IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

DETECTION OF ABNORMAL LEUCOCYTES USING MACHINE LEARNING ALGORITHMS

Mrs. TULASI MIRIYALA [*1], Mr. GANGADHAR DOMA[*2]

[*1] Assistant Professor, Department of Computer Science and Applications,
AQJ CENTRE FOR PG STUDIES,
Gudilova, Anandapuram, Visakhapatnam-531173

[*2] Associate Professor, Department of Computer Science and Applications,
AQJ CENTRE FOR PG STUDIES,
Gudilova, Anandapuram, Visakhapatnam-531173

ABSTRACT

The analysis of a patient's blood sample is a significant responsibility in the medical industry. Blood cell abnormalities create a variety of health issues. White blood cells are one of the most important components of blood. White blood cells are immune cells that fight infections caused by bacteria and viruses in the body. White blood cells can be classified to assist us in identifying various illnesses. Due to a variety of medical conditions, normal white blood cells might change in size, shape, and texture. The objective of this research is to enhance the number of detectable aberrant white blood cells via image processing. For categorization, this study employed various machine learning methods such as Random Forest, XgBoost, and Decision Tree. As a consequence, the algorithm recognised and classified four images of white blood cells. These photos were derived by the system from prior hospital patients. Furthermore, the slides are photographed. The image was then placed into the programme. The system processed and classified the image. In effect, the results indicate the identity of the aberrant white blood cells discovered in the system's image, as well as a soft copy of the checklist.

KEY WORDS:

White Blood Cells, Image Processing, Xgboost, Random Forest, Decision Trees

1. INTRODUCTION

White blood cell are also known as leucocytes. Blood is a fluid that transports nutrients, oxygen, and waste products from metabolic processes to cells. Blood circulates throughout the body, passing via the heart, arteries, and other organs. Blood accounts about 7 to 8 % of total body weight. The three major kinds of human blood cells are platelets, white blood cells (WBCs), and red blood cells (RBCs). White blood cells are immune system cells that protect the body against both infectious illness and external invaders, all white blood cells are created and developed from hematopoietic stem cells, which are multipotent in the bone marrow. Leukocytes can be found in many parts of the body, including the blood and lymphatic systems.

White blood cells, like other blood cells, have nuclei, which distinguishes them froma nucleated red blood cell (RBCs) and platelets. White blood cells are often categorized according to cell lineage (myeloid cells or lymphoid cells). White blood cells are a component of the body's immune system. They assist the body in fighting infection and other disorders. Granulocytes (neutrophils, eosinophils, and basophils), monocytes, and lymphocytes are the three types of white blood cells (T cells andB cells). Myeloid cells (myelocytes) comprise neutrophils, eosinophils, mast cells, basophils, and monocytes. Dendritic cells and macrophages are subtypes of monocytes, Phagocytic cells include monocytes and neutrophils. Lymphoid cells(lymphocytes) comprise T cells (helper T cells, memory T cells, and cytotoxic T cells), B cells (plasma cells and memory B cells), and natural killer cells. Historically, white blood cells were categorised based on their physical properties (granulocytes and agranulocytes), but this categorization method is no longer widely used.

They are produced in the bone marrow and protect your body against infections and illness. When there are too many white blood cells in your body, it typically suggests you have an infection and inflammation. A high white blood cell count may suggest certain blood Malignancies or bone marrow problems. Neutrophils are the most numerous white blood cell, according for 60-70%of all circulating leukocytes. They protect against bacterial or fungal infection. They are typically the initial responses to microbial infection, their activity and death in high numbers generate pus. Neutrophiles are the most prevalent cell type found in the early stages of acute inflammation. Various methods have revealed that the typical lifetime of inactivated human neutrophils un circulation is between 5 and 135 hours.

Eosinophils account for approximately 2-4%of all white blood cells in circulation. This number changes during the day, throughout the year, and during menstruation its rises as a result of allergies, parasite infections, collagen illnesses, and spleen and central nervous system disease. Lymphocytes are far more frequent in the lymphatic system than in the blood. Lymphocytes are recognized by a highly stained nucleus that may be positioned irregularly, as well as a little quantity of cytoplasm. Monocytes, the biggest kind of white blood cell, perform the same "vacuum cleaner" function as neutrophils but survive considerably longer because they play an additional role: they convey pathogen fragments to t cells so that the pathogens can be detected and destroyed again.

