Blood Cells Classified From Blood Smear Images Into White Blood Cells And Red Blood Cells Using Machine Learning Methods

Prof. Kirti Borhade[1], Saurabh Wakase[2], Shiv Yandralwar[3], Pratham Zambre[4],

Department of Computer Engineering[1,2,3,4],

PCET's Nutan Maharashtra Institute of Engineering and Technology[1,2,3,4],

Pune, India[1,2,3,4]

Abstract - The blood cell classification from smear images of peripheral also known as PBS is a crucial step in diagnosing blood-related illnesses such as anemia, leukemia, infection, polycythemia and malignancy. Hematologists regularly utilize a device which zooms to microscopic level to count, shape, and distribute the cells before making a judgment in blood cell-based analysis. Hematology analyzers and flow cytometry give an accurate and precise CBC that identifies and gives the abnormalities in given smear slides of blood. In multiple hospitals, the techniques that are used are costly, least effective in terms of time and are hectic as they are manual. As a result, a reliable, affordable, and automatic method for identifying different sickness through given PBS image is required. The new proposed model automatically does the examination of the provided data, also does in a faster manner. Therefore, in the studied research we have properly and accurately done the classification of images in such a way that the different cells are classified as the WBCs and RBCs. Feature extraction is done to classify the images. These extracted texture features are subsequently inputted into various classifiers, including artificial neural networks (ANN), machines that are SVM and many other model which are in ML and DL. After comparing the performance metrics, it is determined that the logistic regression method is the most appropriate for the task at hand.

Index Terms— Images of microscopic blood smears, ML, RBC, Feature identification and export, WBC, CNN, Neural Networks.

I. INTRODUCTION

In medical diagnostics, one of the important tasks are segregation of WBCs and RBCs from the given smear images of blood. Medical personnel may find it easier to diagnose a variety of illnesses and ailments by quickly and accurately identifying blood cell types with the help of automated picture analysis via machine learning techniques.

This gives a general idea of the significance of the assignment and the approaches taken to ensure correct classification. This research project uses digital images from blood smear samples to build and examine ML based techniques so that the cells of blood cells are accurately classified into erythrocytes i.e. RBCs and leukocytes i.e. WBCs. Blood cells analysis is crucial in medical diagnostics, contributing significantly to the identification and comprehension of various hematologic conditions and ailments. Among the crucial components of blood, leukocytes and erythrocytes serve distinct physiological functions and exhibit characteristic morphological features. The classification and differentiation of these cells from microscopic blood smear images are essential for accurate diagnosis and monitoring of blood-related ailments. Manual examination of blood smears by hematologists or trained technicians is time-consuming and subjective, relying on visual inspection to differentiate between various cell types. The integration of machine learning techniques in analyzing these images offers a promising avenue for automating this process, potentially improving accuracy, consistency, and efficiency in blood cell classification.

II. LITERATURE SURVEY

Considerable research efforts have been devoted to applying machine learning techniques to blood smear classification, with numerous studies highlighting the effectiveness of convolutional neural networks (CNNs) in this specific area .

Transfer learning, a well-established technique in machine learning, has found resonance in the field of whale detection as well. This technique involves leveraging knowledge or features acquired from one related task to enhance the performance of a distinct yet interconnected task. The successful use of transfer learning has been showcased in various domains, one of which is blood smear image classification, highlighting its potential for improving diagnostic accuracy and efficiency.

In order to evaluate frequent fitness conditions in scientific diagnosis, for the necessary assessment the whole part of a blood cell is mandotary. Traditionally hemocytometer was used for the counting of the blood cells, which was done manually and a variety of solvent (chemical) and laboratory substances. This process is expensive as it takes exorbitant

and is highly tedious to experiment with. This paper proposes a machine learning system that can learn how to count blood cells automatically using a computer. Compared to other methods now in use, by using ML and DL achieving the highest level of accuracy of the cells from blood and their count is possible. The classification of images is done using CNN [1]. The models that have been learned are applied generally to the building and examining of various datasets. For practical purposes, this computer-assisted blood cell detection technology is generally more beneficial.

