

Computerization Method to classifying of Red Blood Cells using Boosting Technique

Damini Uike

Electronics Engineering Department,
Government College of Engineering Amravati,
Amravati, India

Snehal Thorat

Electronics Engineering Department,
Government College of Engineering Amravati,
Amravati, India

Abstract: Red blood cells are the rounded and flexible, keep oxygenation to the human body. Along with red blood cells other abnormal cells like sickle cells, elongated cells and Echinocytes present in the body. In present microscopic image are containing touching and overlapping cells, which have to segment properly without any error then proposed automated methodology is established. To identify shape of blood cells and classify in different class according to shape of cells using image processing technique and machine learning approach. XG-boost is the classifier used for classification of the cells in multiclass. Firstly, performed binary classification of cells and multiclass classification of abnormal cells are executed. It is performed in both analysis single as well as multilevel analysis. So testing accuracy for Extreme Gradient Boosting 98% achieved.

Keywords: RBC, Sickle cell, Echinocytes, machine learning, XG-boost

I. INTRODUCTION

The red blood cells (RBC) are highly specific, well adapted for their primary function of transporting oxygen from the lungs to all of the body tissues. This is flexible and round in shape. In Some portion of human other abnormal cells are present e.g. sickle cells, elongated cells and Echinocytes[10]. Sickle cell anaemia is a serious and frequently fatal disease characterized by a genetic defect of haemoglobin. This disease primarily affects RBC's where the RBC will be Sickled in shape or crescent shaped. Sickle cells are stick to the vessels due to lack of flexibility, blockage of blood and experience painful episode of pain called crises. Elliptocytes are also called ovalocytes is an hereditary disorders, this abnormal cells present in patient rather than normal cells such as infectious anemias, thalassemia, iron deficiency anaemia and in newborn babies. Echninocytes are burr-like erythrocyte with narrow, uniform, spike like surface projections The new reagent for sickle test has been adjusted and found optima concentration. However when applied to fresh red blood cell, immediately the RBC transform to echinocytes. A conditions Associated with Echinocytes causes Pyruvate Kinase Deficiency, Uremia, Microangiopathic Hemolytic Anaemia. This type of anaemia happens when red blood cells are destroyed and loss elasticity by an abnormal process in our body before their lifespan is over. As a result, our body doesn't have sufficient RBC to function. The average red cells are lives 100-120 days and sickle cells in human lives 10-20 days. Anaemia is a serious global public health problem that affects approximately 25% of the world's population, particularly west sub-Saharan Africa, Asia and some region in India. There is no cure to prevent from anaemia but early diagnosis is very important. Thus, the simple blood test is a tiresome and erroneous work which should be replaced by an

effective, progressive and correct tool to successfully diagnose the sickle cell disease. One such tool can be prepared by using Image Processing Technique and machine learning algorithm. This tool has to run on open-cv python [2]support to the entire operating system such as Windows, Linux, macOS, FreeBSD, NetBSD etc. It is faster and effective than Matlab. Open-CV is a computer vision library and open source software.

II. LITURETURE SURVEY

Recently, researcher used image processing techniques to identify blood abnormalities by automatic methods. P. Rakshit and K. Bhowmik[3] presented Weiner filter and Sobel Edge detection method to find the edge of the blood cells. Its required small execution time and computationally efficient but not concentrated to segment overlapping cells. Aruna N.S. and Hariharan S.[4] employed different Edge detectors like Canny, Sobel, Roberts, Prewitt and LoG detectors are used to find out the best detector for automatic diagnosis. In this result found that Canny Edge Detection method is more superior for the diagnosis as it gives more details of the original image. Hala Algailani and Musab Elkheir Salih Hamad [5] watershed segmentation used to properly classify the overlapping cells but problem occurred with over segmentation affect quality of detection that reason use nonlocal means filter. Chy and Rahaman[6] uses threshold segmentation to segment RBC containing images but not beneficial for overlapping cells and SVM classifier efficiently trained to testing purpose.

In this studied, some quit problem of segmentation technique to segment microscopic images then proposed system used segmentation and XG-Boost classifier for the classification of RBC images into various categories.

