Model

1. AE Based Feature Reduction: AE is a feed forward neural network (FNN) that serves as an efficient encoding strategy in unsupervised learning. AE can assist reduce the dimensionality of features by learning to represent them. In a real-world implementation of the suggested method, the nonlinear AE is utilized to lower the dimensionality of the NGist feature vector. Feature extraction from both the training and testing datasets is used to generate new features. At this point, some critical AE characteristics must be supplied [25].
2. Blood Cancer Classification Using RELM: The RELM network is a single-hidden-layer feed forward neural network (FNN) with an input layer and an output layer. Input layer weights will be generated at random, while



Figure 1. Algorithm of HD Dimension

output layer weights can be calculated mathematically. Let's pretend R is a training dataset matrix. The following equation describes the RELM with k hidden nodes and an activation function, $g(x)$:

$$Q_i = \sum_{j=1}^k \mu_{kj} (v_{kj} \cdot z_i + d_k) \quad (1)$$

Where $i=1, \dots, M$

where nodes $v_k = [v_{k1}, v_{k2}, \dots, v_{kj}]^T$ is a weight vector connecting each hidden node k to all input nodes, $v_k \cdot z_i$ is the inner product of v_k and z_i , d_k is the threshold of the k th hidden nodes, and $\mu_k = [\mu_{k1}, \mu_{k2}, \dots, \mu_{kM}]^T$

The i th hidden node's weight vector connecting it to the output nodes is denoted by k . It is possible to obtain a perfect approximation to the M data samples using

$$\sum_{l=1}^M \|q_i - s_l\| = 0 \quad (2)$$

Where s_i is an encoding of a target vector. The matching target vector for this encoding is defined as

$$s_i = [d_1, \dots, d_n]^T \quad (3)$$

In a short amount of time, it can rewrite Eq. (3) in the form of matrices by

$$S = G\alpha \quad (4)$$

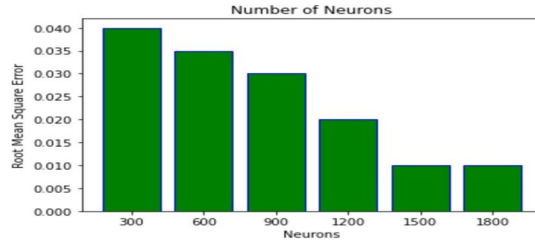


Figure 2. RMSE Comparison

TABLE I. COMPARISON OF THE MODEL

Models	ACCURACY	PRECISION	RECALL
ELM	92.39	89.48	86.54
RELM	95.83	92.74	89.86
AE-RELM	98.45	96.29	94.57

Where

$$G = [g(s_1, s_2, \dots, s_k, c_1, c_2, \dots, c_k, z_1, z_2, \dots, z_k), \dots, g(v_1 \cdot z_1 + d_1) \dots g(v_j \cdot z_1 + d_j) \dots g(v_1 \cdot z_M + d_1) \dots g(v_j \cdot z_M + d_j)] \quad (5)$$

Where G is the hidden layer's output matrix, and each column k represents the hidden node's output for a given set of inputs $z_1, 2, \dots, z_M$.

The parameters of hidden nodes with nonzero activation functions, $(.)$, can be fixed randomly and the output weights can be determined on any input data sets.

For testing phase, the output target matrix can be generated by

$$Z = \tilde{G}\tilde{\mu} \quad (6)$$

where $\tilde{G}$ is a testing dataset weight matrix computed and $\tilde{\mu}$ is a testing dataset hidden layer output matrix

$$\tilde{\mu} = (GS + \beta K) - 1GS \quad (7)$$

where k is an identity matrix, GS is a transpose matrix of G , S is a target output matrix, and μ is a regularize parameter to trade-of between the regularized degree and the estimated error.

IV. RESULT AND DISCUSSION

Images were input and pre-processed before being edited with image modification to increase contrast and filtered with Gaussian and Wiener filters, respectively. The image's uneven and fuzzy contrast. To ensure that the image's contrast was maintained throughout, a contrast enhancement image processing approach was applied, and a Gaussian and Wiener filter was used to eliminate the stepper and pepper noise and blurriness that occurred during the image capturing process. K-Means is used to segment images before extracting GLCM, color, texture, and Hausdroff as features for model training.

