IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125287

IDENTIFICATION AND OBSERVATION OF IMMATURE WHITE BLOOD CELLS USING CNN AND MACHINE LEARNING

Praful Shah[1], Shreya Wavhal[2], Praveen Mahato[3], Amit Kanani[4], Prof. Komal Yadav[5]

Shree Ramchandra College of Engineering, Lonikand Pune[1-4]

Guide, Shree Ramchandra College of Engineering, Lonikand Pune[5]

Abstract: Blood smear samples are microscopic preparations of blood cells that serve as a valuable tool in medical diagnostics. By spreading a thin layer of blood onto a glass slide, staining it, and examining it under a microscope, healthcare professionals can gain essential insights into various physiological and pathological conditions. These samples allow for the visualization and characterization of different types of blood cells, including red blood cells, platelets, and white blood cells (WBCs). The classification and identification of WBCs in blood smear samples play a crucial role in diagnosing and monitoring infectious, hematologic, and immune disorders. This review paper focuses on the development and evaluation of an automated system for identifying and classifying different types of white blood cells in blood smear images. The manual process of WBC classification is labor-intensive, time-consuming, and subject to inter-observer variability. In response to these challenges, researchers have increasingly explored automated approaches that employ image processing techniques and machine learning algorithms. These automated systems aim to enhance the efficiency, accuracy, and consistency of WBC classification.

Keywords: WBC, White Blood Cells, Classification Algorithms, CNN.

I. INTRODUCTION

White blood cell (WBC) classification plays a vital role in medical diagnostics, aiding in the diagnosis and monitoring of various infectious, hematologic, and immune disorders. Traditionally, this classification has been performed manually by skilled medical professionals, which is time-consuming, labor-intensive, and subject to inter-observer variability. However, with advancements in image processing and machine learning, the development of an automated system for identifying and classifying WBCs in blood smear images has emerged as a promising approach. The motivation behind developing an automated system stems from the need for accurate, efficient, and cost-effective solutions in medical diagnostics. By automating the process of WBC classification, we can mitigate the limitations associated with manual methods, such as human error, subjective interpretation, and time constraints. Moreover, an automated system has the potential to improve the accuracy and consistency of results, reduce the workload on medical professionals, and enable timely diagnosis for improved patient care.

II. MOTIVATION

The Person will be able to upload blood smears sample as image as input. The system will perform feature extraction and Provide the output. The White Blood cell Type is an important part of the blood test. The accurate cell type will help for detecting potential diseases and related lesions, and determining the type and development of the patient's diseases.

Faster RCNN was used to detect the location of immature white blood cells in the blood, and then CNN neural network was used to classify type of white blood cells. Some preprocessing methods are also used for accuracy comparisons, such as histogram equalization. The highest precision is 100%.

III. LITERATURE REVIEW

A. JH Chein, S Chang, S Cheng, Yen-Chieh Ouyang

The test data set was obtained from Taichung Tzu Chi Hospital. In order to find the best experimental neural network structure, we have tried many different neural network architectures.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023
DOI: 10.17148/IJARCCE.2023.125287

B. Ranjitha P, Sudharshan, Duth P

we make use of K-means, for identifying cancerous stages and its early detection. Experimentation and results were found to be promising with the accuracy of 90% identification of the cancer cells. The main reason behind occurrence of Leukemia is when bone marrow produces a lot of abnormal white blood cells this happens. Microscopic study on images is done by Hematologists who make use of human blood samples, from which it leads to the requirement of following methods, which are microscopic color imaging, image segmentation, clustering and classification which allows easy identification of patients suffering from this disease.

C. Atul Sharma, Gurbaksh Phonsa

Convolutional neural network(CNN) is a technique which we can use for the image classification. This paper will show how the image classification works in case of cifar-10 dataset. We used the sequential method for the CNN and implemented the program in jupyter notebook.

D. Satvik Dasariraju, Marc Huo, Serena McCalla

A random forest algorithm was trained for the detection and classification of immature leukocytes. The model achieved 92.99% accuracy for detection and 93.45% accuracy for classification of immature leukocytes into four types.

E. DanWei, XuejiaoZeng, Zhangru Yang, Quanyu Zhou, XiaofuWeng

Two artificial neural networks were trained to identify blood vessels and cells in the acquired images. Results and Conclusion: Using this technique, single CTCs and CTC clusters were readily distinguished by their morphology.