2. LITERATURE SURVEY

This section will deal with all the previous information related to brain tumor and several methods for identifying the leucocyte prediction. Literature survey is the most important step in software development process. For any software or application development, this step plays a very crucial role by determining the several factors like time, money, effort, lines of code and company strength. Once all these several factors are satisfied, then we need to determine which operating system and language used for developing the application. Once the programmers start building the application, they will first observe what are the pre-defined inventions that are done on same concept and then they will try to design the task in some innovated manner.

MOTIVATION

Virgnia's Image Classification of Abnormal Red Blood Cells Using a Decision Tree Algorithm [1] Merwin Batitis, Mari E. Batitis II Miccaela D. Dnd iaz, Jhan G. Caballes, Abigail A. Ciudad, Jhan G. Caballes Russel D. Flores published in Proceedings of the Fourth International Conference on Computing Methodologies and Communication (ICCMC 2020) IEEE Xplore Part Number:CFP20K25-ART; ISBN:978-1-7281-4889-2. The purpose of the study is to increase the amount of defective red blood cells observable using image processing. In this work, the Decision-Tree Algorithm was employed as a machine learning algorithm. Algorithms for classification as a result, the system identified and reported 10 aberrant red blood cells. Images in the system came from prior hospital patients. A camera is also used to catch the slides. After that, the image was added.

Mohammad Syahputra [2] and colleagues (2017). Using a Radial Basis Function Network, we can classify abnormal red blood cell shapes. Blood analysis, in addition to a physical examination, is a reliable tool for diagnosing illnesses. This is because blood includes several vital components. Morphological examination of peripheral blood smears is a critical lab examination that must be performed correctly. However, because of precision, concentration, and a lack of understanding, atypical red blood cell shapes detected by a health analyst are not always the same as those discovered by other analysts. Furthermore, morphogical assessment of peripheral blood smears is still done manually by health analyzers, which is regarded inefficient due to the considerable time it takes. To overcome this issue, a classification system was developed.

Pooja Tukaram Dalvi [3] The categorization and counting of red blood cells are critical in diagnosing disorders such as iron deficiency anaemia and vitamin B12 deficiencyanaemia. In this study, we plan to create a standalone application that can categorize red blood cells into four aberrant types: elliptocytes, echinocytes, tear drop cells, and macrocytes. We employed two data mining classifiers, Artificial Neural Network and Decision Tree Classifier, and compared their accuracy in categorizing red blood cells.

Vishwas Sharma Et. Al [4] The purpose of this study is to detect sickle cell anaemia ad thalassemia. The suggested technique entails acquiring thin blood smear microscopic pictures, pre-processing with a median filter, segmentation of overlapping erythrocytes with marker-controlled watershed segmentation, morphological operations to improve the images, extraction of characteristics such as metric value, aspect ratio, radial signature, andits variance, and lastly training the K-nearest neighbour classifier to test the images.

Krishna Kumar Jha, Et. Al [5] The counting of red blood cells, white blood cells, and platelets in human blood in a microscopic picture might give important information about the patient's health. The traditional procedures employed in laboratories to detect anddiagnose the issue were time-consuming and prone to error. Digital image processing can identify and distinguish many disorders, allowing for quick and cost-effective analysis and diagnosis of the real problem. The primary goal of this study is to count the quantity of WBCand identify those that are aberrant. On the basis of counts, the suggested study may identifymany diseases such as COPD, immune system disorders, neutropenia, HIV/AIDS, lymphocytopenia, leukaemia, and so on.

3. EXISTING METHODOLOGY

The current technique, which is explained using a Decision-Tree, can be used to classify aberrant red blood cells discovered in the blood. Based on the data acquired from the 40 photos, which were made up of 600 sample cells, the system was able to categorise the aberrant red blood cells using the Decision Tree Algorithm. Classification errors were caused by minor discrepancies in the attributes used. Because elliptocytes and ovalocytes have nearly identical measurements and characteristics, distinguishing the two aberrant red blood cells is difficult.

