International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Machine Learning-Based Blood Cell Categorization in Smear Images

1st Mr. Rohit Sonawane
Dept. of Computer Science(RMDSSOE)
Savitribai Phule Pune University
Pune, India
rohitsanawaneg602@gmail.com

2nd Mrs. Manisha Darak
Dept. of Computer Science(RMDSSOE)
Savitribai Phule Pune University
Pune, India
manishadarak.rmdssoe@sinhgad.edu

3rd Mr. Abdul Shaikh
Dept. of Computer Science(RMDSSOE)
Savitribai Phule Pune University
Pune, India
shaikhabdullah32939@gmail.com

4th Mr. Sai Yandralwar
Dept. of Computer Science(RMDSSOE)
Savitribai Phule Pune University
Pune, India
saiyandralwar2019@gmail.com

Abstract –

One of the most important tasks in medical diagnostics is the categorization of blood cells from microscopic blood smear images, which helps identify different blood illnesses. This work investigates the use of machine learning approaches for automated image-based blood cell categorization into red blood cells (erythrocytes) and white blood cells (leukocytes). Preprocessing the blood smear pictures to improve contrast and eliminate noise is part of the suggested technique. The important properties, such shape, texture, and color information, are then extracted from the photos using feature extraction algorithms. These characteristics are used as inputs by a number of machine learning methods, such as decision trees, convolutional neural networks, and support vector machines (SVMs). This study advances the creation of automated systems. By advancing automated blood cell categorization systems, this study eventually helps medical personnel identify and track diseases more accurately and efficiently. It also has potential uses in pathology analysis, clinical diagnostics, and medical research.

Key Words*: *Blood Cell Categorization, Machine Learning , Smear Images, CNN , Image Processing.

1.INTRODUCTION

Accurate blood cell classification is essential for the diagnosis and treatment of many hematological illnesses in modern medicine. Hematologists' subjective and time-consuming manual interpretation of blood smear pictures has led to the need for automated treatments that use machine learning approaches. The goal of this research is to automatically classify blood cells in smear photographs using a machine learning-based method. This research intends to expedite the analysis of blood smear pictures by utilizing machine learning, offering faster and more objective findings than manual approaches. With the use of sophisticated algorithms and image processing methods, the system will be able to precisely detect and classify many blood cell kinds, such as platelets (thrombocytes), white blood cells (leukocytes), and red blood cells (erythrocytes). If this research is carried out successfully, it has the potential to completely transform the hematology sector by providing medical personnel with an invaluable instrument for accurate and timely blood cell analysis. The technology can increase productivity, decrease mistakes, and improve patient outcomes by automating this part of diagnostic testing.

A. Smear Image

In the context of classifying blood cells, a "smear image" is usually a digital image that is produced via microscopy and blood smear processing. To see the many kinds of blood cells present, a blood smear is a thin layer of blood put on a glass slide, dyed, and then studied under a microscope. Smear pictures for this project would be digital photos or scans of prepared blood smears that would show the fine features of platelets, white blood cells, and red blood cells. The machine learning algorithms entrusted with classifying the various blood cell types present are fed these photos as raw data. The efficacy of the automatic classification method depends on the quality of the smear pictures.

Proper contrast and low artifacts in high-resolution photos guarantee precise feature extraction and categorization by the machine learning model. Additionally, in order to train the model efficiently and improve its capacity to generalize to novel instances encountered in clinical practice, a varied collection of smear pictures reflecting different blood diseases and physiological situations could be required.

B. Convolutional Neural Network (CNN):

Convolutional Neural Networks (CNNs) are a popular option for blood cell identification in smear pictures because of their capacity to automatically learn hierarchical characteristics from raw pixel data. For this project, a CNN algorithm may operate as follows:

1.Input Layer: The smear image's raw pixel values are sent to the CNN's input layer. The dimensions of the input pictures determine the size of the input layer.