F. Qiwei Wang, Shusheng Bi Minglei Sun, Yuliang Wang, Di Wang, Shaobao Yang

To improve recognition performance, some key factors involving these object detection approaches are explored and the detection models are generated using the train set of 14,700 annotated images.

G. Farhana Sultana, Abu Sufian, Paramartha Dutta

In this paper, We have explained different CNN architectures for image classification. Through this paper, we have shown advancements in CNN from LeNet-5 to latest SENet model.

IV. AIM

The aim of the project is to develop an automated system that can accurately classify different types of white blood cells (WBCs) in blood smear images. The system should leverage machine learning algorithms, such as CNN and XGBoost, to analyze the images and make precise predictions regarding the specific WBC subtypes present. The ultimate goal is to provide a reliable and efficient tool for medical diagnostics, which can assist healthcare professionals in identifying and monitoring various blood-related disorders or conditions.

V. METHODOLOGY

A. Problem Statement

The identification and classification of different types of white blood cells (WBCs) in blood smear samples is a critical step in the diagnosis of various diseases. Traditionally, this task has been performed manually by trained medical professionals, which can be time-consuming and subjective. Therefore, there is a need for an automated system that can accurately and efficiently classify WBCs in blood smear images. The goal of this project is to develop a system that can take a blood smear sample image as input and accurately classify the type of white blood cell present in the image. This system will utilize image processing techniques and machine learning algorithms to automatically identify type of WBC.

B. Software Information

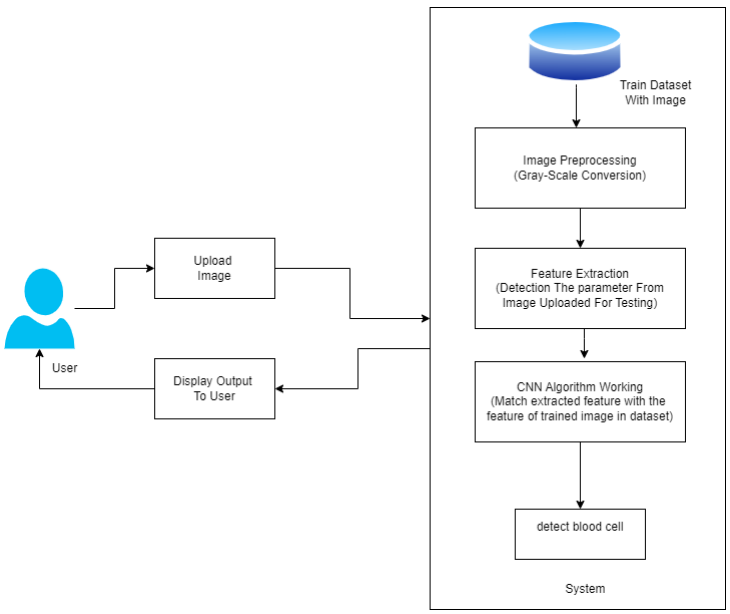
Python is a high-level, general-purpose interpreted programming language. It was created by Guido van Rossum in 1991. Python emphasizes code readability by utilizing whitespace indentation. Its language constructs and object-oriented methods allow programmers to write clean and understandable code for projects of various sizes. Python is dynamically typed and includes garbage collection. Python supports multiple programming paradigms, including object-oriented, structured (procedural), and functional programming. One of the strengths of Python is its extensive standard library, which provides a wide range of modules and functions, making it a "batteries included" language. Python replaced ABC

DOI: 10.17148/IJARCCE.2023.125287

in the late 1980s. Python 2.0, released in 2000, introduced significant features such as list comprehensions and reference counting garbage collection. However, Python 3.0, released in 2008, brought about substantial changes to the language that made it incompatible with Python 2 code. The Python 2 series officially ended in 2020, with version 2.7.18 being the final release. Anaconda is a free and open-source distribution of Python and R that focuses on scientific computing. It simplifies package management and deployment for various applications, including data science, machine learning, and large-scale data processing. Anaconda includes packages for Windows, Linux, and macOS. Anaconda, Inc., founded by Peter Wang and Travis Oliphant in 2012, offers Anaconda Distribution, Anaconda Team Edition, and Anaconda Enterprise Edition as commercial products. Conda, a package management system, is used to manage Anaconda package versions. Conda was later separated as an open-source package due to its usefulness outside of Python. For a minimal installation, there is also Miniconda, which includes only conda, Python, their dependencies, and a few other packages. While Anaconda comes with around 250 packages, the Python Package Index (PyPI) provides access to over 7,500 open-source packages. Anaconda also includes the conda package manager and a virtual environment manager. Anaconda Navigator, a graphical user interface (GUI), provides an alternative to the command-line interface (CLI) for managing Anaconda environments and packages.

