International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

IJIEMR Transactions, online available on 13th Apr 2022. Link :http://www.ijiemr.org/downloads.php?vol=Volume-11&issue=ISSUE-04

DOI: 10.48047/IJIEMR/V11/I04/22

Title **Image Classification of Abnormal Red Blood Cells Using Random Forest Algorithm**

Volume 11, Issue 04, Pages: 137-143

Paper Authors

A. Bhavya, A. Krishna Vamsi, A. Govardhan Reddy, G. Manoj, G. Sridevi

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Image Classification of Abnormal Red Blood Cells Using Random Forest Algorithm

A. Bhavya[1], A. Krishna Vamsi[2], A. Govardhan Reddy[3], G. Manoj[4], G. Sridevi[5]

[1]Student, Department of CSE, Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru, Andhra Pradesh, India,
[2]Student, Department of CSE, Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru, Andhra Pradesh, India,
[3]Student, Department of CSE, Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru, Andhra Pradesh, India,
[4]Student, Department of CSE, Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru, Andhra Pradesh, India,
[5]Associate Professor, Department of CSE, Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru, Andhra Pradesh, India,

E-Mail: bhavyaalokam2102@gmail.com, krishnavamsi61a@gmail.com, annapureddygovardhan@gmail.com, manutarak28@gmail.com, guttasridevi@gmail.com

Abstract:
The purpose of this study is to raise the number of aberrant red blood cells that can be recognized via image processing. To identify the data, this study used the Random Forest Technique as a machine learning model. As a result, ten aberrant red blood cells were discovered and categorized by the system. The system's pictures were captured from previous hospital patients. In addition, the slides are taken using a camera. After that, the image was loaded into the programme. The image was analyzed and classified by the system. The system's outputs, in practice, reveal the names of the aberrant red blood cells discovered in the image, as well as a soft copy of the list.

Keywords: *Random Forest Algorithm, Aberrant Red Blood Cells, Classification*

I. Introduction

Blood is the life-giving fluid that flows throughout the body, passing through the heart, arteries, and other organs. The most significant components of blood include white blood cells, red blood cells, plasma, and platelets. Red blood cell abnormalities include shape or poikilocytosis, size or anisocytosis, color, and the presence of inclusion bodies. The ability to detect irregularities in red blood cell morphologies is critical for one's health because it can determine whether or not one's blood is healthy. To classify abnormal varieties of red blood cells pathologists, hematologists, and medical technicians traditionally used a manual microscopic approach. This procedure is difficult and prone to human mistake in several ways. As a result, new technologies such as image processing are being utilized to classify aberrant red blood cells. Mohammad Syahputra and colleagues (Mohammad Syahputra et al.) claim that (2017),the forms of aberrant cells are unexpected, and Manual morphological evaluation of peripheral blood smears is time-consuming and inefficient. The amount of red blood cells detected differs from analyst to analyst due to the lack of comprehension, concentration and precision factor. A number of steps, including input image processing, pre-processing, and feature extraction, were used in conjunction with the Radial Basis Function Network. According to study, this approach has an accuracy of 83.3 percent in classifying aberrant red blood cell types.

The purpose of this project is to develop a system that can differentiate between ten different types of aberrant red blood cells and determine how accurate each categorization is. Previously, studies looked at only two to four aberrant red blood cells. As a result, proponents attempted to create the most efficient system feasible.

II. Literature Survey

1 The counting and classification of red blood cells is crucial for detecting disorders such as iron deficiency anemia, vitamin B12 deficiency anemia, and so on. As part of our research, we intend to create a solitary software that can classify red blood cells into four aberrant types: tear drop cells , macrocytes, elliptocytes, and echinocytes. The body's total number of red blood cells will also be recorded. Using thirteen geometric parameters, the red blood cells were split into four types based on their irregularity. The Decision Tree Classifier and the Artificial Neural Network are two data mining classifiers that have been studied for their usefulness in categorising red blood cells was investigated. The suggested approach detects elliptocytes with 95.27 percent accuracy, echinocytes with 96.06 percent accuracy, tear drop cells with 85.82 percent accuracy, macrocytes with 85.82 percent accuracy, and normal red blood cells with 89.76% accuracy. RBCs have a disk-shaped and biconcave form. Any change in RBC form indicates the presence of disease. The quantity of red blood cells (RBCs) is also essential in identifying anemia. The existence of blood diseases is clearly demonstrated by a decrease in RBC number and an irregularity in RBC form. Echinocytes, tear drop cells, macrocytes, and elliptocytes are all signs of disorders such myelofibrosis, severe iron shortage, uremia hereditary elliptocytosis, and hemolytic anemia, to name a few. Nearly 24.5 percent of the world's population suffers from anemia and blood disorders.

