**Automated Examination of WBC – Leukaemia using Machine Learning and Deep Learning Algorithm**

Bhavika Vaghela1, Assistant Professor, Parul Institute of Computer Application, Parul University, bhvika.vaghela@paruluniversity.ac.in

Dr. Priya Swaminarayan2, Dean – Faculty of IT & Computer Science, Parul University, [priya.swaminarayan@paruluniversity.ac.in](mailto:priya.swaminarayan@paruluniversity.ac.in)

Aman Kumar Singh3, Student , Parul Institute of Computer Application, Parul University,

[amansingh844123@gmail.com](mailto:amansingh844123@gmail.com)

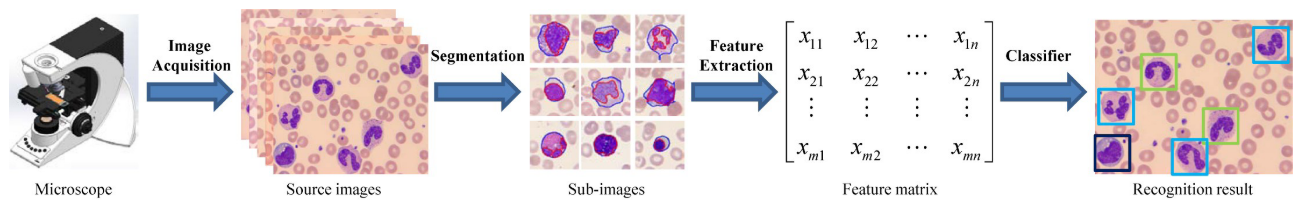
***Abstract*—** In Early Days Diagnosis for any Medical disease is manually so it is Time Consuming task and for getting accurate result there is a need of experts. Over all the medical disease 200 types of cancers are there here in this chapter we are going to discuss about blood cancer. Blood cancer affected the way blood cell behave and work to protect human body. There are three types of blood cell like White Blood cell, Red Blood Cell, Platelets and three types of blood cancer are Leukaemia, Lymphoma, and Myeloma so here in this chapter we are going to discuss depth analysis of Traditional Machine Learning and Deep Learning algorithm to examination of WBC – Leukaemia cancer for Getting accurate diagnosis of disease correctly classification of leukocytes and its sub classes are required. As Machine Learning and Deep Learning plays vital role to Analysis the Medical Image domain like Magnetic Resonance Image, CT-Image, Ultrasound and X-Ray. The main impact of proposed review is to find most suitable Machine Learning and Deep Learning Technique to find White blood cell counts – Leukaemia. These ML and DL algorithms are mostly applied on blood smear images analysis which provide valuable information to the health care specialist for early diagnosis, classification and examination of White Blood Cell – Leukaemia. By Correct analysis of White Blood Cell Medical specialist can diagnose various hematic disease like AIDS and Leukaemia Cancer. In this chapter we are going to discuss ML and DL algorithms like Support Vector Machine (SVM), K- Nearest Neighbor (KNN), Decision Tree (DT), Naive Bayesian, Decision Tree, ANN and Convolutional Neural Networks (CNN) for classification of Leukaemia. Leukaemia makes lot white blood cell which reduce the immunity to fight with infection. The above listed ML and DL techniques normally developed using WEKA, MATLAB or Python. 3, 00,000 new cases of Leukaemia every year diagnosis in all over the world. In United State every 3 minutes one person diagnosis with the Leukaemia. So, with this study our aim to derived future research direction for Analysis of Medical Images.