2. Convolutional Layers: The fundamental components of a CNN are its convolutional layers. Multiple filters, also known as kernels, make up each convolutional layer. These filters go over the input picture, multiplying its elements one at a time, and then combining the results to create feature maps. Different high-level and low-level characteristics, including edges, textures, and patterns, are captured by these feature maps.

3. Activation Function: To provide nonlinearity to the model and help it learn intricate patterns, a nonlinear activation function such as ReLU (Rectified Linear Unit) is typically added to the feature maps.

4.Pooling Layers: The feature maps are down sampled using pooling layers, which lower their spatial dimensions while keeping the most crucial data. A popular method of pooling data that preserves the maximum value inside each pooling zone is called "max pooling."

5.Flattening: The feature maps are flattened into a one-dimensional vector to be fed into the fully connected layers following a number of convolutional and pooling layers.

6. Fully linked Layers: Often referred to as thick layers, one or more fully linked layers are traversed by the flattened vector. These layers classify data by identifying and using the global patterns seen in the feature maps.

7.Output Layer: A single or more neurons, each representing a distinct kind of blood cell (such as red blood cells, white blood cells, or platelets), make up the CNN's output layer. The anticipated class of the input smear picture is indicated by the output neuron with the highest activation value.

8.Training: Using backpropagation and optimization methods like Adam or stochastic gradient descent (SGD), the CNN learns to categorize blood cells by modifying its weights and biases. To increase classification accuracy, the loss function is minimized after the model is trained on a labeled dataset of smear pictures.

9.Evaluation: To measure the trained CNN's effectiveness in classifying blood cells, an independent test dataset is used. To gauge the model's efficacy, metrics including accuracy, precision, recall, and F1-score are computed.

10.Deployment:After achieving acceptable performance, the CNN may be used in clinical settings to classify blood cells in smear pictures automatically, assisting medical practitioners with diagnosis treatment planning.

C. Prediction of diseases

CNN is a deep learning method that produces accurate embeddings characteristics and is used for Image identification. Like MTCNN, this approach generates face embeddings by training on a deep convolutional neural network (CNN) using a triplet loss function . This is how CNN operates:

1. To start, we make use of a sizable dataset of smear images. These pictures feature a range of expressions on lighting, and stances.

2. In the following phase, data Preprocessing is the act of cropping and aligning facial photos such that the face is the primary focus of the picture. Images' pixel values are normalized to provide consistency in the data for processing.

3. To extract features, CNN makes use of the inception module, which aids in identifying the hierarchical features. For every identified face in the picture, consistent embeddings are produced.

4. A key component of CNN model training is the triplet loss function. Three pictures are taken into consideration: an anchor image, a positive image that resembles an anchor image, and a negative image that differs from an anchor image. The primary goal is to increase the space between the anchor and the negative picture and decrease the distance between the anchor and the positive image. L (A, P, N) = max (0, ||f(A) - f(P)||^2 - ||f(A) - f(N)||^2 + margin) is the formulation of the loss function mentioned in. The embedding function is denoted by f(.).

5. In the following stage, the CNN model is trained using the input data and the triplet loss function mentioned before. Accurate face representations require a significant quantity of tagged data. It is possible to employ hyperparameters such learning rate, batch size, and triplet loss margin.

6. To make sure the CNN model can generalize Images that weren't accessible for training, it should be validated using different sets of data.

7. After the model has completed its training, additional Smear Images may be utilized to train the model using this face data.

8. The model's accuracy and precision may be verified using evaluation measures such as accuracy, precision, recall, and F1 score.