Classifying blood cells is an important part of medical diagnostics. In this article, we present a DL-oriented method for classifying cells of blood and assess the effectiveness of a multi-layer neural network model constructed using a combination of recurrent and convolutional neural networks (RNN) for classifying different kinds of white blood cells. The suggested approach produces better outcomes by utilizing the advantages of both CNN and RNN [2].

Consequently, there's an increasing demand for an automatic, cost-effective, and reliable approach to diagnose diseases from PBS images. Automated classification models have the potential to streamline hematological procedures, expedite diagnosis, and improve accuracy in evaluations. In this study, we introduce a brand-new, semi-automated method for sorting cells of blood into RBCs and WBCs [3]. Our goal is to improve diagnostic speed and availability by streamlining the cell identification process.

This work used DL and IP techniques to create a novel mixed attribution extraction method. There are two phases in the suggested method: 1) The CMYK-moment localization approach is utilized to collect the important area of choice and 2) A CNN-oriented attribute fusion method is utilized to extract features based on deep learning. A number of classification methods are applied to assess the importance of the features that were extracted. The main dataset which was chosen for attributes extraction was compared by an external dataset to evaluate recommended method of feature extraction. This technique has created a new option for enhancing WBC detection, which may result in an improved AML diagnosis. [4.]

In the immunological system of human beings, the WBCs guard the system of body from foreign dangers as well as fight infection. They consist of neutrophils, eosinophils, basophils, monocytes, and lymphocytes, each having its unique constitution and having each unique functionality to execute. Traditionally, a CBC exam which helps track an individual's health includes clinical labs process for measuring the various types of WBCs, to further improve Dense Net's performance, more sophisticated optimization methods such label smoothing, mixed-up augmentation, and normalization were applied [5].

Through the use of transfer learning, the CNN component obtains the already developed weighted attributes from the picture dataset. Additionally, we explored two distinct

scenarios. In the first, we used CNN to directly provide us with images. applied SVM on the subsequent instance. Next, we contrast the top category outcomes. To enhance the performance of the proposed system, more classifiers can be incorporated [6].

The hematologist will utilize the classification to identify the many types of white blood cells seen in the human body and identify the underlying causes of illnesses. There is a lot of study being conducted in this area right now. Given the immense potential of WBC classification, Convolution Neural Networks (CNN), a deep learning technique, will be developed and used to label the WBC images into its respective varying subtypes. This project presents the findings from a number of tests conducted using CNN on the BCCD dataset [7].

Because white blood cells, or leukocytes, are crucial for diagnosing diseases, their identification and classification is a popular topic. These days, the morphological examination of blood cells is performed manually by trained personnel, where it has several disadvantages such analysis slowness, non-standard accuracy, and operator reliance. Few publications take into account both WBC classification and detection simultaneously, despite the fact that numerous studies have examined these topics separately [8].

An image-cleaning method for calculating the amount regarding RBCs is proposed in this study. The K-medoids algorithm is utilized to collect WBCs through provided images and it is resistant to outside noise. To distinguish between red and white cells of blood, certain analysis methods are used. Hough's circular transformation method along with the tagging method is performed to calculate the estimation of RBCs [10].

This work provides an image analysis system based on a raspberry pi that uses connected component labelling and the Hue, Saturation, and Value (HSV) thresholding method to separate as well as calculate cells of blood through images at microscopic level. The numbers measured by the suggested system were compared statistically to each patient's real complete blood count test result. [9]

In this study, we suggest detecting ALL by exploiting textural descriptors taken from the nucleus picture. Texture is a result of changes in the nucleus's chromatin distribution brought on by the illness. For the automatic detection of ALL, we have employed two texture attributes: the GLCM and LBP [11].