LIMITATION OF EXISTING SYSTEM

1. Low Accuracy
2. Decision tree has low accuracy of 83%.

4. PROPOSED SYSTEM & ITS ADVANTAGES

The proposed approach employs the Random Forest algorithm to detect irregularities in pictures of white blood cells. The system is also developed by working on all algorithms such as Random forest, Decision tree, and XG boost algorithms and comparing the results to determine which algorithm is providing the best results for the given prediction.

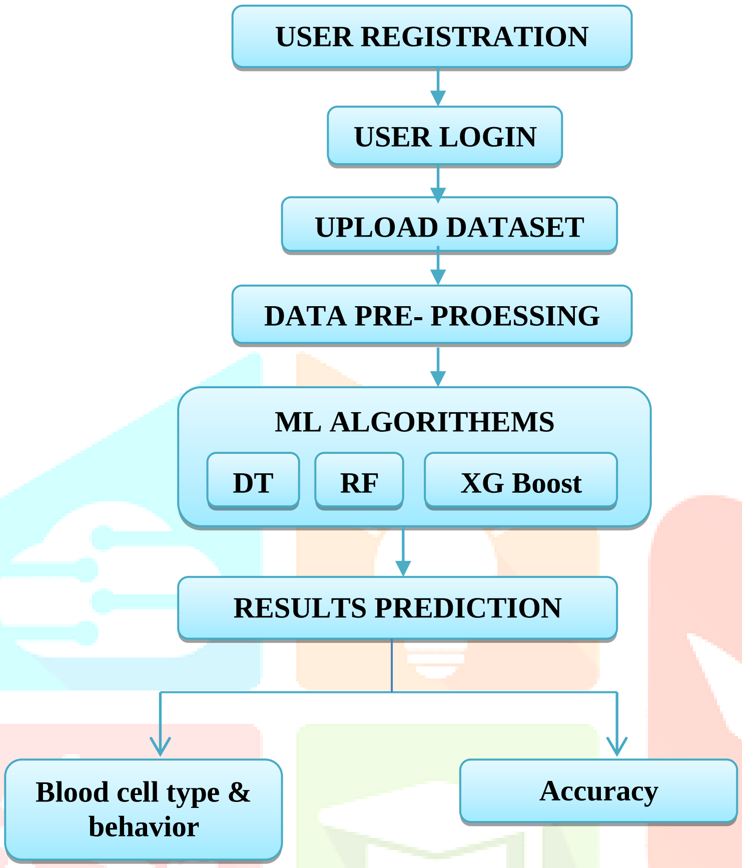


Fig1: Proposed Architecture

ADVANTAGES OF PROPOSED SYSTEM:

The following are the benefits of the proposed system. They are:

1. High accuracy
2. Low complexity
3. High efficient
4. Accurate detection

5. PROPOSED DATASET

In this proposed application we try to implement the dataset collected from KAGGLE website.

Kaggle DATASET:

This dataset contains 12,500 augmented images of blood cells (JPEG) with accompanying cell type labels (CSV). There are approximately 3,000 images for each of 4 different cell types grouped into 4 different folders (according to cell type). The cell types are Eosinophil, Lymphocyte, Monocyte, and Neutrophil. This dataset is accompanied by an additional dataset containing the original 410 images (pre-augmentation) as well as two additional subtype labels (WBC vs WBC) and also bounding boxes for each cell in each of these 410 images (JPEG + XML metadata). More specifically, the folder 'dataset-master' contains 410 images of blood cells with subtype labels and bounding boxes (JPEG + XML), while the folder 'dataset2-master' contains 2,500 augmented images as well as 4 additional subtype labels (JPEG + CSV). There are approximately 3,000 augmented images for each class of the 4 classes as compared to 88, 33, 21, and 207 images of each in folder 'dataset-master'.