***Keywords*—** SVM, KNN, DT, ANN, CNN

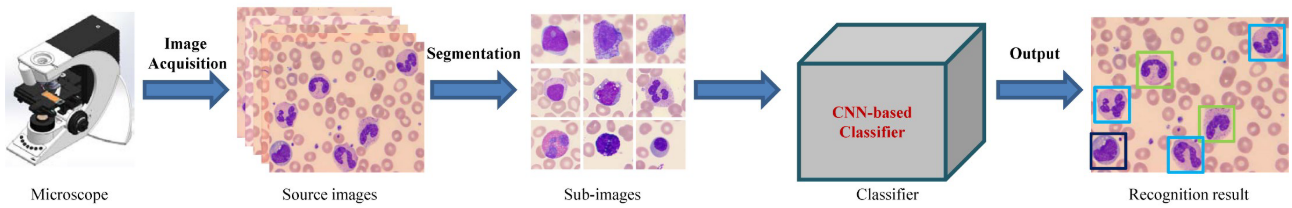
1. **INTRODUCTION**

Leukemia is the term for an increase of hematogenic cells. Hematology tests have been more important in the clinical diagnosis of several deadly illnesses in recent years. The differential count of white blood cells is one of the primary markers in a blood routine test that provides highly helpful information about the health problems of patients [1]⁠. For the malignancy of changing cells and offspring to generate leukemic cells clone, many genetic techniques have been used. In acute leukaemia, the bone marrow contains more than 20% blasts. A rise or fall in the number of WBCs in the peripheral blood signals a problem. Morphological differences, such as shape, size, and color, also aid in diagnosis. As a result, WBC detection and categorization are critical in peripheral blood smear analysis. Stack autoencoders have been used for medical picture segmentation and classification in a number of research. Auto-encoder is a type of deep learning machine learning that works similarly to a neural network. It is a learning approach that converts inputs to outputs with the least amount of error feasible. A feature set or photos can be used as input to the autoencoder. Encoders, decoders, and a loss function are all part of it. Encoder is a neural network that creates output y from given input x based on the size of the hidden layer. Decoder is a type of neural network that generates output x from input y. The decoder receives the encoder's output as input. It makes use of output functions like sigmoid, SoftMax, and so on [2]. Stack autoencoders have been used for medical picture segmentation and classification in a number of research. Auto-encoder is a type of deep learning machine learning that works similarly to a neural network. It is a learning approach that converts inputs to outputs with the least amount of error feasible. A feature set or photos can be used as input to the autoencoder. Encoders, decoders, and a loss function are all part of it. Encoder is a neural network that creates output y from given input x based on the size of the hidden layer. Decoder is a type of neural network that generates output x from input y. The decoder receives the encoder's output as input. It makes use of output functions like sigmoid, SoftMax, and so on [3].

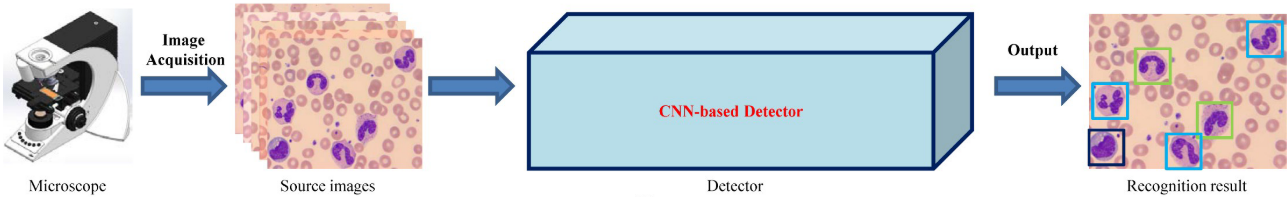
Researchers have always attempted to automate the morphologic leukocyte differential count. They employed shallow machine learning models, which rely on information derived in a similar fashion to morphologists' study. These efforts try to quantify key information derived from digital photos for use as input to prediction systems (see Fig 1). Artificial Neural Networks (ANNs), Supported Vector Machine (SVM), Naive Bayes Classifier, Linear Discriminate Analysis (LDA), and Multi-Layer Perceptron are some of the most prominent shallow machine learning algorithms used in leukocyte categorization (MLP). Many research focus on image pre-processing, object segmentation, and feature extraction and selection, which are the preconditions of classification models, in order to achieve good classification performance. Traditional leukocyte recognition methods generally achieve excellent classification accuracy under highly controlled settings, such as in, or by utilising tiny datasets, such as in.