C. Convolutional Neural Networks (CNN)

It is a type of deep learning algorithm that is specifically designed for analyzing visual data, such as images and videos. CNNs have been widely successful in various computer vision tasks, including image classification, object detection, image segmentation, and more. CNNs make use of convolutional layers, which apply filters or kernels to the input image. These filters detect specific features, such as edges, corners, or textures, by convolving them across the input image. Convolutional layers allow the network to learn hierarchical representations of the input data. CNNs are trained using labeled data through a process called backpropagation. The weights of the network are adjusted iteratively based on the calculated gradients of a chosen loss function, allowing the network to learn and improve its predictions. Pretrained CNN models, such as VGG, ResNet, and Inception, are trained on large datasets, like ImageNet, and are often used as a starting point for transfer learning. Transfer learning involves taking a pretrained model and fine-tuning it on a smaller, task-specific dataset, which can significantly improve performance even with limited training data.



D. XGBoost (eXtream Gradient Boosting)

It is a supervised learning algorithm. It is an enhanced implementation of the gradient boosting machine (GBM) algorithm, which is a popular ensemble learning method for supervised learning tasks. XGBoost works by sequentially adding weak learners, typically decision trees, to an ensemble. It aims to minimize a specified loss function by iteratively fitting new models to the residuals (the differences between the predicted and actual values) of the previous models. Each new model focuses on correcting the errors or deficiencies of the existing ensemble, gradually improving the overall prediction accuracy. XGBoost offers several enhancements and optimizations over traditional gradient boosting algorithms. It incorporates regularization techniques to prevent overfitting, supports efficient parallelization, and utilizes a tree pruning technique called "gradient based one-side sampling" to improve computational efficiency and reduce memory usage.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023
DOI: 10.17148/IJARCCE.2023.125287

VI. SYSTEM REQUIREMENT SPECIFICATION

A. Hardware Requirements

- RAM minimum Required is 8 GB RAM.
- 40GB Hard disk.
- Processor should be intel 5 or above.

B. IDE

- PyCharm
- Best Integrated Development Environment as it gives possible suggestions at the time of typing code snippets that makes typing feasible and fast.

C. Coding Language

- Python Version 3.5

D. Operating System

- Windows 10 or Above

E. Software Interfaces

- Operating System Should be Windows 10 or above.
- IDE : PyCharm, Spyder
- Programming Language : Python

VII. NON FUNCTIONAL REQUIREMENTS

A. Performance Requirements

- All functions and modules within the software should be well-optimized to deliver fast and efficient performance. This includes tasks such as data processing, calculations, and communication with external systems. Careful attention should be given to algorithm efficiency and resource utilization to avoid bottlenecks and delays. The performance of the software as a whole is crucial for user efficiency. This includes factors like response time, loading speed, and overall system resource utilization. Monitoring and performance testing can help identify any areas that require optimization. If the software involves encryption of data, it should be implemented in a way that ensures fast and secure encryption and decryption processes. Efficient encryption algorithms and techniques can be used to minimize the impact on overall performance

B. Safety Requirements

- The application is designed in modules where errors can be detected and fixed easily. This makes it easier to install and update new functionality if required.

C. Software Quality Attributes

- Adaptability: This software is adaptable by all users.
- Availability: This software is freely available to all users. The availability of the software is easy for everyone.
- Maintainability: After the deployment of the project if any error occurs then it can be easily maintained by the software developer.
- Reliability: The performance of the software is better which will increase the reliability of the Software.
- User Friendliness: Since, the software is a GUI application; the output generated is much user friendly in its behavior.
- Integrity: Integrity refers to the extent to which access to software or data by unauthorized persons can be controlled.
- Security: Users are authenticated using many security phases so reliable security is provided.
- Testability: The software will be tested considering all the aspects.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