- Size of the dataset: 113Mb.
- The dataset contains a total of 12,500 images.
- 12,500 images are used.
- The dataset contains 4 different cell types.
- Training data: 10,000 (80%)
- Testing data: 2,500(20%)
- Dataset is collected from the reference of:

https://www.kaggle.com/datasets/paultimothymooney/blood-cells

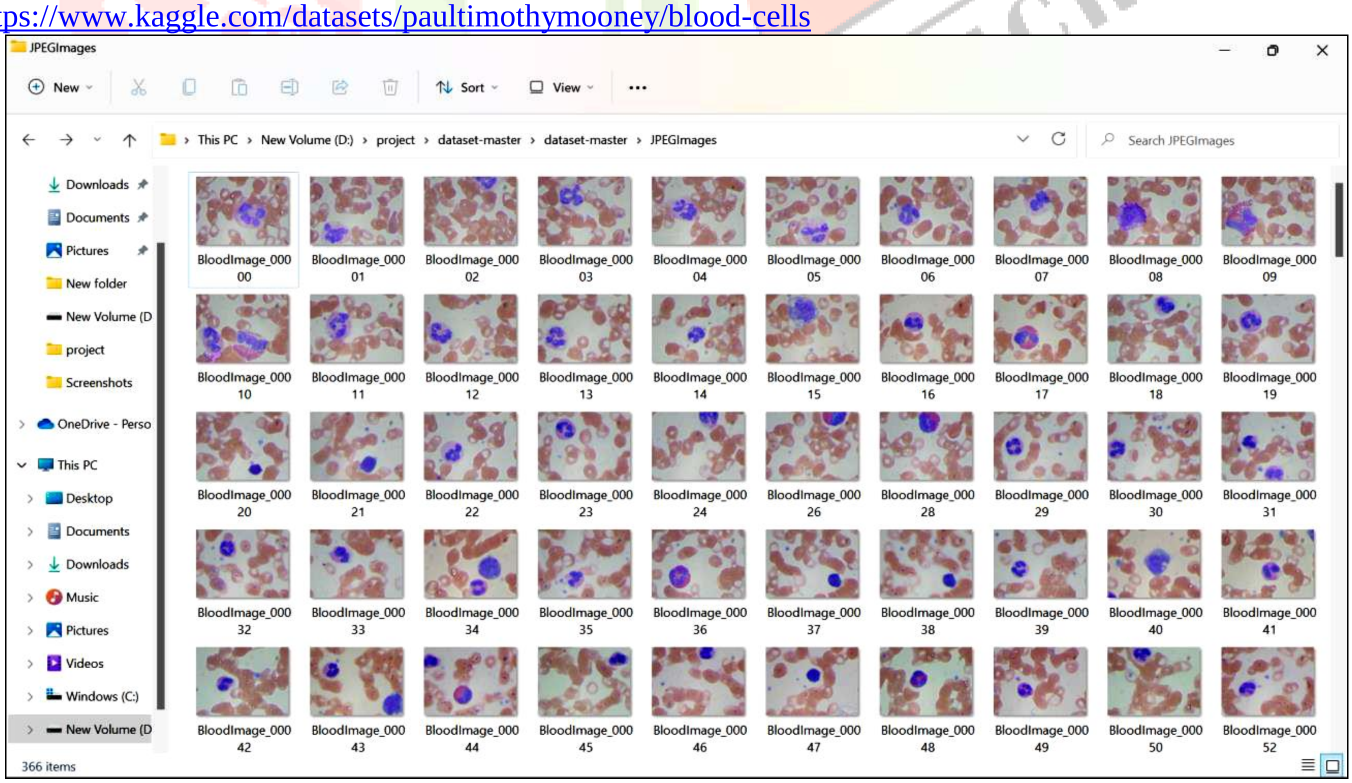


Figure 1: Kaggle white blood cells Dataset

6. IMPLEMENTATION PHASE

The step of implementation is when the theoretical design is translated into a programmatically-based approach. The application will be divided into a number of components at this point and then programmed for deployment. The front end of the application takes Google Collaboratory and as a Back-End Data base we took Kaggle /Blood Cells as input dataset. Python is being used in this instance to implement the present application. The following 4 modules make up the bulk of the application. They are listed below:

1. LOAD THE DATASET

In order to test the efficiency of our application, we try to load the input dataset from KAGGLE website and then try to use required modules as input to propose the current application.