2 Blood is a disc-shaped connective tissue in which red blood cells transport oxygen. Hemoglobinopathies are a type of genetic blood condition that is a big public health issue in India. Sickle cell disease is a series of hereditary illnesses characterized by sickle hemoglobin, anaemia, acute and chronic tissue injury due to abnormally shaped red cells blocking the flow of blood. Sickle cell disease is also known as Sickle Cell Anemia. It's a condition which causes the bodies produce sickle-shaped red blood cells. The crescent shape of red blood cells is referred described as "sickle-shaped." Sickle cell anaemia is also a serious concern in the state of Chhattisgarh. It's very frequent among the scheduled caste, scheduled tribe, and other backward classes. Sahu, Mahar, Gond, Devangan, Kurmi, and Halba, for example, have the highest percentage of sickle cell disease in Chhattisgarh. This study used fractal dimension to find a way to identify sickle-shaped red blood cells in a blood film. Fractal Dimension is used to identify the shape of red blood cells and segment sickle-shaped red blood cells for shape analysis in order to calculate the percentage of sickle cell anaemia. The results demonstrate the technique's future potential, as it outperforms traditional shape recognition and analysis approaches reported in numerous literatures. Sickle cell disease (SCD), also known as sickle cell anaemia, is a life-threatening condition in which the body generates an aberrant form of haemoglobin, the oxygen-carrying protein in red blood cells.

3 In many domains of technology and medicine, microscopic image analysis is applied. Some medicines having known effects on the membranes of red blood cells (RBCs) are utilized in medical research to determine their activity. Several disorders also impact the appearance of R.B.Cs, which can lead to internal organ malfunction. On this study, morphological techniques are used to classify three well-known RBCs, Discocytes, Echinocytes and Elliptocytes in the spread of blood from the periphery. As a result, the distances between each edge pixel and the mass centre are calculated using a simple statistical analysis. Elliptocyte recognition is successful 98.63 percent of the time, normal Discocyte recognition is 96.7 percent of the time, and Echinocyte recognition is 95.36 percent of the time. Microscopy image analysis is useful in a wide range of fields, including medicine and technology blood smear using morphological approaches in this research. As a result, the distances between each edge pixel and the mass centre are calculated using a simple data analysis. Normal identification of eliptocytes is 98.63% successful. Recognizing Discocytesis 96.7 percent successful, and Echinocyte recognition is 95.36 percent successful. Microscopy image analysis is beneficial in a variety of disciplines, including technology and medicine. The small size of R.B.Cs, poor resolution, and shape variety in

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

blood smears make it challenging to detect and classify various cells [1]. There have been numerous attempts to define separate R.B.Cs in blood smears. In 1995, U. Ebel et al. trained a three-layer network for categorising RBCs using a neural network based on a fuzzy classifier [2].JW Bacus and JH Weens employed a digital image processing method in 1977 that provided various features relating to red cell morphology and internal core pallor configurations [3]. J. Dahmen et al. [4] classified Discocyte, Echinocyte, and Stomatocyte in a set of 128128 grayscale pictures by using Gaussian mixture density (GMD) technique in 2000.They were able to come up with a solution that had a 15% error rate [4].One of the most difficult tasks in these efforts is extracting characteristics that are resistant to rotation and scaling. The variance of Euclidean distances between edge pixels and cell mass centres is calculated and used as a feature that is not affected by rotation or scaling throughout this article. There are four sections to this study.The fourth section is on feature extraction. In section 5, the findings are reported. To see how well our strategy works, we put it to the test on a batch of about 200 colored microscopic pictures. This paper comes to a close with Section 6.2. PROCESSING IN ADVANCE It is not required to provide colour photographs because R.B.Cs are categorised based on their look.