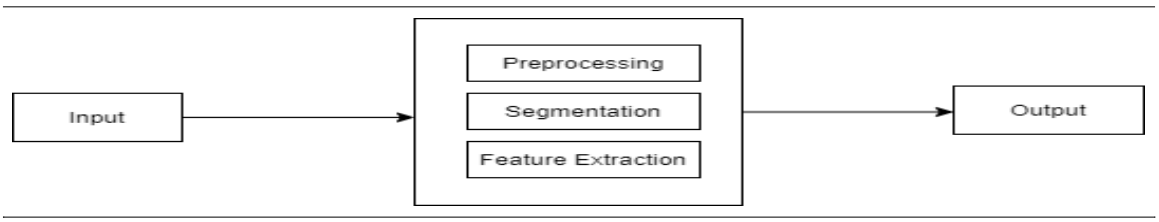
DOI: 10.17148/IJARCCE.2023.125287

VIII. UML DIAGRAMS

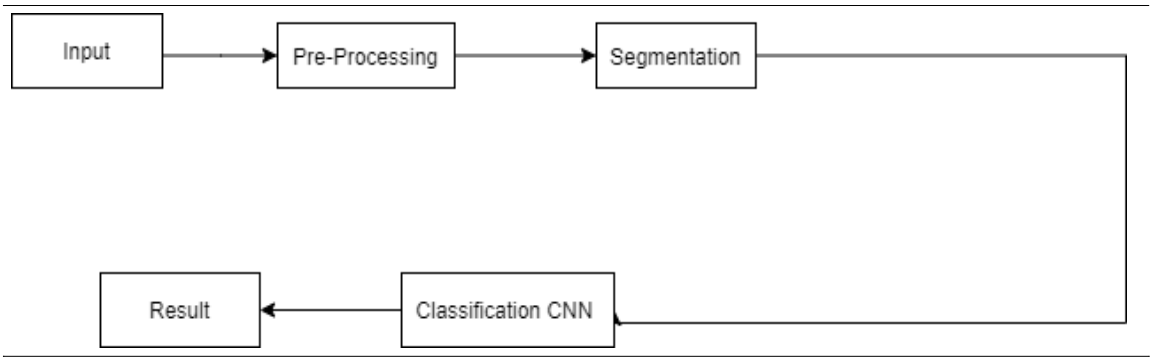
A standardized language for the creation of software blueprints is known as Unified Modeling Language. Although the UML is process agnostic, it should ideally be utilized in a process that is use case driven, architecture-centric, iterative, and incremental. The UML can be used to visualize, specify, develop, and document the artifacts of a software heavy system. The number of UML Diagrams that are at your disposal.

A. Data Flow Diagram (Level 0)

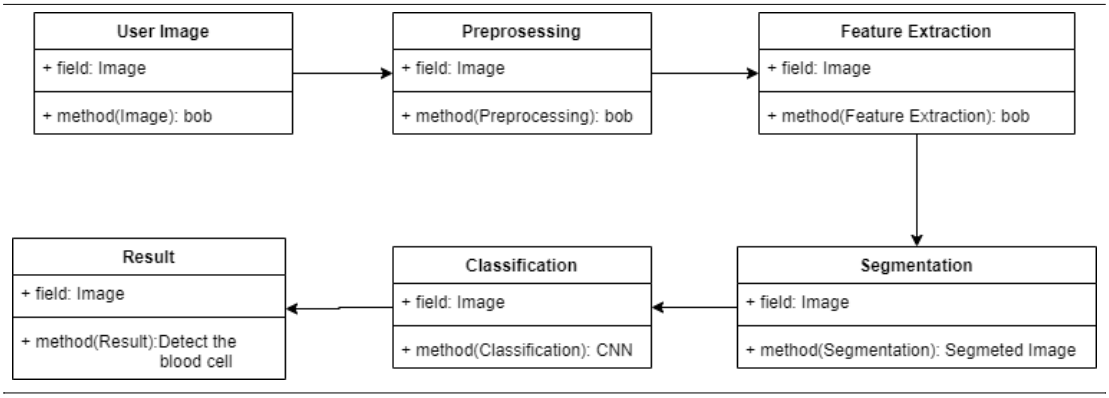
B. Data Flow Diagram (Level 1)



