Classification of Abnormal Red Blood Cells Using Deep Learning

Prof. Usha Kumari V
Department of Information Science and Engineering
Acharya Institute of Technology
Bangalore, India
ushakumari2487@acharya.ac.in

Nandish M S
Department of Information Science and Engineering
Acharya Institute of Technology
Bangalore, India
nandishs.19.beis@acharya.ac.in

Nishchitha J
Department of Information Science and Engineering
Acharya Institute of Technology
Bangalore, India
nishachithaj.19.beis@acharya.ac.in

Nischitha
Department of Information Science and Engineering
Acharya Institute of Technology
Bangalore, India
nishchithan.19.beis@acharya.ac.in

Vijay K V
Department of Information Science and Engineering
Acharya Institute of Technology
Bangalore, India
vijayv.19.beis@acharya.ac.in

Abstract **- The most common and dangerous abnormality in red blood cells (RBCs) is shape aberration. The primary techniques for identifying and validating an anemic stage (shape abnormality) are measuring hemoglobin levels or doing a manual microscopic examination of peripheral blood smears. Under a microscope, manually detecting abnormal cells can be difficult and time-consuming. This leads to subpar outcomes, unnecessary medication that shortens patients' lives, and technician eye fatigue. This program aims to increase the detection of abnormal red blood cells through image processing. The Decision-Tree technique is used as a machine-learning categorization technique. Ten abnormal red blood cells were consequently discovered and identified by the system. Images used in the system had already been collected from hospital patients. The machine altered and classed the image. Results show the names of the abnormal red blood cells that the system identified in the image.**

Keywords **– Decision Tree Algorithm, DeepLearnig, Inception V3 model.**

I. INTRODUCTION

Blood is a crucial fluid that circulates through the heart, arteries, and organs of the human body. It comprises many components, including red and

white blood cells, platelets, and plasma. Any abnormalities in red blood cells, such as changes in size, shape, color, or the presence of inclusion bodies, can indicate poor health. Medical professionals traditionally use manual microscopic methods to identify abnormal shapes of red blood cells, but this method is prone to human error. Consequently, high-tech image processing is used to classify such cells.

According to Mohammad Syahputra Et. Al (2017), the manual morphological examination of peripheral blood smears is less efficient, and the precision and knowledge of each analyst can cause inconsistencies in detecting abnormal red blood cells. To improve the accuracy and reliability of classification, a system that uses Radial Basis Function Network and several stages (input image, pre-processing, and feature extraction) has been developed. This system can classify ten abnormal red blood cells with an 83.3% accuracy rate.

The life-sustaining substance that circulates throughout the entire body, including the heart, arteries, and other organs, is called blood.

Additionally, blood has a variety of components, primarily red blood cells, white blood cells, platelets, and plasma. Red blood cell abnormalities might range in size or anisocytosis, shape or poikilocytosis, color, or even the presence of inclusion bodies. One's health needs to spot these variations in red blood cell shape since it indicates whether or not the blood is healthy. Medical technicians, pathologists, and hematopathologists typically utilize a manual microscopic approach to categorize red blood cells with aberrant forms.

The proponents of this study proposed a maximized system that can automatically classify ten abnormal red blood cells (spherocyte, codocyte, stomatocyte, ovalocyte, elliptocyte, degmacyte, drepanocyte, dacrocytes, acanthocyte, and echinocyte) using the Decision-Tree Algorithm. This algorithm is commonly used in classification and regression analysis, as it divides a dataset into smaller datasets based on descriptive features or attributes until it reaches a specific abnormal red blood cell. This system aims to enhance the reliability and accuracy of identifying abnormal red blood cells.

II. LITERATURE SURVEY

Different types of study and research work have been carried out in earlier days a different time.Blood is the life-maintaining fluidthat flows through the heart, arteries, and other organs and it circulates through the wholebody. Abnormalities of red blood cells vary through size or anisocytosis, through shape or poikilocytosis, in color and even through the presence of inclusion bodies. Detecting 49these irregularities in the shapes of the red blood cells is significant to one's health as itcan determine whether the blood is healthy or not. Medical technicians, Pathologists and Hematologists usually used a manual microscopic method, to classify abnormal shapes of red blood cells. This methodology somehow is difficult and prone to human error. Thus, classifying the abnormal red blood cells using image processing is created using the high technologies. The objective of this study is to create a system that can classify 10abnormal red blood cells and to know the reliability rate of classification of each abnormal red blood cells. Previous studies are usually limited to two to four abnormal red blood cells. Thus, the proponents aimed to create a maximized system.