III. Proposed Solution

The proponents developed a technique that uses the Random Forest Algorithm to automatically classify ten(10) aberrant red blood cells: Codocyte ,ovalocyte, degmacyte, dacrocytes, echinocyte, spherocyte, stomatocyte, elliptocyte, drepanocyte and acanthocyte are some of the different types of cells as shown in Fig.1.The Decision Tree Algorithm was employed in previous techniques to detect aberrant red cells. However, the performance of this technique is less. To improve the accuracy of the system, the author of this study employs a variety of machine learning algorithms.

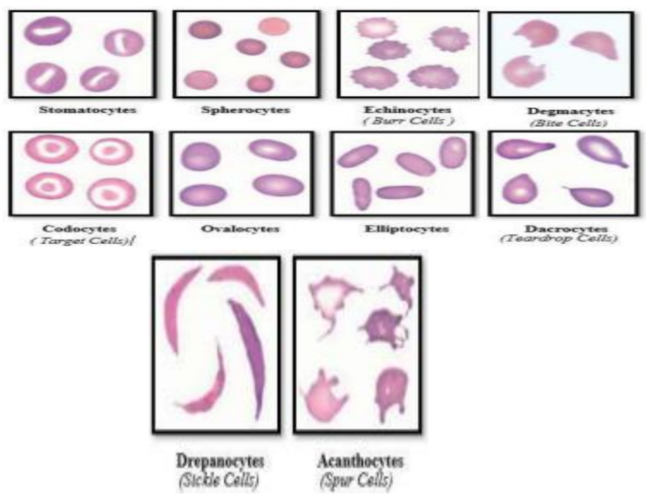

Fig-1: Abnormal Red Blood Cells

IV. Implementation

To implement this project author has used 5 different types of machine learning algorithms such as SVM-Linear, Random Forest, Decision Tree, Adaboost, and Bagging. This project consists of following modules.

1) Upload Dataset: The dataset of aberrant red blood cells is uploaded to the application using this module.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

2) Preprocess Dataset: Dataset often contains missing and null values and non-numeric values and using this module we will replace all such values with 0. This module will divide the dataset into two parts: training and testing. 80% of the dataset will be used to train machine learning techniques, while 20% of the dataset will be used to assess the accuracy of machine learning techniques.
3) Run Algorithms: using above dataset we will train all 5 machine learning algorithms and then calculate the accuracy for different machine learning algorithms.
4) Upload test data &predict the abnormal blood cell: We will submit test results into this module, and the application will determine whether the name of the aberrant red blood cell is correct.

4.2 Dataset Details

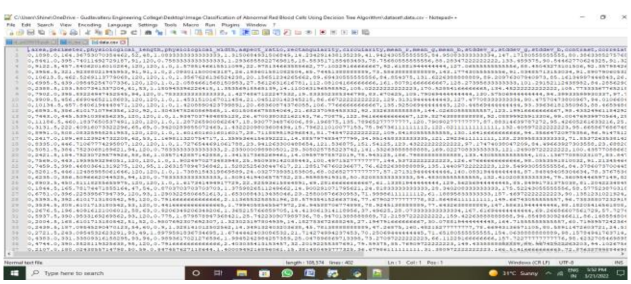

Fig-2: In the above dataset screen first record contains column names and the other records contains column values

5) Performance Graph: Using this module we will plot performance graph between all algorithms.

4.1 Random Forest

Random forest is a popular supervised learning technique for problems like classification and regression. It builds decision trees from numerous samples using a majority of votes for classifying and a mean of regression.

A Random Forest, in essence, creates a random sample of multiple decision trees and then combines them using cross validation to get a more stable and accurate prediction. While constructing random trees, it separates into unique nodes or subgroups. Then it looks for the best result among the random subsets. As a result, the model of the algorithm improves. In a random forest, just the random subset is taken into account.

