Proceedings of the Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV 2021).
IEEE Xplore Part Number: CFP21ONG-ART; 978-0-7381-1183-4

2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV) | 978-1-6654-1960-4/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICICV50876.2021.9388610

Machine learning based Diagnosis and Classification Of Sickle Cell Anemia in Human RBC

Bheem Sen
Dept.Of Electronics and Communication Engineering Amity University Noida,Utter Pradesh,India
senbheem1998@gmail.com

Adarsh Ganesh
Dept.Of Electronics and Communication Engineering Amity University Noida,Utter Pradesh,India
adarshganesh130@gmail.com

Anupama Bhan
Dept.Of Electronics and Communication Engineering Amity University Noida,Utter Pradesh,India
abhan@amity.edu

Shubhra Dixit
Dept.Of Electronics and Communication Engineering Amity University Noida,Utter Pradesh,India
sdixit3@amity.edu

Ayush Goyal
Department of Electrical Engineering and Computer Science Frank H. Dotterweich College of Engineering, Texas A&M University Kingsville, United States.
ayush.goyal@tamuk.edu

Abstract*— Anemia is a disease which is caused by the deficiency of red blood cells. The shape of red blood cell changes to sickle or crescent shape in sickle cell anemia disease. The manual inspection of microscopic images is very difficult and time-consuming process. In this research image processing and machine learning techniques is used to automate the process of detection of sickle cells in microscopic images then classify the RBC into three shapes: circular, elongated (sickle cell) and other shape. The microscopic image is pre-processed and Otsu thresholding technique is used for segmentation. Then, Watershed segmentation is applied to separate the overlapped cells. The geometrical, statistical and textural features are extracted from the images. The machine learning classifier random forest, logistic regression naïve baye sand support vector machine is used. This research describes the comparison among these algorithms.

Keywords—Red blood cell, sickle cell anemia, Otsu thresholding, watershed segmentation, random forest, logistic regression, naïve bayes, support vector machine.

I. INTRODUCTION

Blood is a fluid which transports nutrients, oxygen to the cells and waste products away from the cell. Blood consists of platelets which helps to prevent bleeding, red blood cell helps to carry oxygen and white blood cell are the defenders which stops protects over body from infectious disease [1]. Anemia is most common blood disorder which is caused by lack of red blood cells which unable to deliver enough oxygen throughout the body. When there is a precipitous drop in RBC acute anemia is caused and chronic anemia is caused when their slow decline in red blood cell, it commonly occurs with inflammatory disorders [2]. Sickle cell disorder is a situation in which the RBCs are not produced like they should be. RBCs typically appear like circular discs. But in sickle cell disease, its shaped like a crescent moon, or an ancient agriculture instrument known as a sickle. It is a genetic RBC disease in which there are not sufficiently healthy RBCs to transport oxygen throughout the body [3]. Indication and signs of SCA are typically around five months of age. Sickle cell broke up rapidly and died, leaving relatively few RBC behind. RBC usually survive for almost four months before it needs to be replaced, sickle cells normally die within 2-3 weeks leaving shortage of RBC. Patients experiences swelling, frequent infection, vision problems, delayed growth or puberty [4]. SCA is caused by an irregular hemoglobin type called hemoglobin S. In order for the infant to be infected parent transfers the sickle cell gene to the infant, then child have sickle cell phenotype. The sample of blood in adults is taken from veins. In small children and infants, blood is generally obtained from the fingerer or foot. The basic blood test is therefore an exhausting and incorrect job that could be substituted with an accurate sophisticated and precise instrument to diagnosis sickle cell disease successfully. In this research image processing and machine learning algorithm is used to detect the SCA [5].