III. METHODOLOGY

A. Conceptual Framework.

Figure 1 illustrates how the system works step by step towards its objective to create a system that will classify abnormal red blood cells.

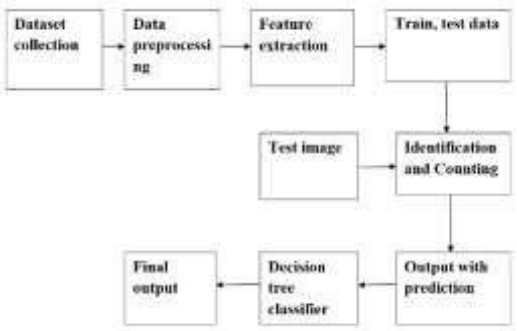


Fig 1. Block Diagram

Collect a dataset of red blood cell images with both normal and abnormal cells.Clean and preprocess the data by removing noise, normalizing the color channels, and resizing the images.Extract relevant features from the preprocessed images, such as shape, size, and texture.Divide the dataset into training and testing sets, with a larger portion allocated for training.Use the training dataset to construct a decision tree classifier that can identify normal and abnormal red blood cells based on their extracted features.Use the trained decision tree classifier to identify and count the number of abnormal red blood cells in test images.Input test images of red blood cells to the system.The decision tree classifier will predict whether each red blood cell in the test image is normal or abnormal.Output the final prediction, which includes the number of normal and abnormal red blood cells in the test image.

B. Classification of abnormal red blood cells.

1). Data Collection

The images used is RBC marphology collected from various sources such as books, journals and websites.Acquiring microscopic images of different types of red blood cells (RBCs) from various sources is a crucial step in research related to hematology and pathology. The process of image acquisition involves the use of a microscope and a digital camera to capture images of RBCs in different states and from different sources. The following is an explanation of the process of image acquisition of microscopic images of RBCs.

To capture images of RBCs, the prepared sample is placed under the microscope and focused on the RBCs using the high-resolution objective lens. The digital camera is then used to capture the images of the RBCs, and the software provided by the camera manufacturer is used to control the camera settings, such as exposure time, gain, and image resolution. The images are typically captured in the RGB format, with a resolution of at least 1024 x 768 pixels. It is essential to adjust the camera settings to

ensure that the RBCs are properly illuminated and that the images are in focus. Multiple images are captured at different magnifications and orientations to ensure that the RBCs are adequately represented in the study.

2). Data Preprocessing

The picture is initially transformed into grayscale, standard size and normalized . The image must be grayscaled to reduce the amount of data that must be sent for each pixel. It penetrates the edge. detection to quickly ascertain the object's form. The system makes use of edge detection techniques like Canny edge detection. The Gaussian filter is used by Canny Edge to smooth out the picture and eliminate noise. By identifying pixel locations where the gradient is greater than its neighbours, edges are found. Additionally, each aberrant red blood cell's edge is detected using edge detection, which also renders the picture in black and white. The backdrop is changed to black, while the edges are changed to white.

Fig 2. Acquired image(left), countered image(Right)

3). Feature Extraction

The pre-processed images are then input into the Inception V3 model for feature extraction.The pre-processed picture is utilised to extract features using the pre-trained Inception V3 model. In order to do this, the picture must be sent through the model's several convolutional layers, which combine the image with a variety of learnt filters to create a collection of feature maps. These feature maps illustrate various spatial sizes as well as abstract ideas like edges, corners, and textures.After passing through a pooling layer, which aggregates the values within each spatial region, the feature maps are often processed to lower their dimensionality. As a result, the feature representation may become even more straightforward, and future analyses may become less computationally complicated. The features extracted from the Inception V3 model are then passed to the decision tree algorithm for classification and training.