III. PROPOSED METHODOLOGY

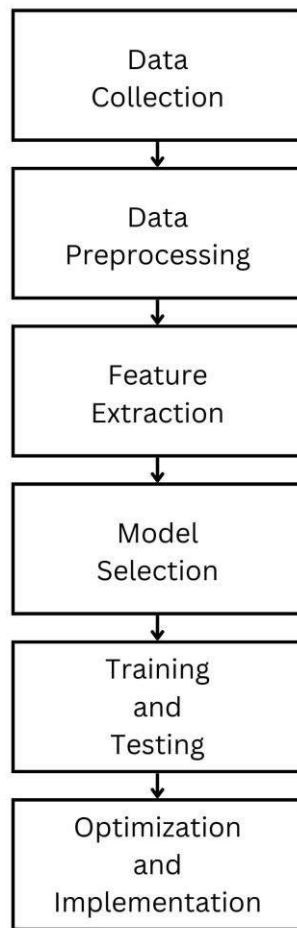The following part describes the suggested methodology represented in Figure 1.

Figure 1: A development structure of the recommended system

A. Image Collection:

Multiple images of stained PBS pictures were obtained from Kaggle and additional online sources. Few of the sample images are as follows.
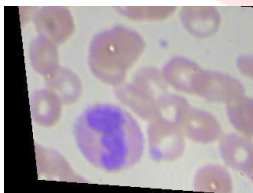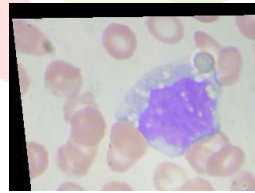

Figure 2: eosinophils

Figure 3: Monocyte

B. Pre-Compiling:

When the image is acquired through many methods, these images are not in clean and labelled format. These images include many artifacts, noise and dust which leads to improper training of the model. To avoid the image is clean that is processed This improves the quality of the image. The RBCs and WBCs which are gathered are individually cropped by the engineer using multiple methods. Few methods such as low pass and high pass filtering are examples to remove noise and artifacts. After the removal of artifacts, the image is then grayscaled, this brings the uniformity in the overall image as it

properly marks the features individually for better computation.

C. Feature Collection:

The images gathered for model training usually contains many features. ML uses extraction of attributes to understand the most important and required attributes for model development. This method is done for set. Feature collection is a process of collecting the important and required attributes for model development. This way we get amount of information held in each and every pixel present in the image. Using multiple methods like texture, features such as contrast, density and many more could be extracted from one selected area to another area of interest.

D. Classification:

The images are then later classified or differentiated into multiple groups. These images are classified by labelling them, Later the features of these images are then given as an input to the ML models and techniques like KNN, SVM and many others that are used for software development. For the objective function, a group of training samples are provided and new attributes are created named as a1 and a2 and so on until an. The new attributes are classified by giving rh best values by classification methods like GLCM, by making sure that no value is missed out. Classification is achieved by categorizing the samples by listing them from the roots of the tree to some of its leaves. Each node is a representative of an attribute, and each drop branch possibly represents the expected values.

Categorization of cells from blood using convolutional neural networks (CNNs) requires a few crucial steps:

Gathering and Preparing Data:

Gather a collection of blood smear pictures that show both red and white blood cells.

Before processing the photographs, resize them to a consistent size, normalize the pixel values, and maybe add more variation to the data by enriching it.

Model Architecture Selection:

Select a CNN design that works well for jobs involving image categorization. Typical options consist of:

- LeNet Alex Net VGG, such as VGG16 and VGG19,
- Res Net (such as ResNet101 and ResNet50) Beginning (for example, InceptionV3)

Training of Model:

From the main dataset, multiple sets named as training set, validation set and test sets are created. This is to ensure that the model developed is accurate enough. The CNN system is designed by applying algorithm methods like backward neural system or backtracking through the network. Also methods like Gradient tools are used to train the system. In ML systems, there are high chances of encountering overfitting. To avoid this, thorough inspection must be done on the validation data set and its performance and adjust the hyper parameters as per the need.

Assessment:
Examine the execution of the trained which used the testing dataset.
Assess model's performance using assessment metrics including recall, accuracy, precision, F1-score, and confusion matrix.