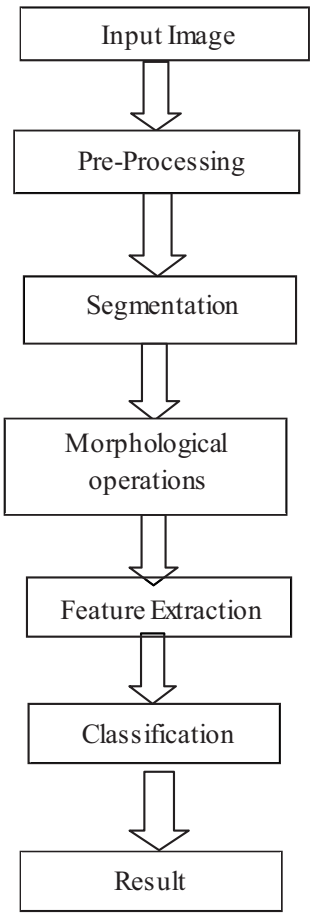
II. LITERATURE REVIEW

The first explanation about sickle cell anemia was shown by Dr. J.B. Herrick, showing the transmitted nature of SCA in his journal [6]. Many techniques have been used to diagnose infected and uninfected cells using image processing and artificial intelligence. Chayashree Patgiri et al. introduced a technique to detect normal and abnormal blood semers by segmenting the pre-processed image using niblack's algorithm [7]. Mohammed Abadulraheem Fadhel et al. circular Hough transform and watershed were used to distinguish infected and uninfected cells of different size and shape [8].Hanny A Elsalamomony proposed a algorithm for certain kind of anaemia detection including sickle and elliptocytosis in human red blood cell, and count them with good red blood semers with circular Hough transform, and some morphological devices [9].Hala Algailani et al. Image recognition method were used to suggest an automatic approach for the identification of sickle cell. For this technique watershed segmentation is a correct option because it functions well with overlapped cells [10]. Some image processing tool and technique are used, in pre-processing wiener filter, sobel edge detection used to find the corpuscles boundary. Then region properties and metric value used to determine the normal and abnormal corpuscles to diagnose the disease [11]. Laplacian of Gaussian edge detection algorithm is used to detect the abnormal cell at the earlier stage by diagnosing patient [12]. fuzzy C mean, morphological operation is applying on the image. geometrical and statistical features are extracted and the extracted feature is feed into the machine learning classifier KNN, SVM and ELM [8]. Vishwas Sharma et al. proposed a method which involves acquisition of microscopic image of blood smear median filter, water shed segmentation,

morphological operation and feature extraction such as aspect ratio, metric value, radial signature after that extracted feature is used to train K-nearest neighbour [13]. By examine the preceding studies, this paper proposes a method which uses Otsu thresholding and watershed technique used for segmentation, distinguishes between circular, elongated and others by geomaterial, statistical and textural features and classifies cells by logistic regression, support vector machine (SVM), random forest and naïve bayes and shows the comparison among the classifier.

III. METHODOLOGY

Block Diagram:



A. *Preprocessing*

The data set is collected by the special hematology department of the general hospital from Santiago de cube. The erythrocytes IDB1 consist microscopic images of blood smears collected from sickle cell patients. In data pre-processing stage the microscopic image is converted to Gray image, median filter is used to eject the noise from the Gray scale image. Final step in data pre-processing is enhancement of the image in this the small object s are removed and sharpening of the image is done to get a better-quality image.

B. *Segmentation*

Otsu thresholding technique is used to separate the three class, this method is one of the best thresholding methods is to detach object from the image. Otsu select the threshold that reduce the interclass variance of the thresholder black and white pixels.

C. *Watershed Segmentation*

Watershed is based on visualizing an image in three dimensions this provides a simple framework for incorporate knowledge-based constraint in segmentation process and utilizes image morphology. The overlapped cell is separated using this method after the separation the cell is labelled and the bounding box is formed on the sickle cell. This segmentation is performed on Gray scale image.

D. *Morphological Operation*

Morphological operation is to reject the undesirable objects from the image. in this the holes of the object are covered; borders of the image are cleared and small objects are eliminated from the image.

E. *Feature Extraction*

Geometrical, statistical and texture feature have been extracted from the circular, elongated and others.

1. *Geometrical Feature*
   Circularity, aspect ratio and eccentricity are applied.

   Circularity- The values of the circularity is by calculated by$(4*\text{Area}*\text{pi})/(\text{Perimeter}^2)$.

   Aspect ratio – The aspect ratio is the ratio of major axis to minor axis which is used to distinguish circular ,other and elongated cell.

   Eccentricity – The eccentricity of an image is the ratio of sqrt (1-(minor axis/2). ^2/ (major axis/2). ^2).

2. *Statical and texture feature*
   In statical and texture features variance, entropy, standard deviation and energy are used.

F. *Classification*

Random forest, Naïve Bayes, SVM**,** Logistic regression are implemented in last step.