**Fig — (1**)



**Fig — (2)**



**Fig — (3)**

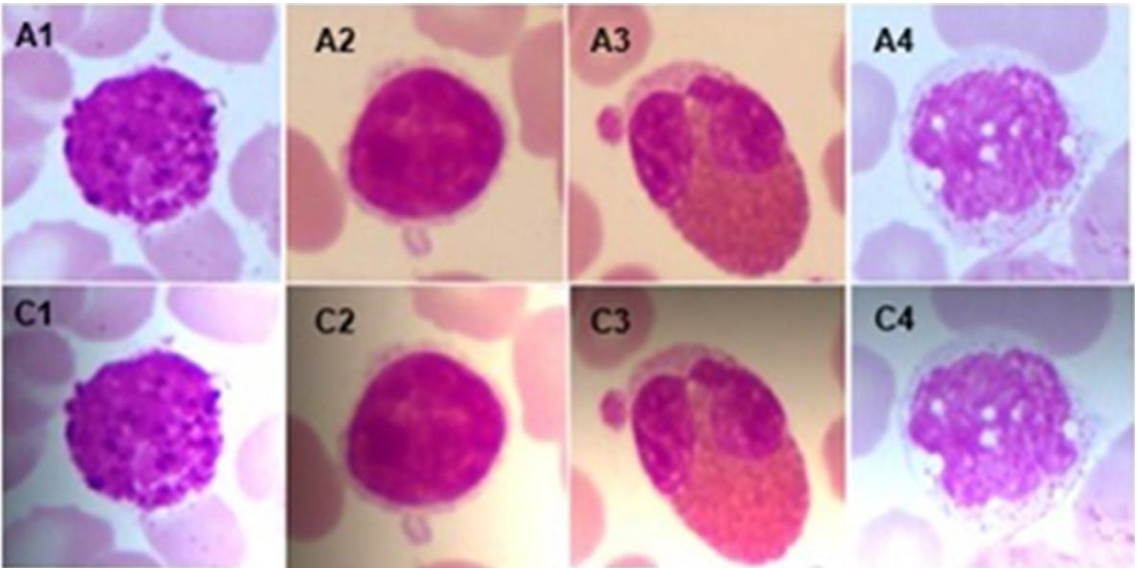
**In above figure** illustrates the identification pipelines for peripheral leukocytes. (A) treat leukocyte recognition as traditional feature engineering: manual segmentation, feature extraction, and selection, followed by a classifier based on the feature matrix; (B) treat leukocyte recognition as object classification: manually or automatically extract patches containing leukocyte candidates from the original image, then feed these patches into a CNN-based deep learning classifier to output leukocyte types; (C) treat leukocyte recognition as object decomposition: manually extract patches containing leukocyte candidates from the original image [3].

1. **MATERIALS AND METHODS**

This section explains how data was collected, features were extracted, and categorization was done. The current work extracts features such as form, color, and texture. The segmentation and retrieved characteristics are described in depth. WBCs were classified into five classes and abnormalities, as indicated in above figure. We tested with classification using three different methods: NN, auto encoders, and CNN.

1. **Data Collection:**

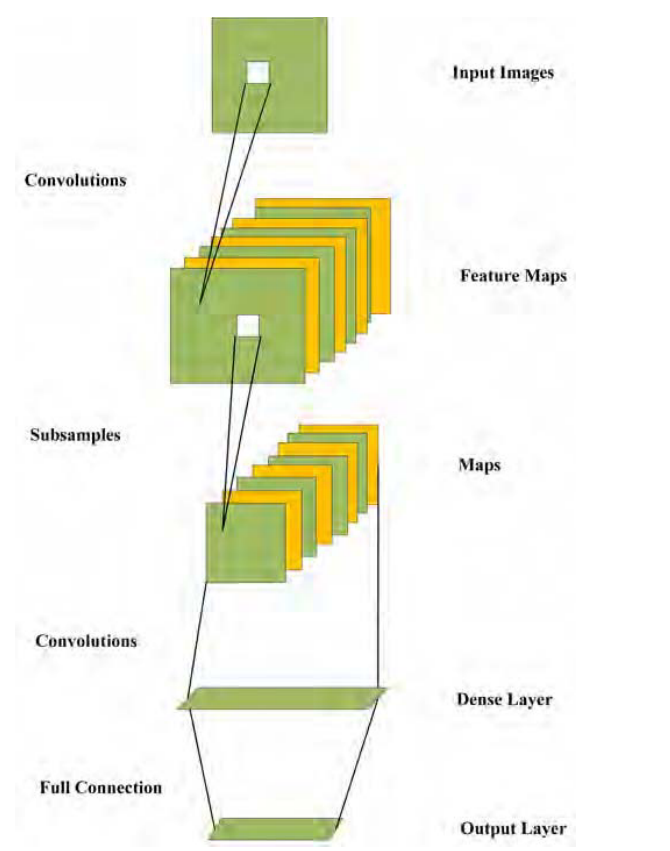
Figure 2 shows a couple of the photos that resulted. Photographs A1–A4 in the figure are the original images, whereas B1–B4 are the resultant images after using the 'color-balancing' procedure. Between the original and enhanced photos, colour difference may be observed. Despite the colour change, the leukocytes retain all of their original information, including nuclei texture and colour.