III. PROPOSED SYSTEM

This section describes the workflow of the research and the machine learning algorithms used in this study. This section also explains how the dataset was generated, pre-processed, and trained and tested. "Fig. 1" shows the entire process used to build the machine learning models in this research. The workflow divided into two phase training and testing phase to build machine learning model. The data are split into training and testing phase, testing phase accounting for 60% of the total data. The remaining 40% are used as the test set. In training phase step by step procedure follows for overall system performance to predict accurate output. Same as like in testing phase, feature extracted from test data to given for the prediction. Feature construction transformed the raw

input into meaningful forms, adding the nonlinearity and introducing the physics of the flow into the machine learning model.

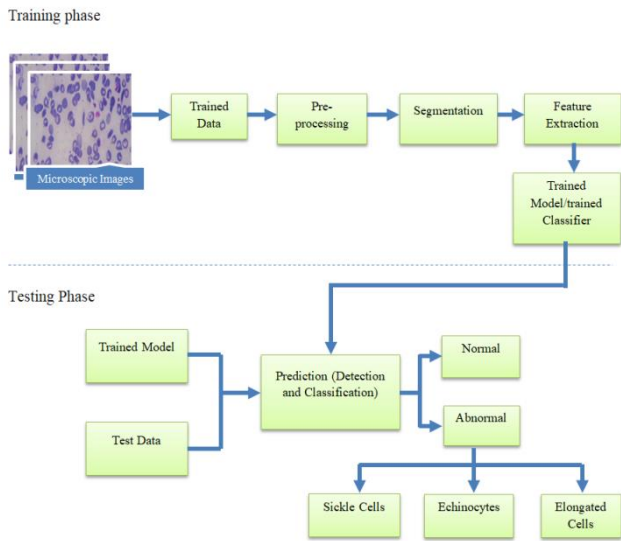


Fig. 1. Proposed Computerized Model

- A. **Database:-** The database present in the study is microscopic image in the Jpg. or Png. format. The dataset collected from the Dr. Panjabrao Deshmukh Memorial Medical College Amravati. The image is first acquired through camera connected to the microscope. The image is captured in the jpg and png form. Another source to collect dataset related RBC images from internet source. ErythrocytesIDB is a standard database [12], which is available at [//erythrocytesidb.uib.es/](http://erythrocytesidb.uib.es/). The images containing dataset in RGB format given to the pre-processing phase for correct detection.
- B. **Pre-processing:-** To prepare the input images according to the standard input of the proposed system, pre-processing is applied on the dataset. The aim of pre-processing is an improvement of the microscopic image data that suppresses unwanted distortions or enhances some image features important for further processing. In pre-processing certain operation performed resize image, remove noise and destroyed unwanted spot or holes to gives wrong meaning. This is beneficial for accurate detection and classification of red blood cells. The acquired image in RGB form and converted into grey. Median filter used to remove noise, find all connected component and clear the border.
- C. **Segmentation:-** Image Segmentation techniques make a MASSIVE impact here. The anaemia blood smear will be segmented by appropriate segmentation methods. Segmentation of images required five main techniques such as pixel intensity, threshold, boundary detection, region-based processing and morphological methods. Here Morphological based segmentation operations are applied on binary images. Closing and opening morphological operation perform by using Erosion and Dilation operator.

D. **Feature Extraction:-** In this step, extracted shape and textural feature from the images. On this basis extract relevant information present in the image.

1. Shape feature

In the shape feature, perimeter, circularity factor, deviation factor, form factor, aspect ratio, equivalent diameter, diameter, major axis and minor axis are computed.

- Area is actual number of pixels in the region. This variation is used to find out abnormalities associated with changes in red cell dimensions.
- Perimeter is the distance around the boundary of the pixel.
- Circularity factor computed as ratio major axis to minor axis.
- Deviation factor is the ratio of circularity factor to area.
- Form factor computed as $(4 * \pi * \text{area} / (\text{perimeter})^2)$
- Aspect ratio is the ratio of the objects to its width.
- Equivalent diameter Computed as $\text{sqrt}(4 * \text{Area} / \pi)$.
- Diameter computed as $4 * \text{area} / \text{perimeter}$.