1. *Random Forest*
   Random forest is a scalable, simple to use machine learning algorithm that delivers a good result much of the time, even without hyperparameter tuning. It is also one of the most commonly used algorithms due to its simplicity and versatility which can be used as both regression and classification algorithms.

2. *Naïve Bayes*
   Naïve Bayes is one of the easy and most powerful classification algorithms which is used for classification and regression problem. It is a supervised machine learning algorithm, which can make fast prediction.

3. *SVM*
   Support vector machine (SVM) is a supervised machine learning algorithm which is used for classification and regression issues.it is mainly used for classification problems,

4. *Logistic regression*
   Another technique borrowed from the world of statics by machine learning in logistic regression. Statisticians developed the logistic equation, also called the sigmoid function.

IV. Result & Discussions

The images obtained in this section are the results of simulation using MATLAB 2019B.

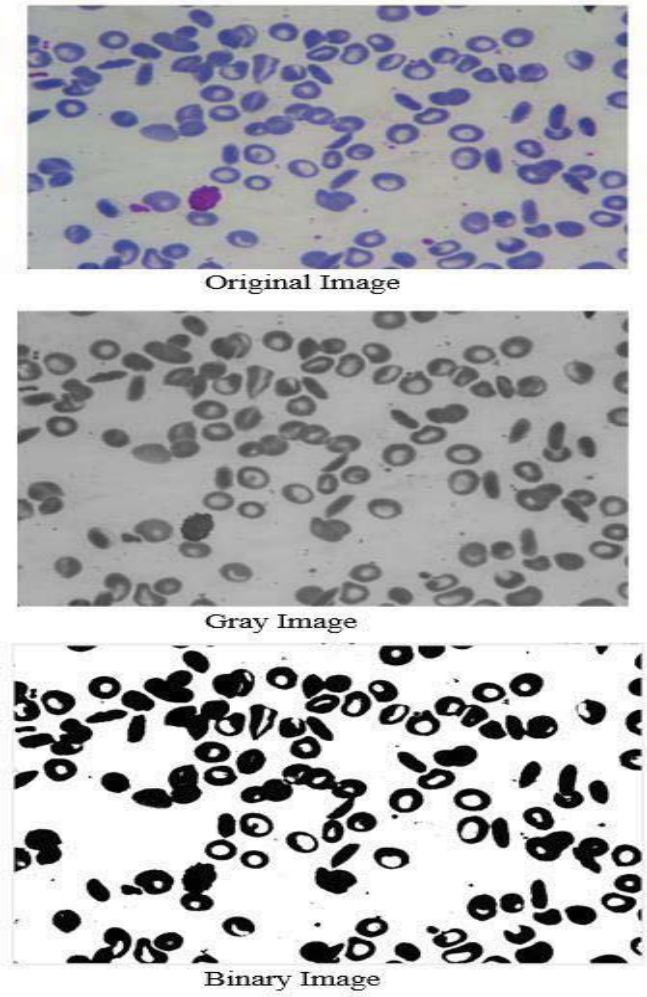


Fig.1. Original image, Gray image, Binary Image

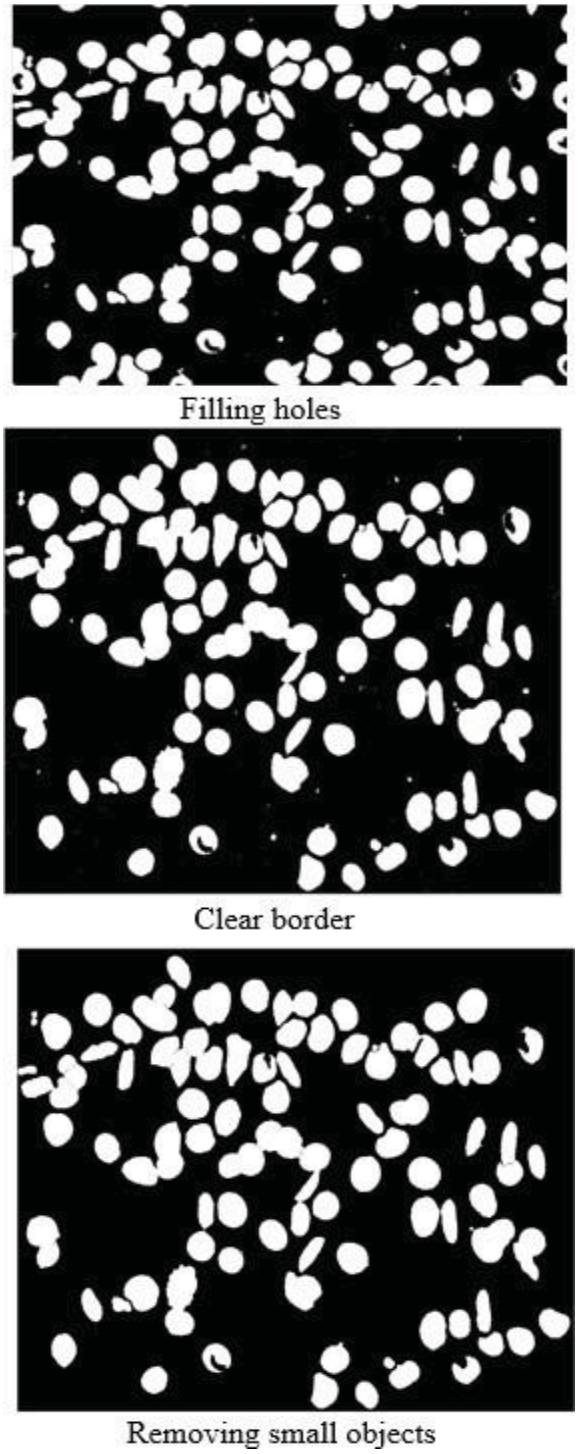


Fig.2. Morphological operation

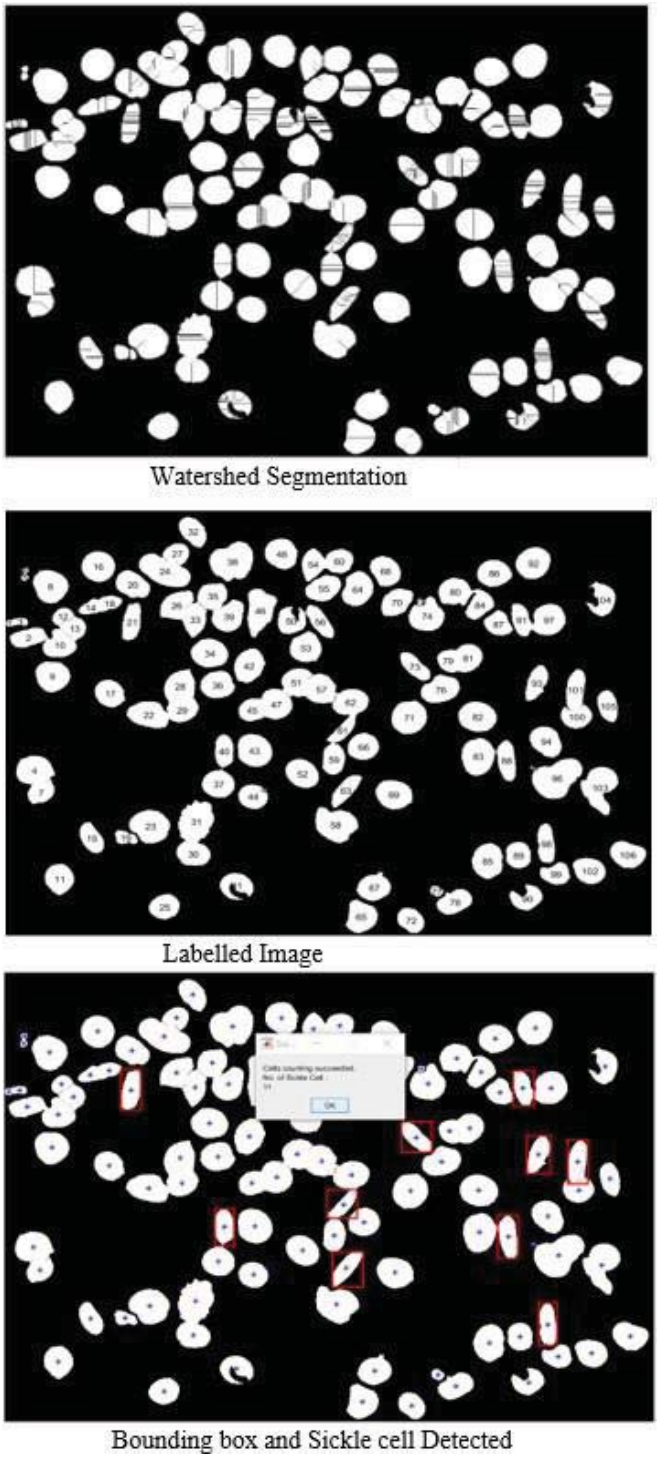


Fig.3. Watershed segmentation, Labelled Image, Bounding box and Sickle cell Detection.