4). Model Training

Model training, where a machine learning model is trained on a dataset to learn patterns and make predictions on fresh data, is an important phase in machine learning. A well-known machine learning technique called the decision tree algorithm creates a tree-like model of decisions depending on the characteristics of the training data. During training, the algorithm divides the data recursively into smaller subgroups depending on the values of the features, and then labels each subset according to the dominant class. Up until every piece of data has been allocated to a leaf node, or class label, the training process continues. The decision tree technique contains a number of hyperparameters that may be adjusted to enhance the model's performance. The maximum depth of the tree, which establishes the maximum number of levels in the tree, is one of the most crucial hyperparameters. However, a deeper tree may overfit to the training data and perform poorly on fresh data. A deeper tree can capture more complicated relationships in the data.

A typical strategy is to run a grid search or random search over a variety of hyperparameters to optimise the decision tree algorithm's hyperparameters. Multiple decision tree models are trained using various hyperparameters, and the effectiveness of each model is assessed using validation data. The validation set, which is a subset of the dataset not used for training, is used to assess how well the model performs when faced with fresh data. Metrics that quantify the trade-off between several sorts of mistakes, including as accuracy, precision, recall, and F1 score, are frequently used to assess the model's performance.By traversing the tree depending on the values of the features, the decision tree model may be used to generate predictions on fresh data after it has been trained and optimised. The leaf node that the data sample has reached is subsequently used to determine the projected class label.

5).Model Testing

To provide a final classification result, the predictions must be post-processed after the model has produced predictions for each new picture. This may entail employing a multi-class classifier to differentiate between several sorts of anomalies, or thresholding the projected probability to provide binary labels (normal or abnormal).

Building the tree structure as the first step in the decision tree method involves choosing a characteristic to divide the data up depending on. Using a metric like information gain or Gini impurity, the process of feature selection entails assessing each feature's ability to divide the data into distinct subsets. The dividing feature is determined by the feature with the greatest score. Once the tree has been constructed, each subset of the data is given a label based on the majority class of the samples in the subset and relates to a leaf node of the tree. This label assignment procedure is based on a classification rule, such as "assign label A to the subset if class A is the majority class."Based on the values of the input characteristics, the decision tree algorithm traverses the tree from the root to a leaf

node and uses the splitting criteria to decide which branch to take at each node in order to generate predictions on fresh data. The anticipated label for the incoming data corresponds to the leaf node that the traversal reaches.The data is then divided into one or more subgroups based on the values of the specified characteristic. The process of choosing a feature and splitting the data is repeated for each subset, which represents a branch of the tree. This process continues until a stopping requirement, such as reaching a maximum depth, is satisfied.

Finally, measures like accuracy, precision, recall, and F1 score are used to assess how well the model performed on the new pictures. These metrics offer a numerical assessment of the model's capacity to detect anomalies in red blood cell pictures and generalise to fresh data.

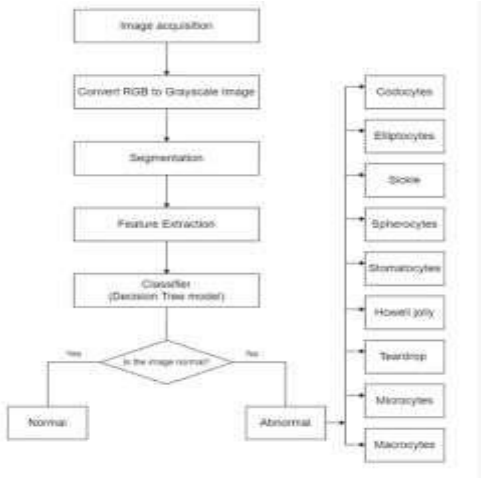


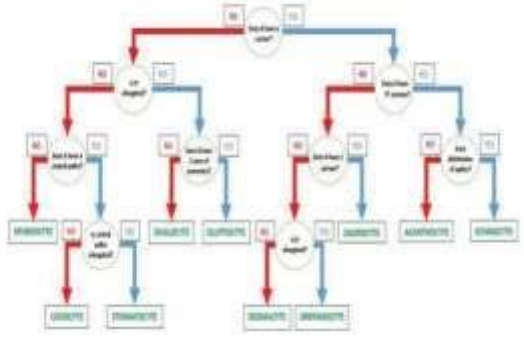
Fig 3. Classification of abnormal red blood cells

IV. RESULTS AND DISCUSSIONS

A. Decision Tree

The determined and the highest node level with regard to the amount of aberrant red blood cells in order to categorise the ten (10) different categories of abnormal red blood cells using the Decision-Tree Algorithm.