2. Textural feature

In this, Haralick and Hu moment are used to find out texture information from the image. Four main haralick features namely Homogeneity, contrast, entropy and energy are used.

E. **Extreme Gradient Boosting (XG-Boost) Classification:-**

For the classification of blood cells XG-Boost is used here. Firstly, RBC is classified in normal cell and abnormal cells. After that abnormal cells are categorised in three different classes which are sickle cells, elongated cells and echinocytes. XG-Boost [7] based on decision-tree ensemble Machine Learning algorithm that uses a gradient boosting framework. In detection problems are not involving structured data. it uses unstructured data (text, images, etc.) artificial neural networks have a tendency to leave behind all other frameworks. However, decision tree based algorithms are considered in structured/tabular best-in-class right now. The algorithm differentiates itself from other because portability, cloud integration, language and Parallelization. In boosting grows trees sequentially and removes error from previous tree. It is learn from predecessors and update the residual error [8] [11].

- An initial model F_0 is defined to predict the target variable y . This model will be associated with a residual $(y - F_0)$
- A new model h_1 is fit to the residuals from the previous step

- Now, F_0 and h_1 are combined to give F_1 , the boosted version of F_0 . The mean squared error from F_1 will be lower than that from F_0 :

$$F_1(x) \leftarrow F_0(x) + h_1(x) \tag{1}$$

IV EXPERIMENT RESULT

Proposed work can be implemented on microscopic images of any size. Microscopic images of red blood cells contain number of normal and abnormal cells. From the image region of interest is extracted and find feature of each cells. Firstly process on each cells and classified various categories.

A. Single Level Analysis

For verifying the accuracy of proposed system, microscopic images used in studies is being utilized. The Linux is utilized as operating environment and Open-CV Python 3.6.9 is adopted for experimentation of proposed work.

“Fig. 2” shows Analysis of normal cells ROI extracted from the microscopic image taken as input in “Fig. 2(a)”. Resize input image are converted into gray scale and enhance quality of image in “fig.2 (d)”. Morphological operation and threshold based segmentation is applied shown in “fig. 2(e)”. obtain gray scale image to extract texture feature are HU-moment and haralick. “Fig. 2(g)” it can be concluded that cells detected as normal cells category. Same result it is concluded for “fig.3 (g)” Sickle cells, “fig.4(g)” Elongated cells and “fig.5(g)” Echinocytes.