TABLE I. FEATURE EXTRACTION

Features	Circular	Elongated	Others
Circularity	0.6334	0.9559	0.8539
Aspect ratio	3.3525	1.1743	1.5693
Eccentricity	0.9545	0.5243	0.7706
Standard Deviation	0.3664	0.234	0.4535
Variance	0.1343	0.4838	0.2056
Entropy	0.6339	0.9536	0.8679
Energy	0.7094	0.5102	0.5697

TABLE II. TABLE OF RANDOM FOREST

	Circular	Elongated	Others
Precision	95%	92%	91%
Recall	95%	95%	89%
F1 score	95%	93%	90%
Accuracy	92%	92%	92%

TABLE III. TABLE OF NAÏVE BAYES

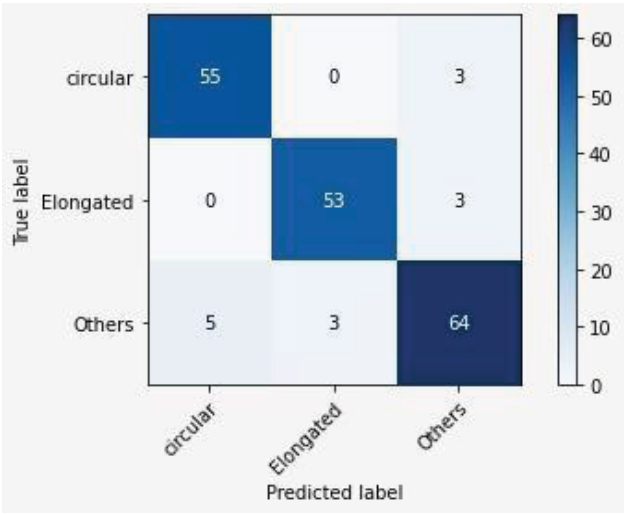
	Circular	Elongated	Others
Precision	84%	88%	90%
Recall	96%	98%	93%
F1 score	90%	70%	80%
Accuracy	88%	88%	88%

TABLE IV. TABLE OF SVM

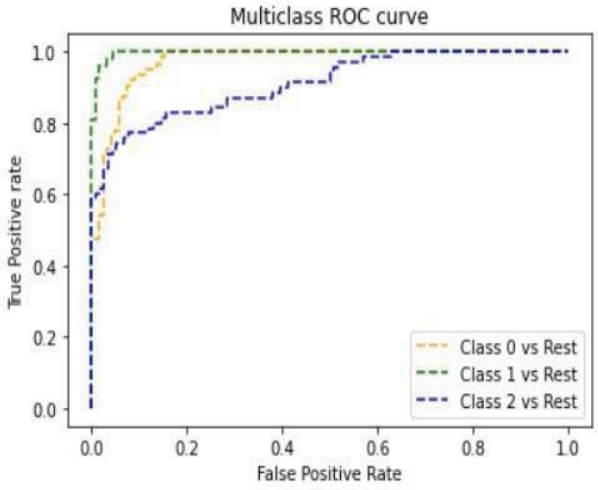
	Circular	Elongated	Others
Precision	97%	86%	89%
Recall	92%	95%	84%
F1 score	94%	90%	86%
Accuracy	90%	90%	90%