2. DATA PRE-PROCESSING

Once the dataset is loaded now we try to apply pre-processing so that dataset can be divided into two parts: One is test and another is train phase. Now the images are pre-processed and if there are any abnormalities present inside the input dataset, they are removed and only valid dataset is kept for processing.

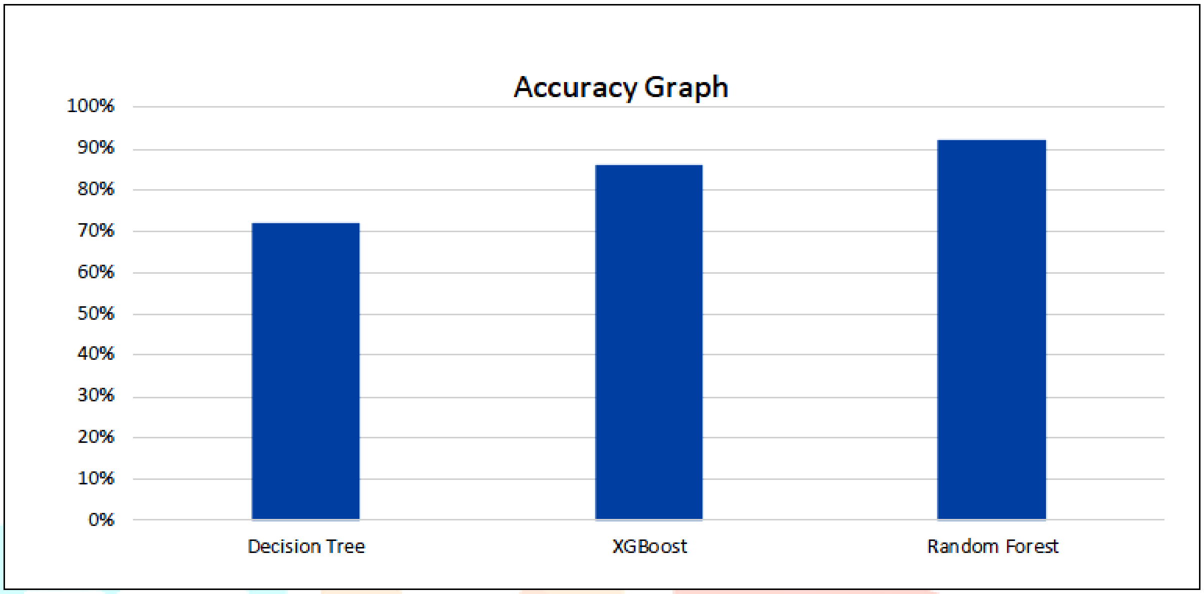
3. PREDICT THE ACCURACY

Finally the dataset is trained and now we try to test one sample blood cell image as input and check the efficiency using one of the ML algorithms, now we can see the accuracy of random forest gives more accuracy compared with other two algorithms. Here along with accuracy we will get the white blood cell type like Eosinophil, Lymphocyte, Monocyte, Neutrophil and blood cell behavior like normal or abnormal as a result.

7. EXPERIMETAL REPORTS

In this proposed application, we try to use google collab as working platform and try to show the performance of our proposed application.