Fig 4. Maximized Decision Tree



$$m < 2^n$$

Where: $n = maximum\ node\ level$

$m = number\ of\ abnormal\ red\ blood\ cells\ need\ to\ classify$

For the computation of Node level:

$$10 < 2^n$$

$$\log 10 = \log 2^n$$

$$\log 10 = n \log 2$$

$$n = \frac{\log 10}{\log 2} = 3.32 \sim 4$$

Therefore, $10 < 2^4$

$$\mathbf{10 < 16}$$

As 10<16 is true , the condition has been satisfied.

B. Results

Number of ARBC	Attributes					
	Corners	Central Pallor	Elongated Central Pallor	Axis of Symmetry	Elongation	Spikes
Spherocyte	0	No	No	No	No	No
Codocyte	0	Yes	No	No	No	No
Stomatocyte	0	Yes	Yes	No	No	No
Ovalocyte	0	Yes	No	No	Yes	No
Elliptocyte	0	Yes	No	Yes	Yes	No
Degmacyte	2	Yes	No	No	No	No
Dreparocyte	2	No	No	No	No	No
Dacrocyte	1	Yes	No	No	No	No
Acanthocyte	Many (No definite count)	Yes	No	No	No	Yes
Echinocyte	Many (No definite count)	Yes	No	No	No	Yes

Table 1. Attributes of abnormal red blood cells.

These are the attributes i.e,table 1 is used to fed into the decision tree for the classification.

Fig 5. Predicted type of abnormal red blood cell.

As shown in the figure 5 , here we have loaded macrocyte rbc image to the system and it is predicted.

B. Evaluation Criteria.

We used the confusion matrix for the evaluation of the designed algorithm. Along with that precision, recall, F1 score, sensitivity, and selectivity were also considered.

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative}$$

$$F1 = \frac{Precision * Recall}{Precision + Recall}$$

$$Sensitivity = \frac{TruePositive}{TruePositive + FalseNegative}$$

$$Specificity = \frac{TrueNegative}{TrueNegative + FalsePositive}$$

V. CONCLUSION AND FUTURE ENHANCEMENT

Classifying RBCs into multiple categories and diagnosing diseases based on them remains a challenging task that typically requires an expert pathologist. To address this issue, we have developed a system that leverages both image processing and Decision tree algorithm to accurately detect normal and abnormal RBCs with a 98% accuracy rate. We have also been able to classify different types of abnormal RBCs into ten distinct categories with an accuracy rate of 98.6%. To accomplish this, we utilized RBC cell images in conjunction with additional features such as shape, size, and texture of RBCs. In our ongoing research, we plan to further refine this approach by developing a deep learning method capable of classifying RBCs from microscopic images without any preprocessing. Ultimately, this approach could prove useful not only for RBC classification and disease diagnosis but for any multiclass segregation problem.By giving medical professionals a precise and effective way to spot potential patient health issues, the use of machine learning algorithms for the classification of abnormal red blood cells has the potential to improve healthcare outcomes. The classification of red blood cells as normal or abnormal can be accomplished by gathering a dataset of red blood cell images, preprocessing the images, and extracting features. A decision tree model can then be trained and tested for its effectiveness in this task. Once trained, the classifier can be used to classify fresh red blood cell images and deliver precise classification outcomes.

Future enchancement:

Augmenting the training dataset with more data can make it easier for the model to generalize to previously unexplored data sets.

Advanced feature extraction means Experimenting with more complicated and subtle properties of red blood cells that are challenging to detect with conventional feature extraction techniques, such as deep learning or transfer learning, can be helpful.

Combining several classifiers, such as decision trees and SVMs, can increase classification accuracy and decrease overfitting. This is known as ensemble learning.

Creating a decision tree model that can explain its predictions in a way that is comprehensible to both medical professionals and patients can help to increase trust and comprehension.

Real-time classification by creating a system for categorizing red blood cell pictures in real time.

REFERENCES

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

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

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

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

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

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

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

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

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

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