The recognition performance of a RELM network is particularly sensitive to the number of neurons in its hidden layer. Fig 2 depicts the recognition results for several neuron combinations used to compute activation function and select best architecture. Because increasing the number of neurons in the network has no significant effect on the Root Mean Square Error, we limit the network size in this article at 1800 (RMSE). Table 1 shows the comparison of all models. Our suggested AE-RELM based Blood cancer detection technique obtains higher accuracy than ELM (92.39%) and RELM (95.83%). Several studies have demonstrated that our proposed AE-RELM based Blood cancer detection method can succeed where others have failed.

V. CONCLUSION

The majority of the human body is made up of blood. Red blood cells, white blood cells, platelets, and other artefacts make up blood. Blood samples are used in the majority of diagnostic testing for illnesses such as malaria, anaemia, and leukaemia. Blood cells are very important in certain disorders. Because of parasite attack or other anomalies in their sizes, shapes, and colours, blood cells are the primary focus of these illnesses. Following the receipt of the input image, the model is trained using Adaptive K-Means clustering, colour features, texture features, GLCM, and Hausdroff, followed by residual entropy learning and radial basis function models.

Extended AE and RELM are offered as a potential approach for diagnosing blood cancer. To extract the blood cancer features, an alternative to the Gist descriptor (NGist) was used, and AE was applied to the issue of high dimensionality using NGist capabilities.

REFERENCES

- [1] E. Barbé, M. de Boer, and D. de Jong, "A practical cytological approach to the diagnosis of breast-implant associated anaplastic large cell lymphoma," *Cytopathology*, vol. 30, no. 4, pp. 363–369, 2019, doi: 10.1111/cyt.12678.
- [2] R. Willemze *et al.*, "The 2018 update of the WHO-EORTC classification for primary cutaneous lymphomas," *Blood*, vol. 133, no. 16, pp. 1703–1714, 2019, doi: 10.1182/blood-2018-11-881268.
- [3] A. B. Kul'chyn's'kyi, V. M. Kyjenko, W. Zukow, and I. L. Popovych, "Causal neuro-immune relationships at patients with chronic pyelonephritis and cholecystitis. Correlations between parameters EEG, HRV and white blood cell count," *Open Med.*, vol. 12, no. 1, pp. 201–213, 2017, doi: 10.1515/med-2017-0030.
- [4] H. Dathar and A. M. Abdulazeez, "A Modified Convolutional Neural Networks Model for Medical Image Segmentation," *TEST Eng. Manag.*, vol. 83, no. July, pp. 16798–16808, 2020.
- [5] S. Agaian, M. Madhukar, and A. T. Chronopoulos, "Automated screening system for acute myelogenous leukemia detection in blood microscopic images," *IEEE Syst. J.*, vol. 8, no. 3, pp. 995– 1004, 2014, doi: 10.1109/JSYST.2014.2308452.
- [6] F. Kazemi, T. Najafabadi, and B. Araabi, "Automatic recognition of acute myelogenous leukemia in blood microscopic images using K-means clustering and support vector machine," *J. Med. Signals Sens.*, vol. 6, no. 3, pp. 183– 193, 2016, doi: 10.4103/2228-7477.186885.
- [7] A. Husham, M. Hazim Alkawaz, T. Saba, A. Rehman, and J. Saleh Alghamdi, "Automated nuclei segmentation of malignants using level sets," *Microsc. Res. Tech.*, vol. 79, no. 10, pp. 993–997, 2016, doi: 10.1002/jemt.22733.
- [8] S. Iqbal, M. U. Ghani, T. Saba, and A. Rehman, "Brain tumor segmentation in multi-spectral MRI using convolutional neural networks (CNN)," *Microsc. Res. Tech.*, vol. 81, no. 4, pp. 419–427, 2018, doi: 10.1002/jemt.22994.
- [9] N. Patel and A. Mishra, "Automated Leukaemia Detection Using Microscopic Images," *Procedia Comput. Sci.*, vol. 58, pp. 635–642, 2015, doi: 10.1016/j.procs.2015.08.082.