2. Body of Paper

An overview of earlier studies on automated blood cell segmentation and categorization is given in this paper. This study investigated a novel object identification technique that counts blood cells by using the YOLO algorithm. Using 364 annotated pictures, the approach obtained 96.1% accuracy for red blood cells (RBC) and 86.89% accuracy for white blood cells (WBC). WBC and RBC segments were produced with an average accuracy of 98.4% for WBCs and 95.3% for RBCs in 100 pictures using the iterative circle approach. Cruz et al. suggested a technique for Hue Saturation value (HSV) component RBC counting that uses watershed transform and blob analysis. Using ten blood samples, an average accuracy of 96% was attained. Wei and colleagues Developed a technique to distinguish between white blood cells and red blood cells using hue and saturation components. The work's author achieved an accuracy of 92.9% for 100 Wright-Giemsa stained images. The author suggested a method for blood cell counting by utilizing Nearest Neighbor and SVM techniques and manually cropping each cell. With 368 images from the ALL-IDB dataset, this method achieved an average accuracy of 98% for RBCs and 99.2% for WBCs.
Kindly rephrase this section using a new style. I'm grateful. In order to achieve 94.93% and 91.11% accuracy for RBC and WBC segmentation, respectively, we adopted a deep learning semantic segmentation technique. Furthermore, we demonstrated a Faster R-CNN model that achieved a 98% accuracy for the BCCD dataset in the classification of WBC and RBC variations. While many machine learning methods and deep neural networks are used to classify blood cells. We have created a simple and accurate method for identifying red blood cells (RBC) and white blood cells (WBC) in pictures that have numerous stains.

1. "Classification of blood cells into white blood cells and red blood cells from blood smear images using machine learning techniques"

The automated categorization model increases the accuracy of the assessment process, speeds up the diagnosis process, and improves hematological operations. Therefore, in this research, we classified and divided blood cells into Red Blood Cells (RBC) and White Blood Cells (WBC) using a semi-automated technique. Using the Gray Level Co-occurrence Matrix (GLCM), texture characteristics from a cell are retrieved and fed into classifiers such as the Naive Bayes classifier, K-nearest neighbors, decision tree, K-means clustering, random forest, logistic regression, ANN, and SVM. After comparing the performance metrics, it is discovered that, with a 97% accuracy rate, logistic regression is the most appropriate method for the job.

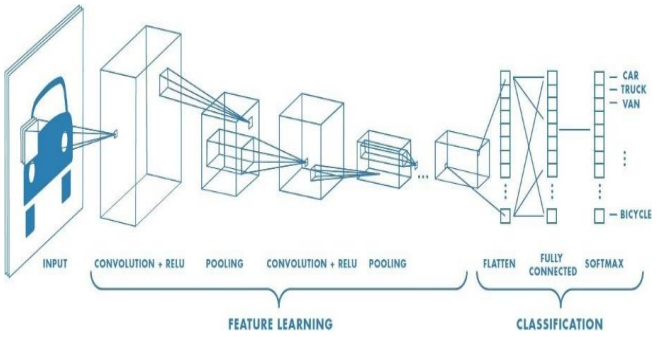
2. "Feature Extraction of White Blood Cells Using CMYK-Moment Localization and Deep Learning in Acute Myeloid Leukemia Blood Smear Microscopic Images"

Medical diagnosis has been transformed by artificial intelligence, especially for cancer. The diagnostic process for acute myeloid leukemia (AML) is laborious and prone to both human and technological mistake. Even after a thorough examination by a skilled pathologist, it might be challenging to reach a definitive determination in a number of cases. Nonetheless, AML-related mistakes and time might be decreased with the use of computer-aided diagnostics (CAD) diagnosis. A crucial stage in the diagnosis of AML is the identification of white blood cells (WBCs), and deep learning is regarded as a cutting-edge method for this purpose. However, the quality of the retrieved features that were utilized to train the pixel-wise classification models has a substantial correlation with the accuracy of WBC identification. Understanding the various patterns of changes connected to WBC counts and characteristics is essential for diagnosing CAD. In this work, deep learning and image processing techniques were combined to create a novel hybrid feature extraction approach. The suggested technique is comprised of two steps: 1) the CMYK-moment localization method is used to extract an area of interest (ROI); and 2) a CNN-based feature fusion method is used to extract features based on deep learning.

