Journal of Artificial Intelligence and Computing (JAIC)
ISSN (e): 3005-3358
Volume: 2, Number: 1, Pages: 29- 32, Year: 2024

Blood Cell Disease Prediction with Machine Learning

Ateeqa Arshad[1], Arooj Fatima[1], Asad Tahir[1], Rida Fatima[2], Mudasira Khalil[3]

[1]Department of Computer Science, Islamia University Bahawalpur, Bahawalpur, Pakistan (Email: Seharzari411@gmail.com)
[2]Department of Computer Science, Virtual University, Pakistan
[3]Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Kha, Pakistan

***Abstract*—Our body has cells that provide oxygen to our cell nucleus, called red blood cells (RBC). The other cells that are fighting the disease in our body are called the White Blood Cell (WBC), and some types of cells prevent bleeding with clotting processes. There are three main types of blood cells. If we have some disorder in these cells, then our body faces diseases like leukemia, lymphoma and myeloma. Blood cell diseases can be identified using blood cell image symptoms, causes, and present conditions. Blood smear images with surrounding images of red, platelet, and white cell breakdown play an important role in assessing and diagnosing a large range of illnesses, including infection, disease, and leukemia. Due to the segmentation or breakdown procedures, some image processing techniques are performed to improve picture detection quality. Blood cell segmentation remains a serious challenge. That's why deep learning technology is used to get the blood cell images clearly with the application of cutting-edge technology for the white blood cells, platelets and red blood cells in blood smear images. The accuracy of blood cell smear images required for automatic illness identification has been demonstrated in previous research experiments. To tackle this blood cell disease issue and optimize the system's performance to detect the diseases present in blood cells, we can find through their images by proposing deep learning.**

Index Terms*— *Blood Cell, Machine learning, Deep learning, Blood cell Disease.

I. Introduction

Blood cells are classified as platelets (thrombocytes), white (leukocytes) and red (erythrocytes) blood cells. Hematopoietic cells in the blood make up the term "blood cell.". Blood plasma, on the other hand, accounts for 55% of the tissue. Males have a red blood cell volumetric ratio of 45%, and females have a red blood cell volumetric ratio of 40%. To get oxygen into tissues, hemoglobin, an iron-containing protein, has to transfer oxygen from the lungs to the hemoglobin. Pulmonary circulation is the movement of blood from the heart to the lungs and back again. They all are up to 45% of the blood tissue by volume; the remaining is plasma and liquid components of blood. These elements are found in some tissues such as bone marrow, spleen, lymph nodes, blood and bleeding clotting. Blood disorders mean abnormality in the Blood cells. The most commonly found Blood disorders include anemia, bleeding disorders (e.g., hemophilia, blood clots) and some blood cancers (e.g., leukemia, myeloma and lymphoma). There are some symptoms of red blood cell diseases which appear in the body, such as fatigue, Due to lack of oxygenated blood there is trouble concentrating, shortness of breath, weakness of muscles and a fast heartbeat. There are white blood cell disorder symptoms such as chronic infection, mysterious weight loss and malaise (or unwellness generally). White blood cells are vital to the body's ability to fight against harmful foreign elements (bacteria and viruses). A shortage of healthy white blood cells or an imbalance in the number and types of white blood cells can lead to diseases that require medical attention. Examining white blood cell types and their defects is critical to establishing the person's health. White blood cell counts are generally performed in a lab setting under the guidance of a microscope and with staining techniques. This is a gruelling process, and exhausted examiners are likelier to commit errors. Cell-counting devices that may be bought commercially don't rely on morphology or images to do their work. Blood cells are destroyed during analysis. Using images of white blood cell types as a means of non-destructive classification is an interesting option. The lack of large amounts of training data hampers white blood cell classification using images and machine learning. Some common symptoms of platelets disorders are easy to detect, such as cuts and sores on the body that are difficult to heal or slow in healing, in injury or cut cases, blood doesn't clot, easily skin bruises, mysterious gums and nose bleeding, simply blood flowing not easy to stop in large cases. There are many types and subtypes present of different blood cell disorders, which can cause big trouble if not cured on time, such as blood cancer. Deep learning is one of the most trending subsets of artificial intelligence, especially used for heavy data sets. CNN models are used widely, coupled with Alex Net, Resnet50, Densenet201 and Google Net, and trained with the Kaggle Dataset. Automated image analysis may include acquisition, preprocessing, segmentation, and classification. The most critical step in image processing is the segmentation of the image. The fact that the segmented image should retain the area of interest and discard unwanted information makes the process critical.