**Fig 3:** Results of non-uniform brightness variation A1–A4: original images C1–C4: augmented images [2].

1. **Convolution Neural Network:**

The Convolution Neural Network (CNN) is a fundamental network structure for performing machine learning tasks. CNN has a good capacity to adapt in computer vision tasks since it is extremely invariant to tilting, translation, and scaling. CNN can extract numerous local characteristics from an image using various convolution kernels. These characteristics are then loaded into a standard neural network, which produces an accurate output. The output result reflects the likelihood that a photograph falls into one or more specific categories. Figure 5 depicts a typical CNN architecture.



**Fig 5:** CNN’s usual Architecture [1].

1. **Image acquisition:**

Microscopic photographs of bone marrow stained with Leishman stain are observed optically and shot using a Euromax digital camera microscope under normal lighting conditions and oil immersion with the 100 lens. Every grab image is kept in three colours: red, green, and blue in its original form (RGB) [4].

1. **Extraction of features using conventional neural networks:**

The primary concept behind transfer learning with very deep CNNs is to take a pretrained deep network that has previously been fitted to a large dataset like ImageNet and modify it to handle a new image classification problem22. The network has a foundation of characteristics that may be utilised to focus on a specific picture type to perform a classification assignment since it learnt important image features from a general training dataset. We employed the VGGNet CNN architecture, which is a popular and stable CNN design. The number of channels (width of the conv layers) is quite minimal, ranging from 64 in the initial layer to 512 after each max-pooling operation, growing by a factor of two. The input layer's size is set at 224 x 224 pixels. A stride is added to each picture as it passes through a stack of conv layers to retain spatial resolution. Over a specified window, 5 max-pooling layers conduct pooling, with stride following some but not all conv layers [5].

1. **EXPERIMENTAL RESULT**

We compare the prediction performance of our CNN model to that of numerous classic models to assess our system's generalization capacity. The train pictures are used to train these models, and the test set is used to evaluate them. The performance of several models in the experiment is shown in Table I. In the classification challenge, our CNN model attained an accuracy of roughly 88.5 percent, which is greater than any other traditional technique.

Because the time to process all of the iterations is recorded during the trials, the accuracy is shown on the basis of iteration. All experiments, even those with a low learning rate and fewer epochs, had a high accuracy rate. As a result, it is evident that as the learning rate and epochs grow, so will the accuracy [4].

We compare ConVNet against the dominating technique of SVM-GA and two standard machine learning methods, namely MLP and random forest, to assess the performance of our deep learning approach. The accuracy results from these methodologies are shown in the table, which includes 10 test sets and the average with standard deviation over the ten performance estimations. The two classic techniques cannot reach an average accuracy of more than 80%, but ConVNet and SVM-GA achieve an average accuracy of more than 80% and generate comparable results with a tiny margin of difference. The majority of the results returned by both ConVNet and SVM-GA from the 10 set runs are above 80%, with the number ranging between 78 and 86 percent [6].

We also compare the classification results of NN with those of deep learning techniques such as stacked autoencoders and CNN for WBC classification. For comparison, the performance of SVM and ensemble classifiers was assessed using the retrieved features. The ensemble classifier and the multi-class SVM with quadratic kernel had average accuracies of 92.5 percent and 97.2 percent, respectively. NN, on the other hand, was determined to be superior, with an average accuracy of 99.8%. The NN classifier's performance was further assessed using 5-fold cross validation. With an overall accuracy of 99.6% and a sensitivity of 98.9%, the results were impressive. With 80 percent of the data for training and the remaining 20% for testing, the extracted features were utilised to train the autoencoder. With an average accuracy of 96.72 percent, we were able to achieve our goal. For 100 iterations, the average accuracy was calculated. We investigated creating a CNN from scratch, as well as using a transfer learning strategy to classify WBCs. The 'training-set,' which included 2697 cropped photos, was used to create and train the CNN from the ground up, while the 'test-set,' which had 816 cropped images, was used to test the designed CNN [2], [4], [7].