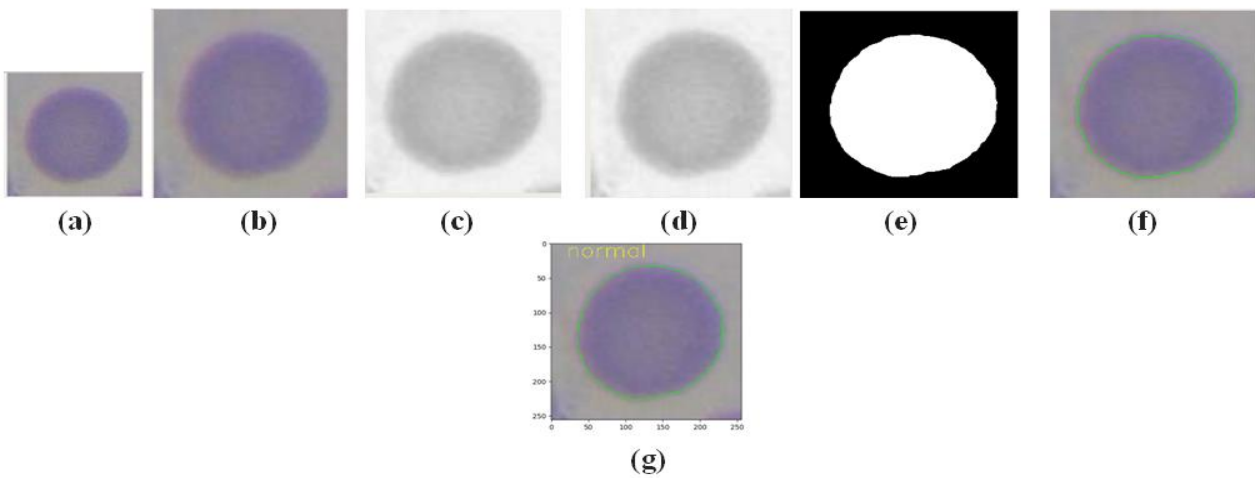


Fig.2. Analysis of Normal cells (a) Input image. (b) Resize image. (c) Gray image (d) Enhance image (e) Segmented image. (f) Image segmented cells. (g) Output of Normal cells

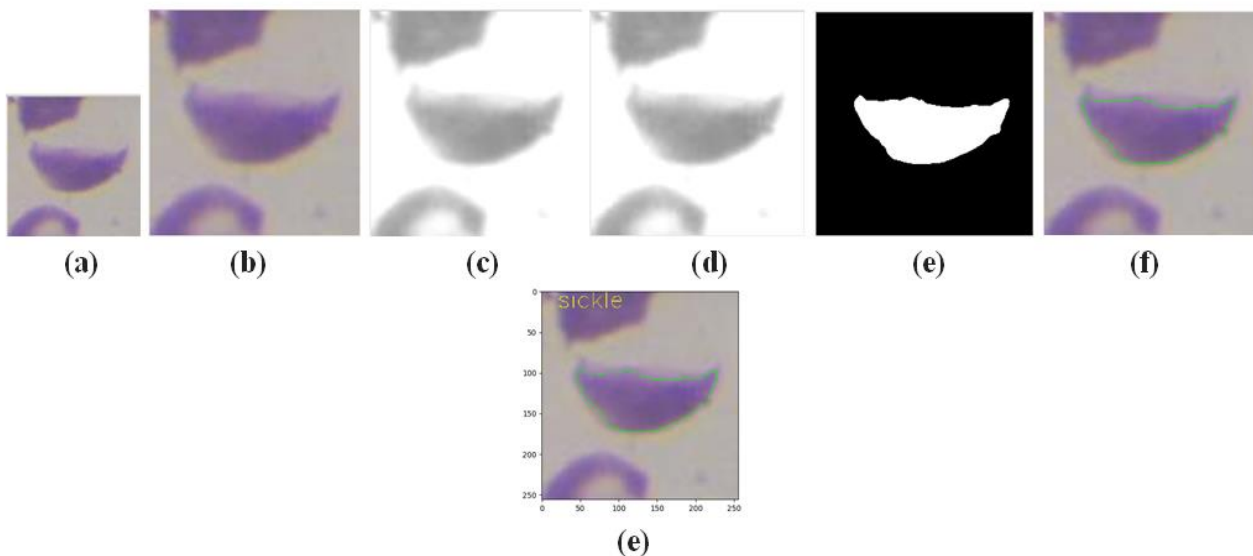


Fig. 3. Analysis of Sickle cells (a) Input image. (b) Resize image. (c) Gray image (d) Enhance image (e) Segmented image. (f) Image segmented cells. (g) Output of Sickle cells