A. Objectives

- The main objective is the implementation of a model to be used to detect and classify blood cells.

- Improve the precision with which blood cells are screened.
- Screening costs can be reduced.
- Detect blood cells using deep learning algorithms.

B. Research Questions

- How can we process the dataset images?
- What kind of diseases are predicted in this study?
- What Machine learning model can be fit for the proposed framework?
- What type of Blood cells are used for disease prediction?
- What are the approximate results of the model?

C. Problem Statement

- Blood cell disorders are difficult to identify and require a long test procedure like CBC to confirm the condition and start the procedures to cure it. If it is late, then the cases become rare to cure, like blood cancer.
- The system can detect blood cell disorders on time and fast, and it can help diagnose and treat diseases as quickly as possible, thereby allowing the body's natural healing process.
- There is a CBC test, the Blood test that helps to understand the RBC and WBC status but does not recognize the kind of disease patients may have.
- Diagnosis is not initially possible at the stage of only blood, which is used to detect malaria during screening processes.
- Initially, we cannot diagnose the blood cell disease with additional malaria and non-malaria patients in a single model.

II. LITERATURE REVIEW

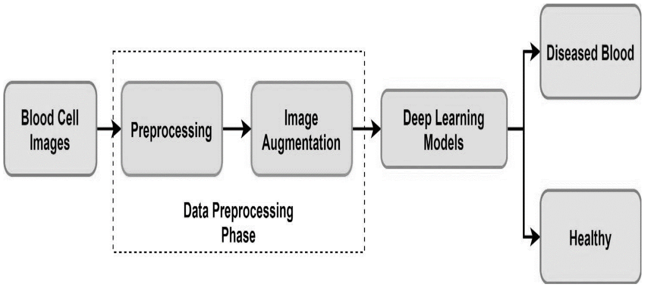
Blood cells are categorized into three primary types: Red Blood Cells (RBCs), White Blood Cells (WBCs), and Platelets. RBCs, primarily produced in the liver, transport oxygen from the lungs to cells throughout the body [1]. WBCs, generated mainly in the bone marrow, defend the body against diseases. Platelets facilitate blood clotting to prevent excessive bleeding [2]. Sickle cell anemia, or Sickle Cell Disease (SCD), results from a disorder in RBCs, causing them to become sickle-shaped and obstruct blood vessels, impeding oxygen delivery [3].

Diagnosing blood cell disorders often begins with examining RBCs. A study addressing the challenge of limited training datasets utilized transfer learning and data augmentation techniques to process blood cell images. By employing ImageNet for transfer learning, the model achieved a classification accuracy of 99.54% and a multi-class SVM classifier accuracy of 99.98% on the erythrocytes IDB dataset and 98.87% on a collected dataset, aiding in the early diagnosis of sickle cell anemia. Another study applied morphological information from 236 blood samples to diagnose COVID-19 using deep learning, achieving 79% classification accuracy and an ROC AUC of 0.90 [5].

The Complete Blood Count (CBC) is a manual laboratory test that counts RBCs, WBCs, and platelets. Image processing techniques, such as RetinaNet, have been employed to train models for detecting and classifying these blood cells. The dependency on confidence thresholds and deep learning epochs was examined, and results were compared with other object detection methods[3] . Leukemia, a cancer affecting mainly WBCs, poses diagnostic challenges. Machine learning (ML) techniques have been reviewed for their efficacy in identifying leukemia from peripheral blood smear (PBS) images, with an average success rate of over 97%. ML, especially deep learning (DL), has demonstrated high precision and sensitivity in detecting various leukemia types [6]. An enhanced hybrid fuzzy C-means clustering algorithm has been proposed for segmenting WBC nuclei from images, followed by feature extraction and classification using a Support Vector Machine, resulting in high image reconstruction quality [7].