Segmented neutrophils were misclassified as band neutrophils, lymphocytes as variant lymphocytes, band neutrophils as meta-myelocytes, meta-myelocytes as myelocytes, promyelocytes as myelocytes, and big platelets as thrombocyte aggregations, according to the DLS. We used t-distributed Stochastic Neighbor Embedding to evaluate the intrinsic features learnt by the DLS to deconstruct such misclassifications in the confusion matrixes (t-SNE). The blasts are surrounded by three types of cells: granulocytes, lymphocytes, and monocytes. From the most developed segmented neutrophils (top) to the most preterm promyelocytes, granulocytes are dispersed to the left of the blasts (bottom). Lymphocytes, on the other hand, are seen to the right of the blasts, and range in age from premature mutant lymphocytes (top) to mature lymphocytes (bottom) (bottom). Eosinophils, basophils, and monocytes are all located in different parts of the body [8].

1. **CONCLUSION**

This study showed a full architecture for the categorization of ALL that was built on deep learning techniques and obtained 97.78 percent accuracy with a short processing time. To train the model, the system uses convolution layers, max-pooling layers, and a fully connected layer, SoftMax, and classification layer. The suggested method takes the bone marrow picture as input, does segmentation, and classifies the marrow as normal if it is not impacted or as subtype L1, L2, and L3 if it is. The segmentation approach, which has never been used previously, is a new contribution of this work. The researchers have not before used automated methods to segment the nucleus and cytoplasm of entire cells. While segmentation is critical for effective categorization of L2 and L3 blasts based on their morphology, the application of deep learning approaches for ALL classification is considered innovative in this study. We offer a deep learning method for recognizing normal lymphocytes and ALL subtypes according to WHO categorization. We employ a CNN called Con- VNet, which takes raw pictures and uses a series of layered architecture to automatically uncover valuable characteristics. The performance of our deep learning model is compared to two traditional machine learning approaches, MLP and random forest, as well as a popular SVM classifier, SVM-GA. We also used both a classical image processing technique and a deep learning strategy to classify WBCs. Both approaches worked equally well, with a 99 percent overall accuracy and sensitivity. The accuracy of classification in a typical image processing technique is dependent on the precision of segmentation and feature extraction. This is no longer an issue because to the use of deep learning techniques. It learns the feature on its own, regardless of image alterations, but it requires a big amount of labelled data and enough infrastructure. Because of the availability of data and the reduced amount of cell pictures required by the network, CNN may be utilised to classify WBCs.

**ACKNOWLEDGMENT**

Here we have to write acknowledgment of this research paper.

**REFERENCES**

[1] W. Yu *et al.*, “Automatic classification of leukocytes using deep neural network,” *Proc. Int. Conf. ASIC*, vol. 2017-Octob, pp. 1041–1044, 2017, doi: 10.1109/ASICON.2017.8252657.

[2] R. B. Hegde, K. Prasad, H. Hebbar, and B. M. K. Singh, “Comparison of traditional image processing and deep learning approaches for classification of white blood cells in peripheral blood smear images,” *Biocybern. Biomed. Eng.*, vol. 39, no. 2, pp. 382–392, 2019, doi: 10.1016/j.bbe.2019.01.005.

[3] Q. Wang, S. Bi, M. Sun, Y. Wang, D. Wang, and S. Yang, “Deep learning approach to peripheral leukocyte recognition,” *PLoS One*, vol. 14, no. 6, pp. 1–18, 2018, doi: 10.1371/journal.pone.0218808.

[4] A. 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.

[5] T. S. Ahmed, K. Philip, and A. E. Ahmed, “Efficient Classification of White Blood Cell Leukemia with Improved Swarm Optimization of Deep Features,” pp. 1–11, 2020, doi: 10.1038/s41598-020-59215-9.

[6] T. Pansombut, S. Wikaisuksakul, K. Khongkraphan, and A. Phon-On, “Convolutional neural networks for recognition of lymphoblast cell images,” *Comput. Intell. Neurosci.*, vol. 2019, 2019, doi: 10.1155/2019/7519603.

[7] S. I. Lee *et al.*, “A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia,” *Nat. Commun.*, vol. 9, no. 1, 2018, doi: 10.1038/s41467-017-02465-5.

[8] K. Kimura *et al.*, “A novel automated image analysis system using deep convolutional neural networks can assist to differentiate MDS and AA,” *Sci. Rep.*, vol. 9, no. 1, pp. 1–9, 2019, doi: 10.1038/s41598-019-49942-z.