C. Data Flow Diagram (Level 2)



D. Class Diagram



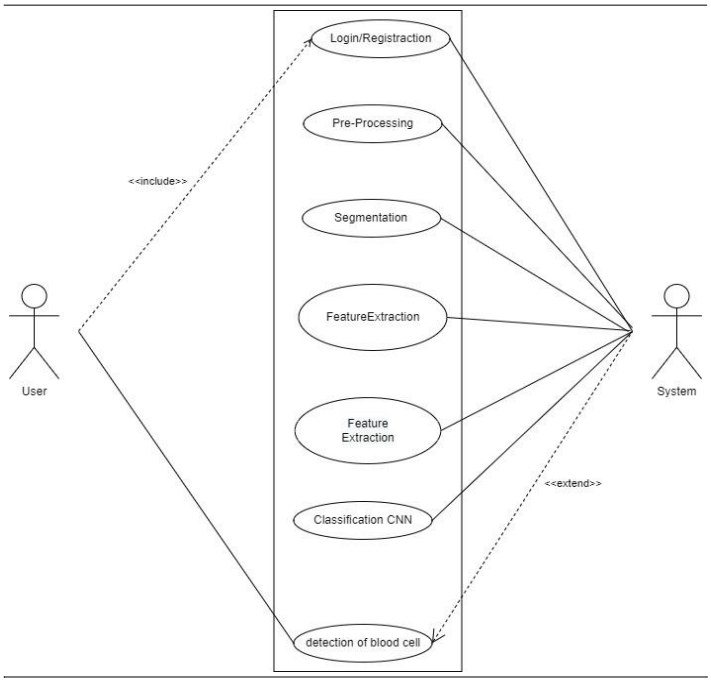
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

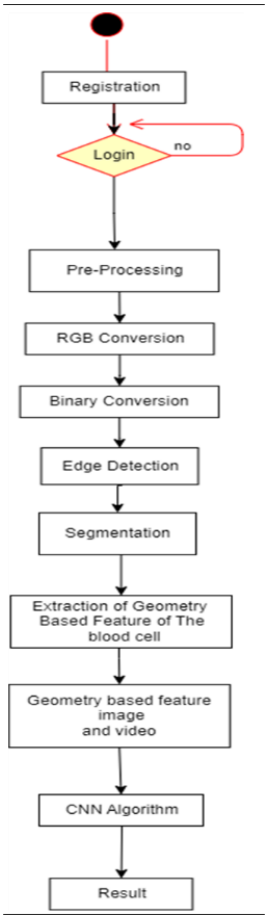
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125287

E. Use Case Diagram



F. Activity Diagram



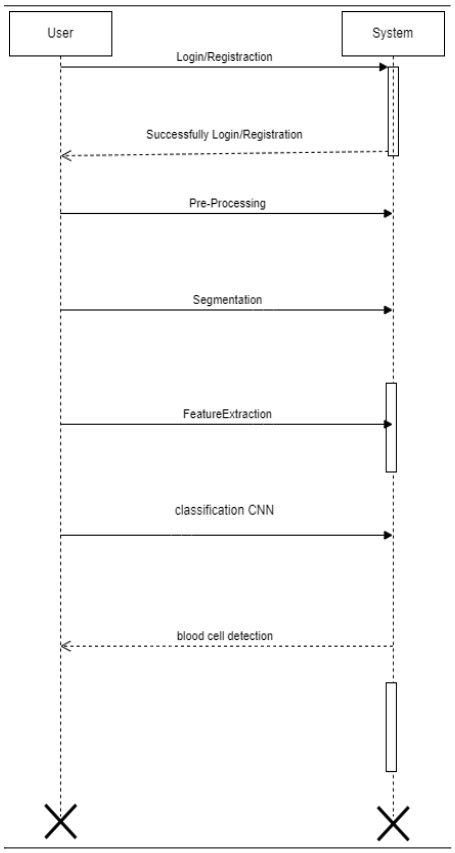
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125287