3. "Lightweight EfficientNetB3 Model Based on Depth wise Separable Convolutions for Enhancing Classification of Leukemia White Blood Cell Images"

A kind of cancer known as acute lymphoblastic leukemia (ALL) is brought on by an excess of immature white blood cells (WBCs) in the bone marrow. Many categorization models that make use of deep learning

(DL) and machine learning (ML) techniques have recently been created. The research paper proposes a new



robust model with DL assistance that uses depth-wise separable convolutions and EfficientNet-B3 to distinguish between normal cells and acute lymphoblastic leukemia in the dataset of white blood cell pictures. The suggested model is small and seeks to increase the classification task's accuracy. Less trainable parameters are used by the lighter and new EfficientNet-B3 to increase leukemia classification performance and efficiency. In addition, we are evaluating the efficacy and applicability of the suggested lightweight EfficientNetB3 using two publicly accessible datasets. Furthermore, our findings show that the proposed lightweight EfficientNet-B3 model is reliable and useful for aiding medical practitioners and clinical researchers in the diagnosis of leukemia.

4. "Automatic Detection of White Blood Cancer From Bone Marrow Microscopic Images Using Convolutional Neural Networks"

Leukocytes, which are produced in the bone marrow, comprise approximately one percent of the total blood cell count. The excessive proliferation of these white blood cells results in the development of blood cancer. The proposed study offers a strong method for categorizing Acute Lymphoblastic Leukemia (ALL) and Multiple Myeloma (MM) using the SN-AM dataset, in the field of cancer classification. Acute lymphoblastic leukemia, also known as ALL, occurs when the bone marrow produces an excessive amount of lymphocytes, a type of white blood cell. Conversely, Multiple myeloma (MM) is a distinct type of cancer that results in the accumulation of cancer cells in the bone marrow instead of their release into the bloodstream. By utilizing convolutional neural networks, the proposed model eliminates the potential for errors in the manual process using advanced deep learning techniques. Therefore, the model can effectively be utilized as a tool for identifying the cancer type present in the bone marrow.