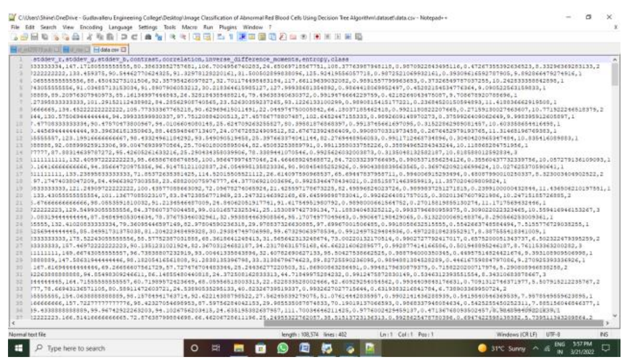

Fig-3: Each record in the above screen has 402 column values, with class being used to identify aberrant blood cells in the last column. We will use the following dataset to train all 5 machine learning algorithms.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

V. Results and Discussion

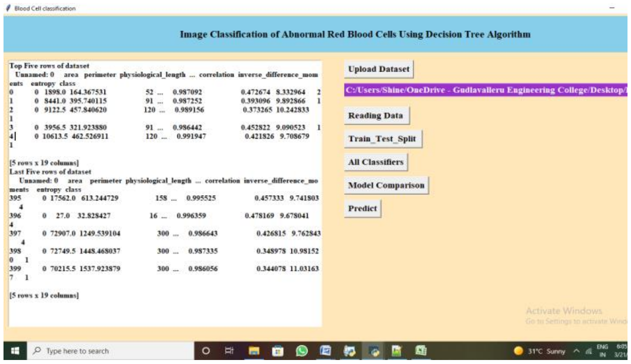

Fig-4: In the above screen dataset is loaded and to preprocess the data we have to click on reading data

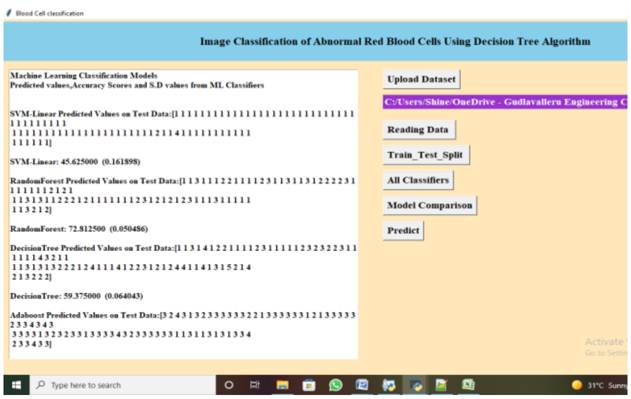

Fig-5: In the above screen the I clicked on All Classifiers button and then we got the accuracy values of each algorithm and in all algorithms Random Forest is giving better performance result.

VI. Conclusion

Defective cells of the blood can be classified using the Random Forest classification approach. Using the Random Forest Algorithm, the system had the ability to categorise aberrant cells of the blood based on data obtained from 40 images with 600 sample cells. Classification

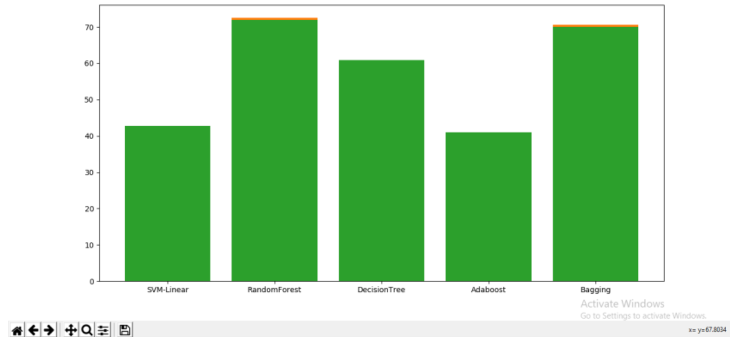

Fig-6: The x-axis in the above graph shows the algorithm name, while the y-axis reflects the accuracy numbers for each algorithm. Based on the above graph, we can conclude that Random Forest produces the best results.