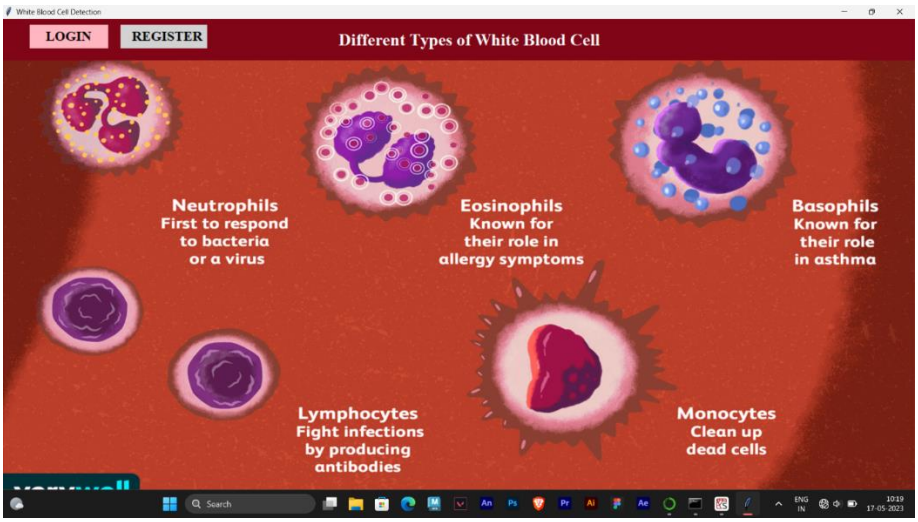
G. Sequence Diagram



IX. OUTPUT AND RESULTS

The following steps should be taken in order to classify blood smear image sample images on the basis of Type of White Blood Cells.

- Step 1 : Open the Application.



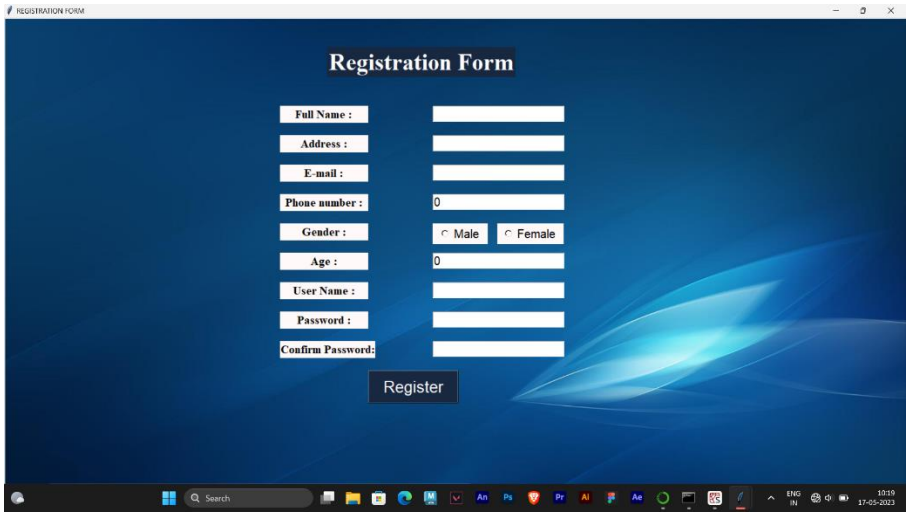
- Step 2 : Register with Personal Details.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

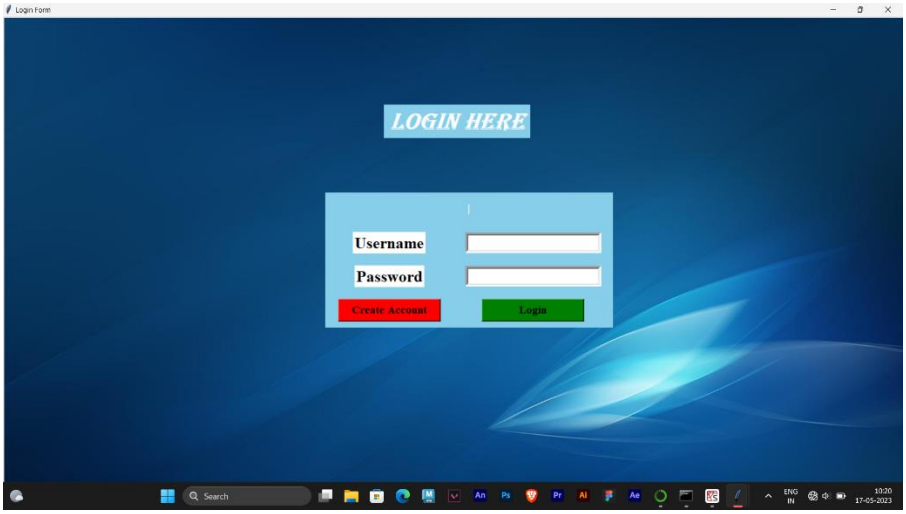
International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