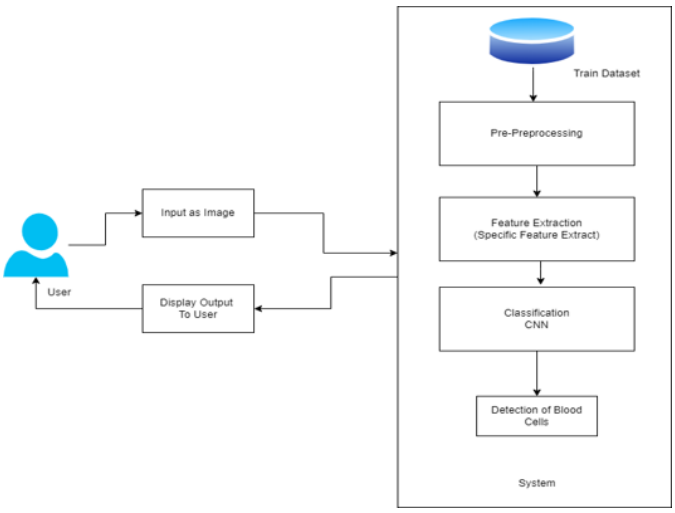


Figure 1. System Architecture

Figure 2: CNN Algorithm

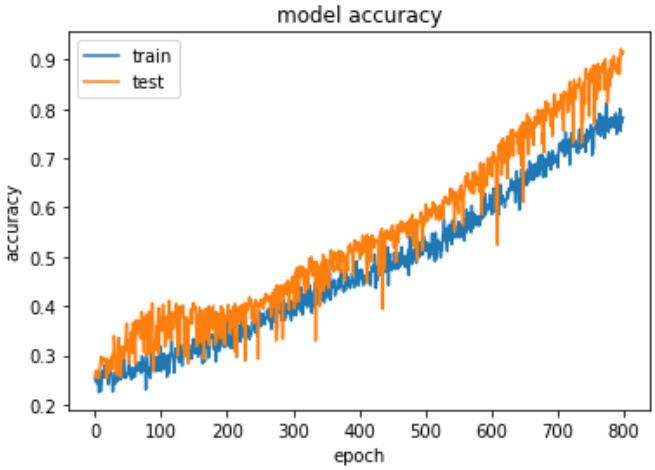


Chart 1: Model Accuracy

3. CONCLUSIONS

One potential direction in medical diagnostics is the automated categorization of blood cells from blood smear images using machine learning algorithms. The goal of this work was to create a digital image analysis system that could reliably and efficiently identify between erythrocytes, or red blood cells, and leukocytes, or white blood cells. Medical diagnostics might be completely transformed by integrating machine learning algorithms for blood cell categorization from blood smear images. With its excellent blood cell identification and classification, the created automated system shows promise for major improvements in medical research, pathology analysis, and clinical diagnostics.

ACKNOWLEDGEMENT

We are grateful to our project guides, Mrs. Vina Lomte, HOD, RMDSSOE, and Mrs. Manisha Darak, Assistant Professor, RMDSSOE, for their unwavering support, tolerance, and inspiration as well as for their insightful advice and insightful throughout the Research process.

REFERENCES

[1] Navya K.T, Keerthana Prasad, Brij Mohan Kumar Singh. " Classification of blood cells into white blood cells and red blood cells from blood smear images using machine learning techniques."
[2] Tusneem Ahmed Elhassan, Mohd Shafry Mohd Rahim, Tan Tian Swee, Siti Zaiton Mohd Hashim and, Mahmoud Ahjurf. “Feature Extraction of White Blood Cells Using CMYKMoment Localization and Deep Learning in Acute Myeloid Leukemia Blood Smear Microscopic Images”
[3] Amreen Batool and Yung Cheol Byun, “Lightweight EfficientNetB3 Model Based on Depth wise Separable Convolutions for Enhancing Classification of Leukemia White Blood Cell Images”
[4] Deepika Kumar, Nikita Jain, Aayush Khurana, Sweta Mittal, Roman Sekerik and Suresh Chandra Satapathy " Automatic Detection of White Blood Cancer From Bone Marrow Microscopic Images Using Convolutional Neural Networks"
[5] Jagrit Mitra, Kartik Vijayran, Kartikeya Verma & Anurag Goel.“ Blood Cell Classification using Neural Network Models”.
[6] Najmeddine Dhieb, Hazim Ghazzai, Hichem Besbes, Yehia Massoud Wei Qi Yan “An Automated Blood Cells Counting and Classification Framework using Mask R-CNN Deep Learning Model Guided.’
[7] Virgnia Mari E. Batitis, Merwin Jhan G. Caballes, Abigail A. Ciudad, Miccaela D. Diaz, Russel D. Flores and Roselito E. Tolentino.“ Image Classification of Abnormal Red Blood Cells Using Decision Tree Algorithm”.
[8] Allaparthi HemaSri, Mopuru Devi Sreenidhi , Valluru Venkata Krishna, Gamidi Vasanth, V.Murali Mohan and T.Satish “Detection of RBCs, WBCs, Platelets Count in Blood Sample by using Deep Learning”.
[9] Hersh Abdulrahman Muhamad, Shahab Wahhab Kareem, and Amin Salih Mohammed, “A Comparative Evaluation of Deep Learning Methods in Automated Classification of White Blood Cell Images”.
[10] J.Setu Sai Sowmya Kumari, K,Chandra Sekhara Chari, M.Leelavathi, K.Pavan Kumar and, Mrs.N.Rajeswari “Classification of white blood cells from microscopic images using CNN”.
[11] Thinam Tamang, Sushish Baral, May Phu Paing. “Classification of White Blood Cells: A Comprehensive Study Using Transfer Learning Based on Convolutional Neural Networks”