Whole Slide Imaging (WSI) represents a significant advancement in pathology, digitizing glass slides for broader clinical practice, education, and research applications. Despite its benefits, WSI's complex implementation remains a challenge. The technology's advantages include improved workflows, reproducibility, and collaboration, though adoption hurdles persist [8]. In a study on cereal grain classification, an area scan camera captured images of various grains, with morphological, colour, textural, and wavelet features extracted for classification. The best results were obtained using a linear discriminant classifier, achieving high accuracy for different grain types [9]. A new blood sample dataset has been created to test and compare segmentation and classification algorithms, providing a valuable resource for advancing research in image processing and pattern matching [10].

Figure 1: Flow process.



III. METHODOLOGY

A. Dataset Collection:

The data set is collected from Kaggle, and we can also download it from the NCBI Library's publically available website. This collection includes 12,500 enhanced blood cell images (JPEG) and cell type designations (CSV). There are over 3,000 photos for each major cell type, organized into four separate folders (according to cell type). Eosinophils, lymphocytes, monocytes, and neutrophils are the different cell types. The 410 original images (pre-augmentation), two additional subtype labels (WBC vs WBC), and bounding boxes for each cell in each of the 410 images (JPEG + XML information), and this dataset are all included in a separate dataset.

Here are sample images of the dataset.

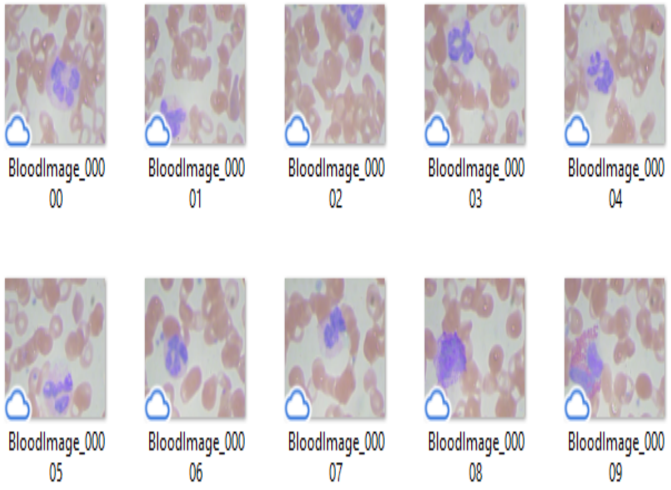


Figure 2: Dataset sample images

Since the dataset consists of red blood cells, White Blood cells, and Platelets, this is the sample from the dataset with all types of blood cells. The Sample blood cell below indicates where RBC, WBC, and Platelets are in the images.

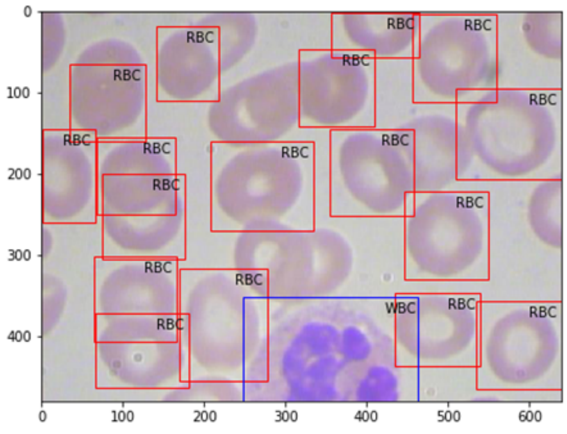


Figure 3: Here is the indication of the RBC, WBC, and Platelets.

B. *Data Visualization*

The Dataset is in the image format. We view and extract the RBC, WBC, and Platelets from the images and view the status of the extracted data in Fig. 3.

Here, we see the sample image of the extracted images from the dataset with blood cell disease. In this image, we see images like monocytes, neutrophils, lymphocytes, eosinophils, malaria, and non-malaria. We are going to find this disease in the blood cell disease.