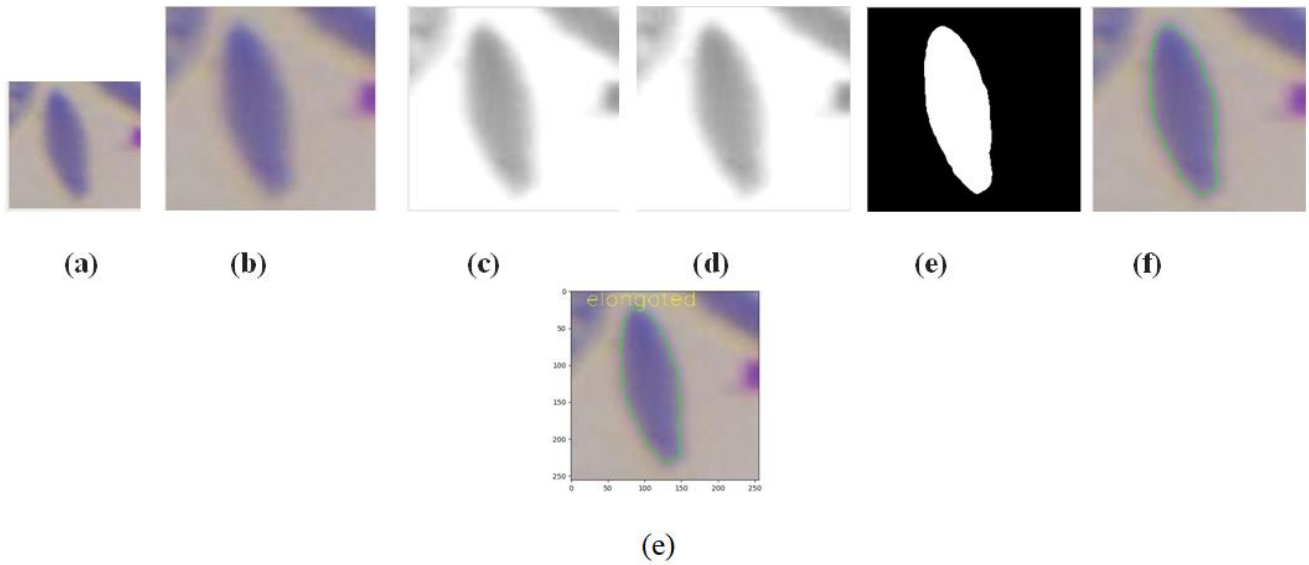


Fig.4. Analysis of Elongated cells (a) Input image. (b) Resize image. (c) Gray image (d) Enhance image (e) Segmented image. (f) Image segmented cells. (g) Output of Elongated cells.

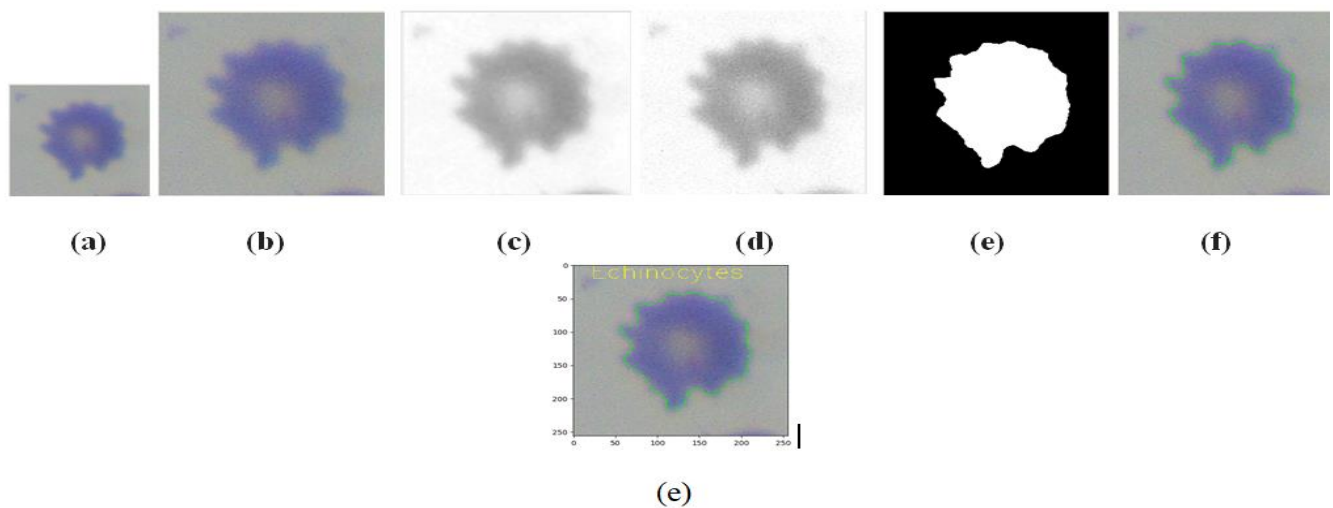


Fig.5. Analysis of Echinocytes (a) Input image. (b) Resize image. (c) Gray image (d) Enhance image (e) Segmented image. (f) Image segmented cells. (g) Output of Echinocytes