1) ACCURACY OF PROPOSED MODEL USING ML ALGORITHMS



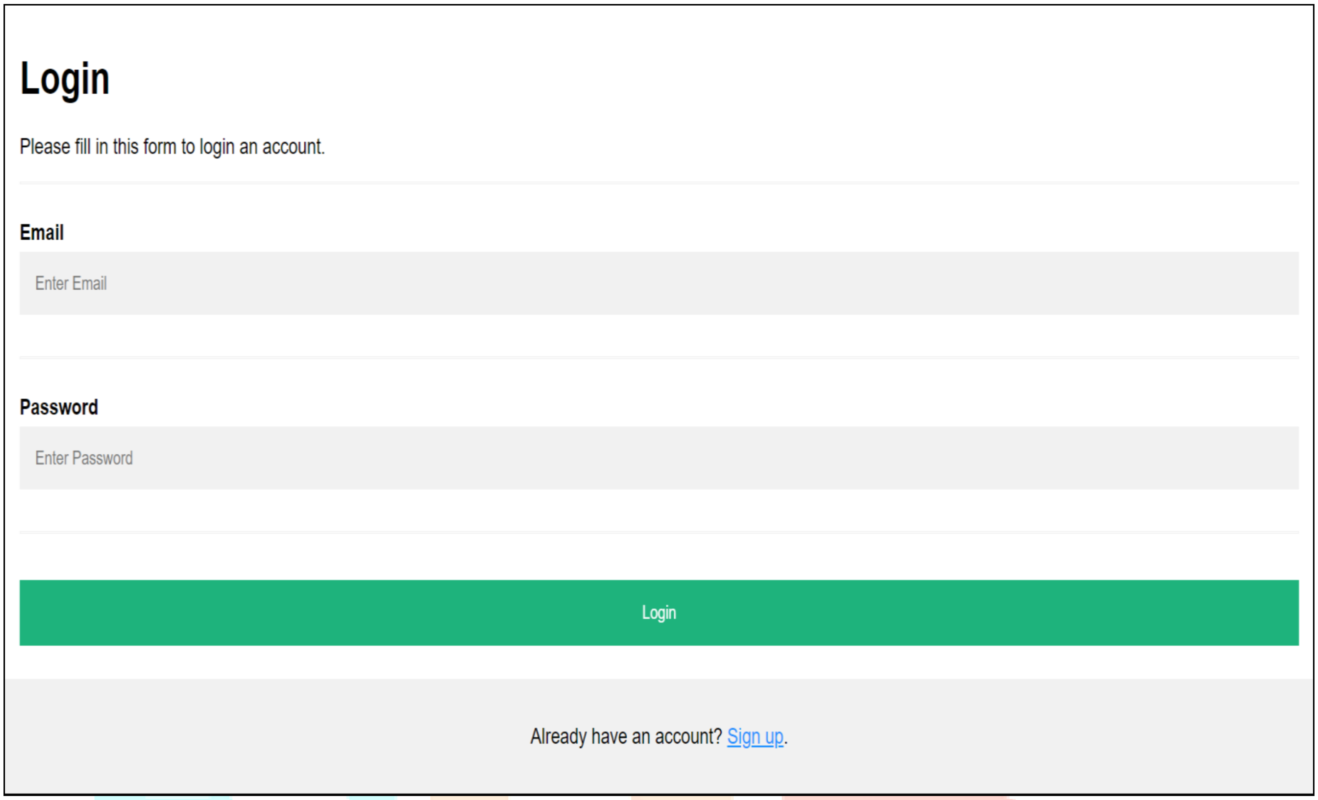
2) ACCURACY OF ALGORITHMS

```
XG Boost Accuracy: 0.864533141210375

[[   0    0    0    0    0]
 [536    6   25   74    0]
 [   3  640    8    0    0]
 [  12    2  540   13    0]
 [  55    6   19  490    0]]
```

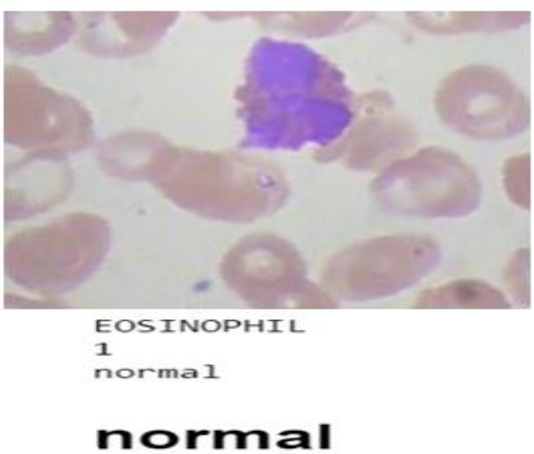
```
DecisionTree Test accuracy 0.7229312474269246
[[388  28  75 124]
 [ 23 552  22  21]
 [ 43  15 441  77]
 [133  28  84 375]]

RF Test accuracy 0.9180732811856731
[[523   3  37  52]
 [  2 609   5   2]
 [  5   1 564   6]
 [ 54   6  26 534]]
```