- [10] S. Ravikumar, "Image segmentation and classification of white blood cells with the extreme learning machine and the fast relevance vector machine," *Artif. Cells, Nanomedicine Biotechnol.*, vol. 44, no. 3, pp. 985–989, 2016, doi: 10.3109/21691401.2015.1008506.
- [11] T. Dharani and S. Hariprasath, "Diagnosis of Leukemia and its types Using Digital Image Processing Techniques," *Proc. 3rd Int. Conf. Commun. Electron. Syst. ICCES 2018*, no. Icces, pp. 275–279, 2018, doi: 10.1109/CESYS.2018.8724075.
- [12] L. H. Nee, M. Y. Mashor, and R. Hassan, "White blood cell segmentation for acute leukemia bone marrow images," *J. Med. Imaging Heal. Informatics*, vol. 2, no. 3, pp. 278–284, 2012, doi: 10.1166/jmihi.2012.1099.
- [13] L. Malihi, K. Ansari-Asl, and A. Behbahani, "Malaria parasite detection in giemsa-stained blood cell images," *Iran. Conf. Mach. Vis. Image Process. MVIP*, no. September, pp. 360–365, 2013, doi: 10.1109/IranianMVIP.2013.6780011.
- [14] M. M. Alam and M. T. Islam, "Machine learning approach of automatic identification and counting of blood cells," *Healthc. Technol. Lett.*, vol. 6, no. 4, pp. 103–108, 2019, doi: 10.1049/htl.2018.5098.
- [15] A. Acevedo, S. Alf  rez, A. Merino, L. Puigv  , and J. Rodellar, "Recognition of peripheral blood cell images using convolutional neural networks," *Comput. Methods Programs Biomed.*, vol. 180, p. 105020, 2019, doi: 10.1016/j.cmpb.2019.105020.
- [16] Rehman, N. Abbas, T. Saba, S. I. ur Rahman, Z. Mehmood, and H. Kolivand, "Classification of acute lymphoblastic leukemia using deep learning," *Microsc. Res. Tech.*, vol. 81, no. 11, pp. 1310–1317, 2018, doi: 10.1002/jemt.23139.
- [17] K. N. Khamael AL-Dulaimi, Jasmine Banks, Vinod Chandran, Inmaculada Tomeo-Reyes, "Classification of White Blood Cell Types from Microscope Images: Techniques and Challenges," *Microsc. Sci. Last Approaches Educ. Programs Appl. Res.*, vol. 8, pp. 17–25, 2018.
- [18] D. Umamaheswari and S. Geetha, *Review on image segmentation techniques incorporated with machine learning in the scrutinization of leukemic microscopic stained blood smear images*, vol. 30. Springer International Publishing, 2019, doi: 10.1007/978-3-030-00665-5_163.
- [19] M. N. Sharath, T. M. Rajesh and M. Patil, "Analysis of Secure Multimedia Communication in Cloud Computing," 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICT), Kannur, India, 2019, pp. 136–144, doi:10.1109/ICICT46008.2019.8993352.
- [20] Sharath M N, Dr. Rajesh T M, Dr. Mallanagouda Patil. "Novel video Steganography approach with the combination of knight tour and 7th bit model of embedding data". EEO. 2021; 20(4): 657-669. doi:10.17051/ilkonline.2021.04.70
- [21] Sharath, M.N., Rajesh, T.M. & Patil, M." Design of optimal metaheuristics based pixel selection with homomorphic encryption technique for video steganography".Int.j.inf.tecnol.14, 2265–2274(2022). <https://doi.org/10.1007/s41870-022-01005-9>.
- [22] Sharath M N, Dr. Rajesh T M, Dr. Mallanagouda Patil. "A Novel Encryption with Bacterial Foraging Optimization Algorithm based Pixel Selection Scheme for Video Steganography". Multimedia Tools and Applications. <https://doi.org/10.1007/s11042-023-14420-2>
- [23] K. Dese Gebremeskel, T. Chung Kwa, K. Hakkins Raj, G. Ayana Zewdie, T. Yemaneh Shenkute, and W. Addisu Maleko, "Automatic Early Detection and Classification of Leukemia from Microscopic Blood Image," *Abyssinia J. Sci. Technol.*, vol. 3, no. 1, pp. 1–10, 2021, [Online]. Available: <https://journals.wu.edu.et/index.php/ajec/article/view/160>
- [24] A. Krishnan and K. Sreekumar, "A Survey on Image Segmentation and Feature Extraction Methods for Acute Myelogenous Leukemia Detection in Blood Microscopic Images," vol. 5, no. 6, pp. 7877–7879, 2014.
- [25] A. Gumaei, R. Sammouda, A. M. S. Al-Salman, and A. Alsanad, "An Improved Multispectral Palmprint Recognition System Using Autoencoder with Regularized Extreme Learning Machine," *Comput. Intell. Neurosci.*, vol. 2018, 2018, doi: 10.1155/2018/8041609.