B. Multilevel Analysis

After analysis of each cells or single blood cells, it is performed same operation on whole image. In this analysis, the proposed system is used microscopic images database containing abnormal cells with sickle cells, Echinocytes and elongated cells. Following figure shows (a) Input Image (b) Gray Image (c) binary segmented Image (d) Test Image Cells. Acquired input image from database, resize image for appropriate dimension and convert it into gray scale. Median filter is used to blur the noise and threshold segmenting perform on it. After that morphological application applied on the image to obtain binary segmented image. In the test image cells, rectangular boxes cover to the abnormal cells which are available like elongated cells “Fig.6. (d)”, sickle cells “Fig.7 .(d)” and Echinocytes “Fig.8. (d)” based on shape feature.

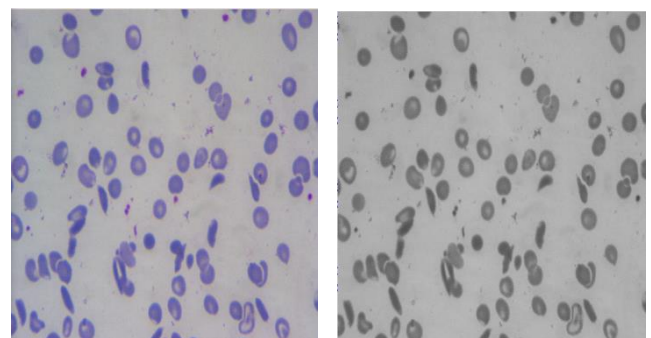


Fig.6. (a) Input Image

Fig.6.(b) Gray Image



Fig.6.(c) Binary Segmented Image

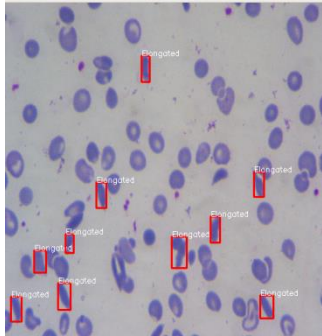


Fig.6. (d) Test Image Cells

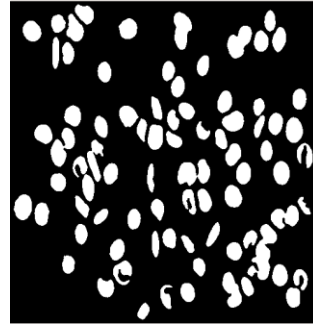


Fig.8.(c) Binary Segmented Image

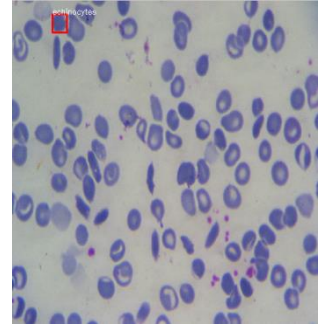


Fig.8. (d) Test Image Cells

Table-I. Geometrical or Shape Features

Features	Normal Cells	Sickle Cells	Elongated Cells	Echinocytes
Area	29623.0	8704.5	9107.0	18328.0
Parameter	648.78	547.91	513.95	605.12
Aspect Ratio	1.01	0.5	0.70	0.87
Diameter	182.63	63.54	70.87	121.15
Equivalent Diameter	194.20	105.27	107.68	152.76
Circularity Factor	0.95	0.67	0.22	0.89
Form Factor	0.88	0.36	0.43	0.62
Deviation Factor	3.23	7.72	2.511	4.89