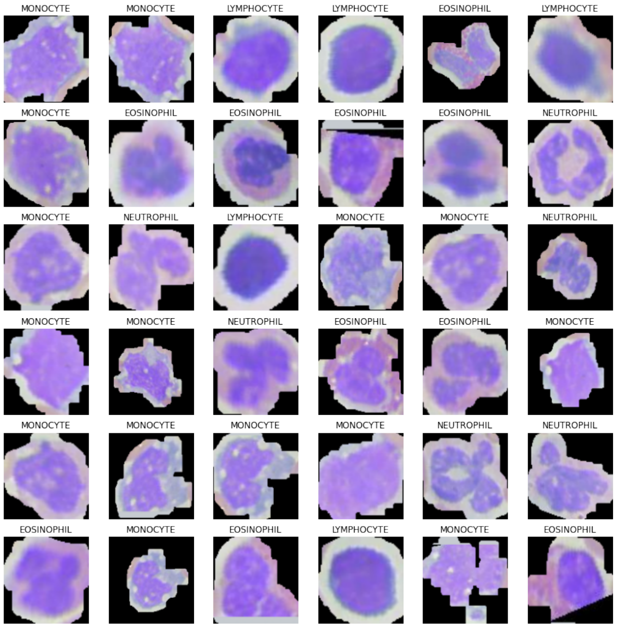


Figure 4: Sample images after loading the dataset with different types of disease label

Here, we see the sample image of the extracted images from the dataset with blood cell disease. In this image, we see images like monocytes, neutrophils, lymphocytes, eosinophils, malaria, and non-malaria. We are going to find this disease in the blood cell disease.

C. *Feature Engineering*

Feature extraction: We use different base techniques to get the qualitative features that lead us to the effective results expected to reach more than 80%, which is acceptable for the proposed study. We use the below-listed features extraction methods:

- Select Image Edges
- Do the Gaussian Blur
- Do the Gray Scale
- Do the Dilate Edges
- Do the eroded edges
- Get the contours of the image
- Get the Boxes of the images
- Load the Data and do all the functions

D. *Train and Test:*

The *train and test* are the processes in which we train the algorithm.

The *test or validation* process consists of some measurement tools after the dataset has been trained. The results are based on the True Positive, True Negative, False Positive, and False.

TP = True Positive, **TN** = True Negative, **FP** = False Positive, **FN** = False Negative

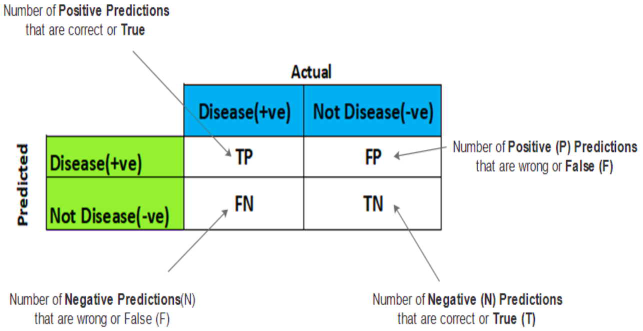

Number of Positive Predictions that are correct or True
Actual
Disease(+ve)
Not Disease(-ve)
Predicted
Disease(+ve)
Not Disease(-ve)
TP
FP
FN
TN
Number of Positive (P) Predictions that are wrong or False (F)
Number of Negative Predictions(N) that are wrong or False (F)
Number of Negative (N) Predictions that are correct or True (T)

Figure 5: Confusion Matrix Predictions validations processes.

E. *Model Classification Accuracy:*

The model validation equations are Precision, Recall, F1 Score, Accuracy, and specificity.

$$precision = \frac{TP}{TP+FP}$$

$$recall = \frac{TP}{TP+FN}$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall}$$

$$accuracy = \frac{TP+TN}{TP+FN+TN+FP}$$

$$specificity = \frac{TN}{TN+FP}$$

F. *Deployment and build the model:*

After fitting and training the model, the deployment will give us two things: one is model training accuracy, and 2nd is model validation accuracy.

G. *Visualization of training and validation of the model:*

After the model's effective accuracy of classification, the model is saved for further system development.
At this step, we visualize the results, as in the sample below.