Enhancement:
To enhance the model's performance, adjust hyper parameters such batch size, dropout rate, and learning rate. To further improve training efficiency and avoid overfitting, take into account strategies like learning rate scheduling and early stopping.

IV. FUTURE SCOPE

While the paper explores many different methods algorithms of the machine learning, with the help of more and highly advanced techniques it is possible to extend this research. There are many places where affording the medical fees is not feasible, so this research can provide an alternative to the problem of regions. Machine learning techniques provided with proper training will definitely extend the functionality of many other researches. By extending this research, further analysis of many medical issues like disease forecasting, understanding the different diseases, etc. in possible.

V. CONCLUSION

Utilizing machine learning techniques for the automated categorization of cells from the blood smear images shows great potential in advancing medical diagnostic capabilities. The primary objective of this research is to develop an exact and effective software which can understand and differentiate between WBCs and RBCs through digital image analysis. The integration of machine learning algorithms for blood cell categorization from smear images of blood holds immense potential in revolutionizing medical diagnostics. The developed automated system showcases promising results in accurately identifying and categorizing blood cells, offering significant advancements in clinical diagnostics, pathology analysis and medical research.

VI. ACKOWNLEDGEMENT

We extend our sincere gratitude to all the people and researchers involved in the project, by providing their valuable contributions in enriching this research and helping in completing the research paper. Also, we extend our sincere gratitude and thank to Nutan Maharashtra Institute of Engineering and Technology for providing all the necessary support required in fulfilling this research.

We also deeply thank to all the researchers who have been previously and actively involved in the machine learning and deep learning field for their effortless and spontaneous contribution to the medical as well as the research field advancement.

Lastly, we will like to thank to all our family members and friends for supporting and believing in the project.

VII. REFERENCES

[1] A. HemaSri, M. D. Sreenidhi, V. V. K. Chaitanya, G. Vasanth, V. M. Mohan and T. Satish, "Detection of RBCs, WBCs, Platelets Count in Blood Sample by using Deep Learning," 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2023.

[2] J. Mitra, K. Vijayran, K. Verma and A. Goel, "Blood Cell Classification using Neural Network Models," 2023 2nd International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN), Villupuram, India, 2023.

[3] K. T. Navya, K. Prasad and B. M. K. Singh, "Classification of blood cells into white blood cells and red blood cells from blood smear images using machine learning techniques," 2021 2nd Global Conference for Advancement in Technology (GCAT), Bangalore, India, 2021.

[4] 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, 2022.

[5] Tamang T, Baral S, Paing MP. Classification of White Blood Cells: A Comprehensive Study Using Transfer Learning Based on Convolutional Neural Networks. Diagnostics (Basel). 2022 Nov 22.

[6] H. A. Muhamad, S. W. Kareem and A. S. Mohammed, "A Comparative Evaluation of Deep Learning Methods in Automated Classification of White Blood Cell Images," 2022 8th International Engineering Conference on Sustainable Technology and Development (IEC), Erbil, Iraq, 2022.

[7] Zhao J, Zhang M, Zhou Z, Chu J, Cao F. Automatic detection and classification of leukocytes using convolutional neural networks. Med Biol Eng Comput. 2017.

[8] William Franz Lamberti, Blood cell classification using interpretable shape features: A Comparative study of SVM models and CNN-Based approaches, Computer Methods and Programs in Biomedicine Update, Volume 1, 2021.

[9] D. Cruz et al., "Determination of blood components (WBCs, RBCs, and Platelets) count in microscopic images using image processing and analysis," 2017 IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), Manila, Philippines, 2017.

[10] Acharya, V., Kumar, P. Identification and red blood cell automated counting from blood smear images using computer-aided system. *Med Biol Eng Comput* **56**, 483–489 (2018).

[11] Fei Long, Jing-Jie Peng, Weitao Song, Xiaobo Xia, Jun Sang, BloodCaps: A capsule network based model for the multiclassification of human peripheral blood cells,

Computer Methods and Programs in Biomedicine, Volume 202, 2021, 105972, ISSN 0169-2607.