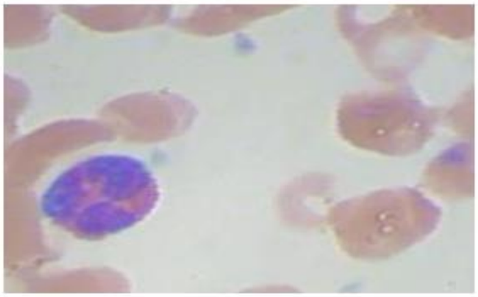
Here we can clearly see the accuracy of Random Forest is high compared with other algorithms and hence we finally conclude that RF is best among several ML algorithms.

3) USER LOGIN PAGE TO UPLOAD BLOOD SAMPLES



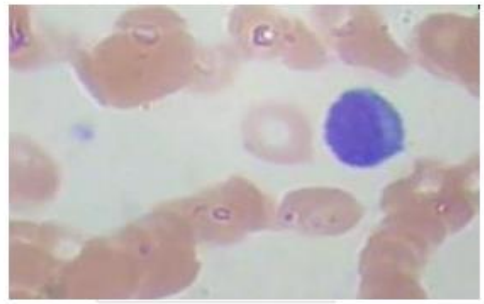
4) RESULTS OF BLOOD SAMPLES & TYPE





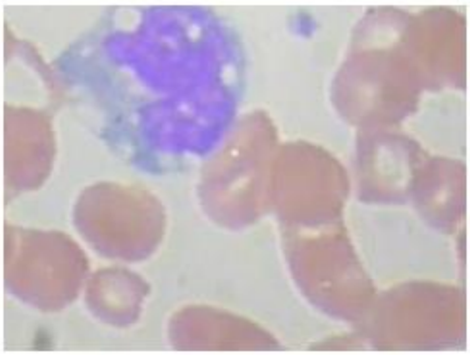
LYMPHOCYTE
2
normal

normal



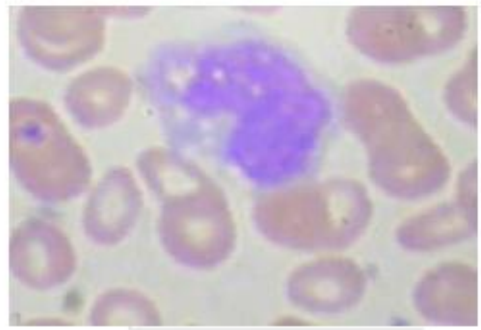
LYMPHOCYTE
2
abnormal

abnormal



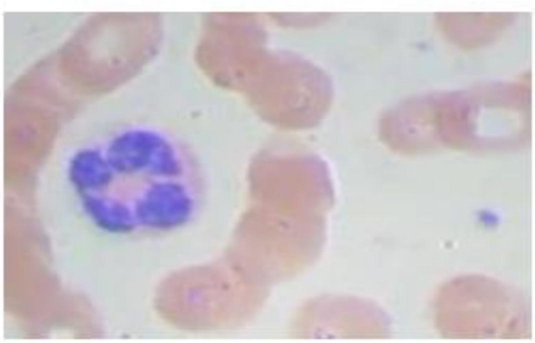
MONOCYTE
3
normal

normal



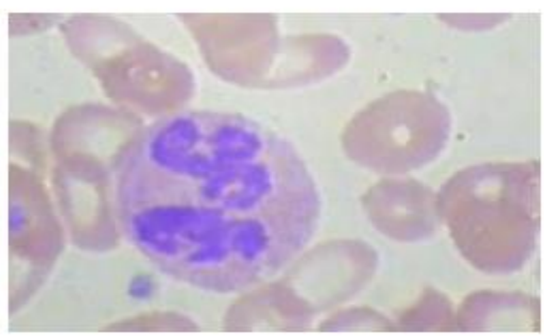
MONOCYTE
3
abnormal

abnormal



NEUTROPHIL
4
normal

normal



NEUTROPHIL
4
abnormal

abnormal

8. CONCLUSION

The system is able to classify and identify abnormal white blood cells by using three techniques like Random Forest (RF), Decision tree (DT), XG-Boost. Out of three algorithms, Random forest gives best accuracy. As we can see RF achieved approximately 91.8% accuracy ,we want to extend the same application on more number of images by taking deep learning models so that accuracy and efficiency can be increased more than RF.As a future work we want to gather large dataset as input and test the efficiency of our proposed application using deep learning models.