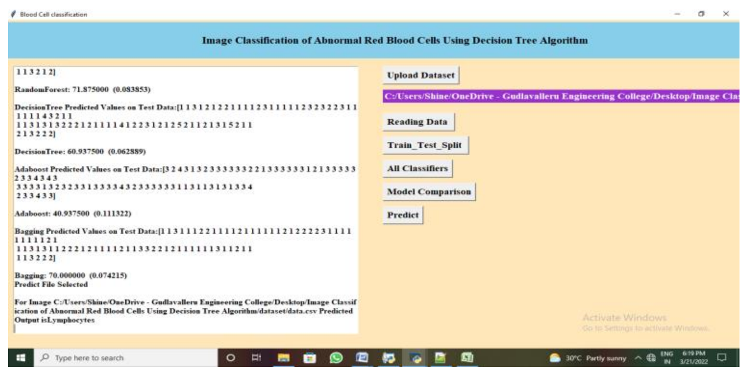

Fig-7: We can submit the dataset to the Predict button in the above interface, and the ML system will predict the defective red blood cell

errors were caused by minor differences in the attributes employed. The mean efficiency rate is 89.31 percent. The mean error margin was 10.69 percent, with node H being the source of the majority of the problems. Because the irregularity of the codocyte's core pallor was poorly identified and noticed, the categorization of aberrant red blood cells was inaccurate. The poor quality of hospital blood slides was frequently the cause of the detection error rate. The approach was also proven to be helpful in classifying abnormal red blood cells.

References

1. Pooja Tukaram Dalvi and Nagaraj Vernekar, "Computer Aided Detection of Abnormal Red Blood Cells," IEEE Int. Conf. Recent Trends Electron. Inf. Commun. Technol., pp. 1741–1746, 2016.
2. Vishwas Sharma Et. al. Detection of Sickle Cell Anaemia and Thalassaemia Causing Abnormalities in Thin Smear of Human Blood Sample Using Image Processing 2010.
3. Hossein Rabbani and RaminSoltanzadeh (2018). Red blood cells in a peripheral blood smear are divided into three categories based on their shape.
4. H. A. Aliyu, "Detection of Accurate Segmentation in Blood Cells Count – A Review," Int. J. Sci. Dev. Res., vol. 2, no. 8, pp. 28–32, 2017.
5. S. C. Webster, J. G. and Cazzanti, BioinstrumentationHematology. John Wiley and Sons, 2004.
6. Mohammad Syahputraet. al, Abnormal Classification on the Shape of Red Blood Cells Using Radial Basis Function Network 2017.
7. M. Sheikh, H., Zhu, B. & Tzanakou, "Blood cell identification using neural networks.," Twenty-Second Annu. Northeast Bioeng. IEEE Conf. Proc., pp. 119– 120, 1996.
8. J. Bronkorsta, P.J.H., Reinders, M.J.T., Hendriks, E.A., Grimbergen and G. . Heethaar, R.M. & Brakenhoff, "On-line detection of red blood cell shape using deformable templates," Pattern Recognit. Lett. 21, pp. 413– 424, 2000.
9. J. P. and C. Neatpisarnvanit, "Red Blood Cell Extraction and Counting," 3rd Int. Symp. Biomed. Eng., 2008.
10. H. A. Elsalamony, "Healthy and unhealthy red blood cell detection in human blood smears using neural networks," Micron, vol. 83, no. 4, pp. 32–41, 2016.
11. D. K. Das, C. Chakraborty, B. Mitra, A. K. Maiti, and A. K. Ray, "Quantitative microscopy approach for shape-based erythrocytes characterization in anaemia," J. Microsc., vol. 249, no. 2, pp. 136–149, 2013. [10] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection," Front. Plant Sci., vol. 7, no. September, pp. 1–10, 2016.
12. I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," J. Mach. Learn. Res., vol. 3, no. 3, pp. 1157–1182, 2003.
13. I. G. and A. Elisseeff, "An introduction to variable and feature selection," J. Mach. Learn. Res. 3, pp. 1157– 1182, 2003.