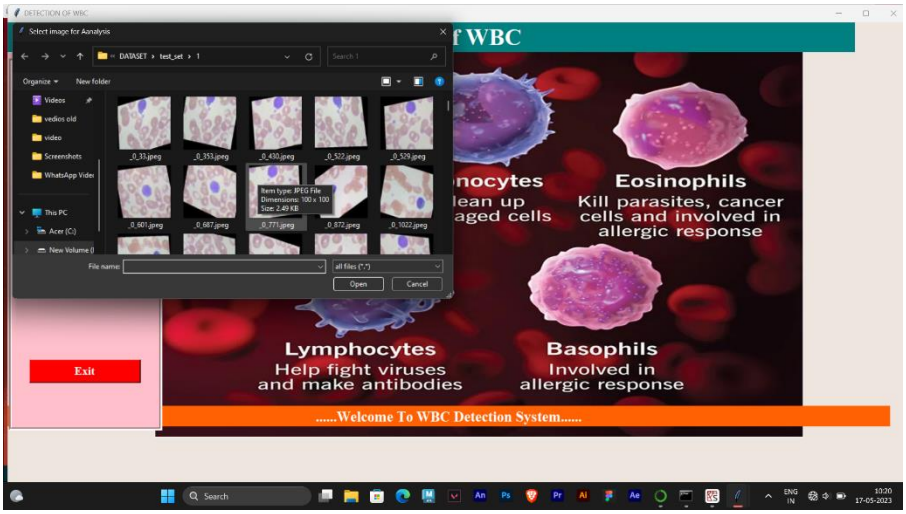
DOI: 10.17148/IJARCCE.2023.125287



- Step 3 : Login to your Account.



- Step 4 : Click on select image option and Choose a blood smear sample image from the Computer.



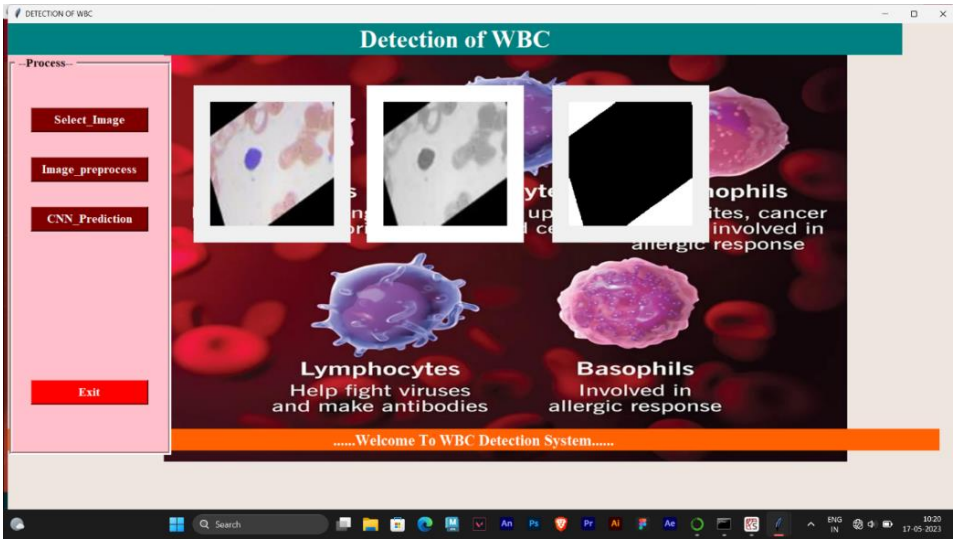
- Step 5 : Click on Image Preprocess option to start the Preprocessing of the sample Image.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

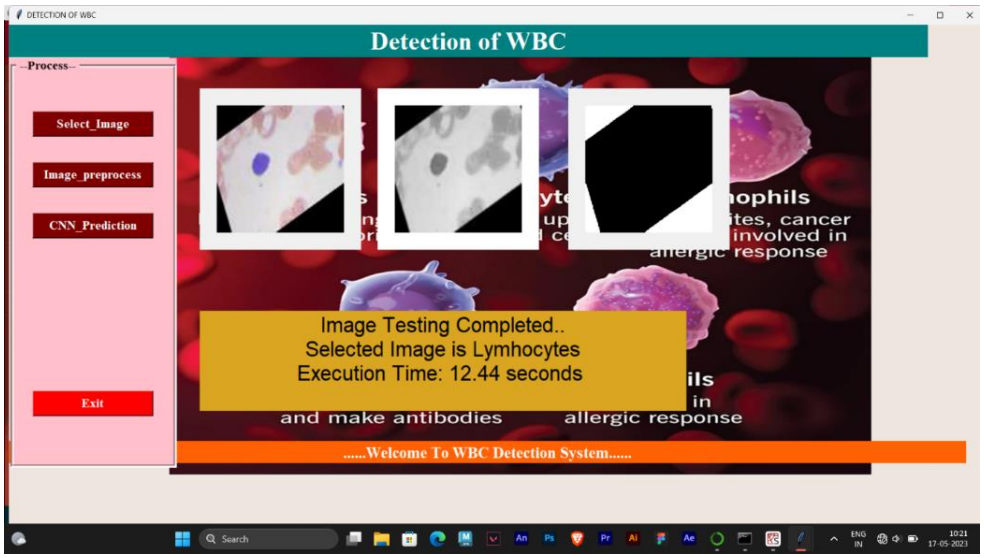
International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125287