9. REFERENCES

[1] Virgnia Mari E. Batitis II , Merwin Jhan G. Caballes ,Abigail A. Ciudad , Miccaela D. Dnd iaz Russel D. Flores Image Classification of Abnormal Red Blood Cells Using Decision Tree Algorithm in Proceedings of the Fourth International Conference on Computing Methodologies and Communication (ICCMC 2020) IEEE Xplore Part Number:CFP20K25-ART; ISBN:978-1-7281-4889-2 .

[2] Mohammad Syahputra Et. Al (2017). Abnormal Classification on the Shape of Red Blood Cells Using Radial Basis Function Network. [2] Pooja Tukaram Dalvi and Nagaraj Vernekar (2016). Computer Aided Detection of Abnormal Red Blood Cells.

[3] Pooja Tukaram Dalvi and Nagaraj Vernekar (2016). Computer Aided Detection of Abnormal Red Blood Cells.

[4] Vishwas Sharma Et. Al (2010). Detection of Sickle Cell Anaemia and Thalassaemia Causing Abnormalities in Thin Smear of Human Blood Sample Using ImageProcessing.

[5] Ramin Soltanzadeh and Hossein Rabbani (2018). Classification of Three Types of Red Blood Cells in Peripheral Blood Smear Based on Morphology.

[6] Vishwas Sharma Et. Al (2010). Detection of Sickle Cell Anaemia and Thalassaemia Causing Abnormalities in Thin Smear of Human Blood Sample Using Image Processing.

[7] Ramin Soltanzadeh and Hossein Rabbani (2018). Classification of Three Types of Red Blood Cells in Peripheral Blood Smear Based on Morphology.

[8] Jameela Ali Akrimi ET. Al (2014). Classification Red Blood Cells Using Support Vector Machine.

[9] Pranati Rakshit, Et. Al (2013). Detection of Abnormal Findingss in Human RBC in Diagnosing G-6-P-D Deficiency Haemolytic Anaemia Using Image Processing.

[10] Ahmad Sabry Mohamad ET. Al (2017). Sickle Cell Disease Verification via Sobel Edge Algorithms for Image Processing.

[11] Krishna Kumar Jha, Et. Al (2014). Detection of abnormal blood cells on the basis of nucleus shape and counting of WBC.

[12] Nicoleta Safca, Et. Al (2018). Image Processing Techniques to Identify Red Blood Cells.

[13] Hajara Abdulkarim Aliyu, Et. Al (2018). Red Blood Cell Classification: Deep Learning Architecture Versus Support Vector Machine.

[14] S.F. Bikhet, Et. Al (2000) Segmentation and Classification of White Blood Cell

[15] Sarach Tantikitti, Et. Al (2015). Image Processing for Detection of Dengue Virus Based on White Blood Cell Classification and Decision Tree

[16] Vasundhara Acharya Et. Al (2017). Identification and Red Blood Cell Classification using Computer Aided System to Diagnose Blood Disorder.

10. ABOUT THE AUTHORS

Mrs. TULASI MIRIYALA is currently working as an Assistant Professor in Department of Computer Science and Applications at AQJ CENTRE FOR PG STUDIES, Gudilova, Anandapuram, visakhapatnam, 531173. She has more than 8 years of teaching experience. Her research interest includes Machine Learning, Network Security, Python, Data Science, and Artificial Intelligence.

Mr. GANGADHAR DOMA is currently working as an Associate Professor and Head of the Department in Department of Computer Science and Applications at AQJ CENTRE FOR PG STUDIES, Gudilova, Anandapuram, Visakhapatnam, 531173. He has more than 10 years of teaching experience. His research interest includes Data Science, Computer Networks, Machine Learning, Network Security & Cryptography, Artificial Intelligence and Python.