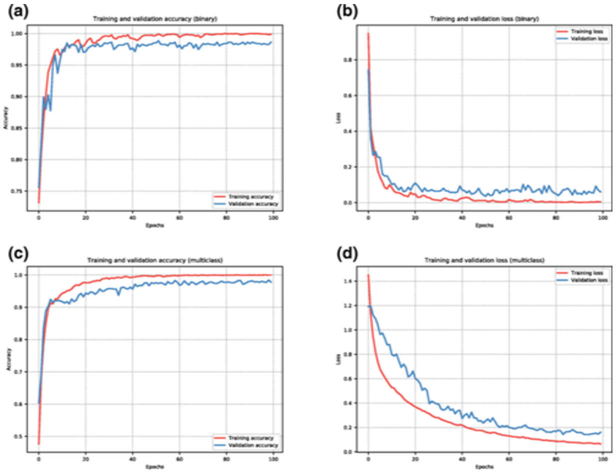


Figure 6: Results visualization.

IV. Conclusion

The process is based on the classification of the blood cell disease in terms of RBC, WBC, and Platelets based on the smear images of the blood cells. The assessment of the proposed study shows that the data availability and the possibility of the model designing ingredients are almost perfect for the synopsis. We try our best to make things possible and build a good productive, and efficient model for the development of blood disease prediction.

References

[1] Tavakoli, S., Ghaffari, A., Kouzehkanan, Z. M., & Hosseini, R. (2021). New segmentation and feature extraction algorithm for classification of white blood cells in peripheral smear images. Scientific Reports, 11(1), 1-13.

[2] Bernard, N. J. (2021). Another function for red blood cells. Nature Immunology, 22(12), 1469–1469.

[3] Drałus, G., Mazur, D., & Czmil, A. (2021). Automatic Detection and Counting of Blood Cells in Smear Images Using RetinaNet. Entropy, 23(11), 1522. https://doi.org/10.3390/e23111522

[4] Alzubaidi, L., Fadhel, M. A., Al-Shamma, O., Zhang, J., & Duan, Y. (2020). Deep learning models for classification of red blood cells in microscopy images to aid in sickle cell anemia diagnosis. Electronics, 9(3), 427.

[5] Cooke, C. L., Kim, K., Xu, S., Chaware, A., Yao, X., Yang, X., Neff, J., Pittman, P., McCall, C., Glass, C., Jiang, X. S., & Horstmeyer, R. (2021). Deep Optical Blood Analysis: COVID-19 Detection as a Next Generation Blood Screening Case Study. Hematology. https://doi.org/10.1101/2021.07.18.21259553

[6] Ghaderzadeh, M., Asadi, F., Hosseini, A., Bashash, D., Abolghasemi, H., & Roshanpour, A. (2021). Machine Learning in Detection and Classification of Leukemia Using Smear Blood Images: A Systematic Review. Scientific Programming, 2021, 1–14.

[7] Begum, A. R., & Razak, T. A. (2017). A Proposed Novel Method for Detection and Classification of Leukemia using Blood Microscopic Images. International Journal of Advanced Research in Computer Science, 8(3).

[8] Zarella, M. D., Bowman;, D., Aeffner, F., Farahani, N., Xthona;, A., Absar, S. F., Parwani, A., Bui, M., & Hartman, D. J. (2019). A Practical Guide to Whole Slide Imaging: A White Paper From the Digital Pathology Association. Archives of Pathology & Laboratory Medicine, 143(2), 222–234. https://doi.org/10.5858/arpa.2018-0343-RA

[9] Choudhary, R., Paliwal, J., & Jayas, D. S. (2008). Classification of cereal grains using wavelet, morphological, colour, and textural features of non-touching kernel images. Biosystems Engineering, 99(3), 330–337. https://doi.org/10.1016/j.biosystemseng.2007.11.013

Labati, R. D., Piuri, V., & Scotti, F. (2011). All-IDB: The acute lymphoblastic leukemia image database for image processing. 2011 18th IEEE International Conference on Image Processing, 2045–2048.