- Step 6 : Click on CNN prediction or XG Boost Prediction in order to get prediction output of respective algorithms.



X. CONCLUSION

In this paper, we use deep learning methods to detect and classify immature white blood cells In conclusion, an automated system for identifying and classifying different types of white blood cells in blood smear images offers several advantages that can improve the efficiency, accuracy, and cost-effectiveness of diagnostic procedures.

It has a wide range of potential applications in the field of healthcare and medical research, including clinical diagnostics, disease monitoring, research, education, telemedicine, and quality control. However, the development and implementation of such a system require careful consideration of its limitations, including the quality and size of the dataset, the quality of the images, heterogeneity in morphology, computational requirements, interpretability, regulatory approval, and maintenance.

Despite these challenges, the development of an automated system for identifying and classifying different types of white blood cells in blood smear images holds great promise for improving patient care and advancing medical research

DOI: 10.17148/IJARCCE.2023.125287

REFERENCES

[1]. JH Chein, S Chang, S Cheng Yen-Chieh Ouyang, "IDENTIFICATION AND DETECTION OF IMMATUR WHITE BLOOD CELLS THROUGH DEEP LEARNING", Publisher IEEE 3rd Global Conference of life Sciences and Technologies(Life Tech), 2021.

[2]. Ranjitha P, SudharshanDuth P, Department of Computer Science, Department of Computer Science, Amrita school of Arts and Sciences, Mysuru Amrita school of Arts and Sciences, Mysuru Amrita VishwaVidyapeetham, India Amrita VishwaVidyapeetham, " Detection of Blood Cancer-Leukemia using K-means Algorithm", IEEE 5th International conference on intelligent computing and control systems (ICICCS) 2021.

[3]. Atul Sharma, Gurbaksh Phonsa "Image Classification Using CNN" Proceedings of the International Conference on Innovative Computing & Communication (ICICC) 2021.

[4]. Satvik Dasariraju, Marc Huo, Serena McCalla, "Detection and Classification of Immature Leukocytes for diagnosis of acute Myeloid Leukemia using Random Forest Algorithm", 2020.

[5]. DanWei, XuejiaoZeng, Zhangru Yang, Quanyu Zhou, XiaofuWeng, Hao He, WenyuanGao, " Visualizing Interactions of Circulating Tumor Cell and Dendritic Cell in the Blood Circulation Using In Vivo Imaging Flow Cytometry*", IEEE Transections on Biomedical Engineering*, 2019.

[6]. Qiwei Wang, Shusheng Bi, Minglei Sun, Yuliang Wang, Di Wang, Shaobao Yang , " Deep Learning approach to peripheral leukocyte recognition", *Plos one – journals.plos.org* , 2019.

[7]. Farhana Sultana, Abu Sufian, Paramartha Dutta, "Advancements in image classification using convolutional neural network" Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN) 2018.

[8]. R. Girshick, J. Donahue, T. Darrell, J. Malik, "Region-based Convolutional Networks for Accurate Object Detection and Segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 142- 158, 2016.

[9]. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified real-time object detection", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779-788, 2016.

[10]. Travis Williams, Robert Li "Advanced image classification using wavelets and convolutional neural networks" 15th IEEE international conference on machine learning and applications (ICMLA), 2016.

[11]. R. Girshick, "Fast R-CNN", IEEE International Conference on Computer Vision, pp. 1440-1448, 2015.

[12]. F. Schroff, D. Kalenichenko and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 815-823, 2015.

[13]. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", *28th International Conference on Neural Information Processing Systems*, vol. 1, pp. 91-99, 2015.

[14]. Taigman, M. Yang, M. A. Ranzato and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1701-1708, 2014.

[15]. Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation Applied to Handwritten Zip Code Recognition", Neural Computation, vol. 1, no. 4, pp. 541- 551, 1989.