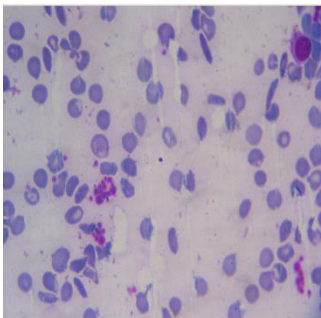


Fig.7.(a) Input Image

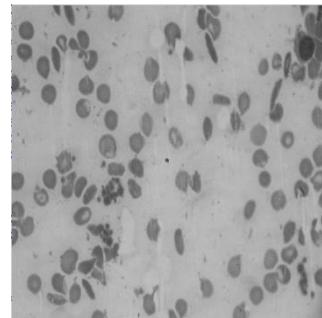


Fig.7.(b) Gray Image

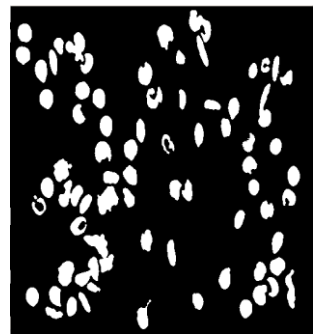


Fig.7.(c) Binary Segmented Image

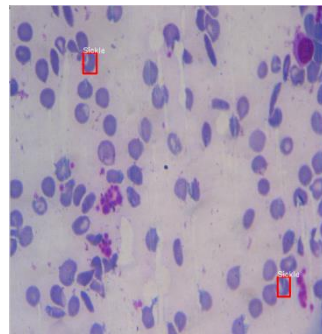


Fig.7. (d) Test Image Cells

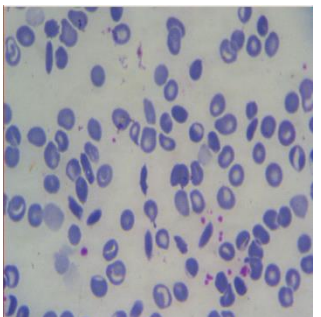


Fig.8. (a) Input Image

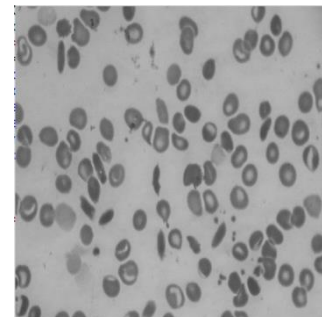


Fig.8. (b) Gray Image

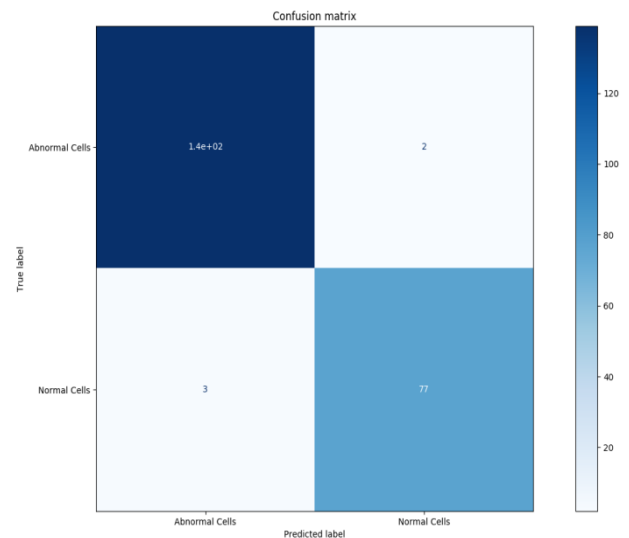


Fig 9: Confusion matrix of the result

Above “fig.9” shows the confusion matrix of XG-Boost classifier for testing dataset. From the data set some numbers of images are of abnormal patients and remaining is of normal patients. After performed classification by XG-Boost classifier gives So testing accuracy for Extreme Gradient Boosting 98% achieved. As number of image samples increases the for training as well as testing phase, it will help

to increase the rate of accuracy and other some parameter like sensitivity, specificity, precision.

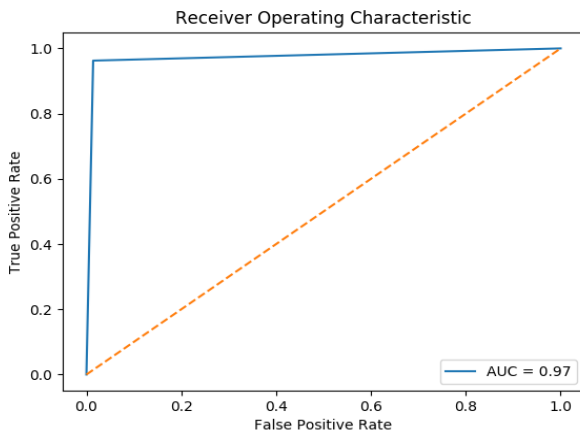


Fig.10. Receiver Operating Characteristic Curve

Receiver Operating Characteristic Curve (ROC) shows “fig. 10” the plot of sensitivity against the false positive rate. It shows the relationship between true positive rate and false positive rate. Area under ROC curve (AUC), which is commonly used as an important metric of performance, always results in the values between 0 and 1; the closer the AUC to 1, the better the performance.

Table- II. Performance of Normal Cells and Abnormal Cells

parameter	Precision	Recall	F1-Score	Support
Abnormal Cells	0.98	0.99	0.98	141
Normal Cells	0.97	0.96	0.97	80
Accuracy	0.98			221
Macro avg.	0.98	0.97	0.98	221
Weighted avg.	0.98	0.98	0.98	221

V. CONCLUSION

The present study demonstrates the effectiveness of different features and the Extreme Gradient Boosting Algorithm for Red Blood cells identification and classification. The region of interest is segmented upon the base of the shape and texture feature. Form factor, Aspect ratio, Circularity factors and Deviation Factors are features utilized to examine the shape of the cells and perform a vital role in the classification process. The processing time of the XG-Boost classifier for

training is 3.06 sec. and for testing is 5.40 sec. The XG-Boost classifier achieves the classification accuracy of 98%, 97% recall and 98% f1-score. Another system can be developed with large dataset of abnormal cells in multiclass to high accuracy.

REFERENCES

- [1] Barpanda, S.S., "Use of Image Processing Techniques to Automatically Diagnose Sickle-Cell Anemia Present in Red Blood Cells Smear." 2013.
- [2] P. Rakshit, and K. Bhowmik, "Detection of Abnormal Finding in Human RBC in Diagnosing Sickle Cell Anaemia Using Image Processing," ScienceDirect, vol. 10, pp. 28-36, 2013.
- [3] Aruna N.S. , Hariharan S. "Edge Detection of Sickle Cells in Red Blood Cells", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, Pages 4140-4144, ISSN 0975-9646.
- [4] Hala Algailani and Musab Elkheir Salih Hamad, "Detection of Sickle Cell Disease Based on an Improved Watershed Segmentation" , 2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCEEE).
- [5] T. S. Chy and M. A. Rahaman, "Automatic Sickle Cell Anemia Detection Using Image Processing Technique," 2018 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE), IEEE, pp. 1–4, Nov. 2018.
- [6] Pradeep Kumar Das and Rutuparna Panda, "A Review of Automated Methods for the Detection of Sickle Cell Disease", IEEE Reviews in Biomedical Engineering, pp.1937-3333 (c) 2019.
- [7] Tianqi Chen: Introduction to Boosted Trees: <https://homes.cs.washington.edu/~tqchen/pdf/BoostedTree.pdf>
- [8] Tianqi Chen, Carlos Guestrin: XGBoost: A Scalable Tree Boosting System <https://arxiv.org/abs/1603.02754>
- [9] https://en.wikipedia.org/wiki/Sickle-cell_disease.
- [10] <https://www.analyticsvidhya.com/blog/2019/03/opencv-functions-computer-vision-python/>
- [11] <https://medium.com/@gabrielziegler3/multiclass-multilabel-classification-with-xgboost-66195e4d9f2d>
- [12] M. Gonzalez-Hidalgo, F. A. Guerrero-Pena, S. Herold-Garca, A. Jaume-iCapo, and P. D. Marrero-Fernandez, "Red Blood Cell Cluster Separation from Digital Images for use in Sickle Cell Disease," IEEE Journal on Biomedical and Health Informatics, pp. 2168-2194, 2013.