TABLE V. TABLE OF LOGISTIC REGRESSION

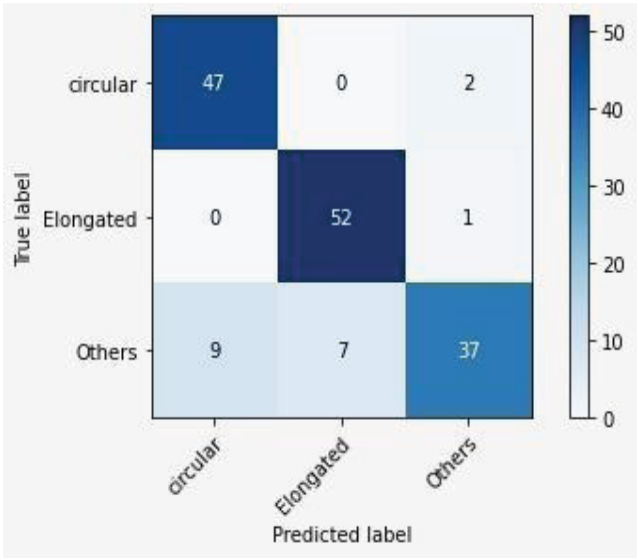
	Circular	Elongated	Others
Precision	98%	86%	86%
Recall	97%	86%	87%
F1 score	86%	98%	87%
Accuracy	90%	90%	90%



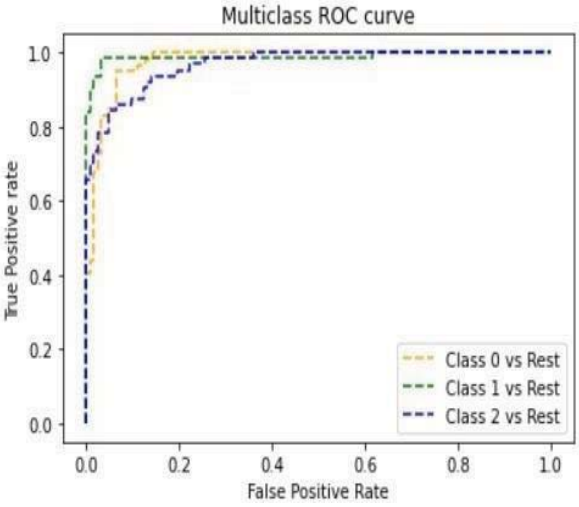
Confusion Matrix of Random Forest



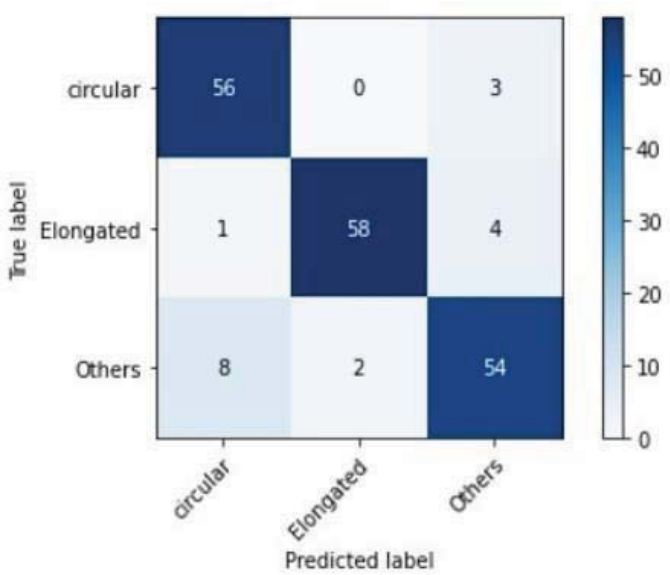
Random forest ROC curve



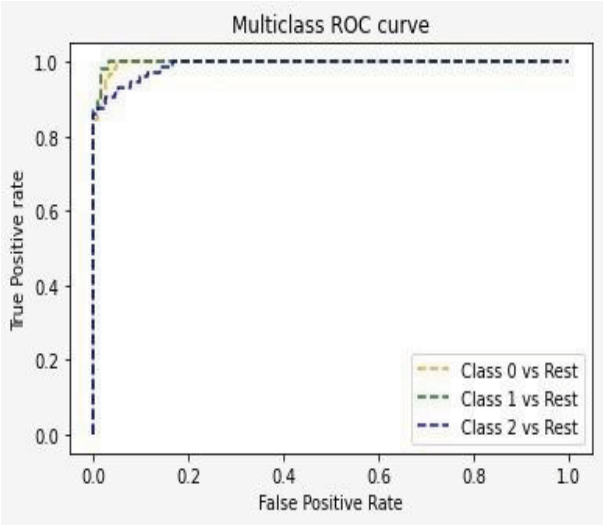
Confusion Matrix of Naïve Bayes



Naïve Bayes ROC curve

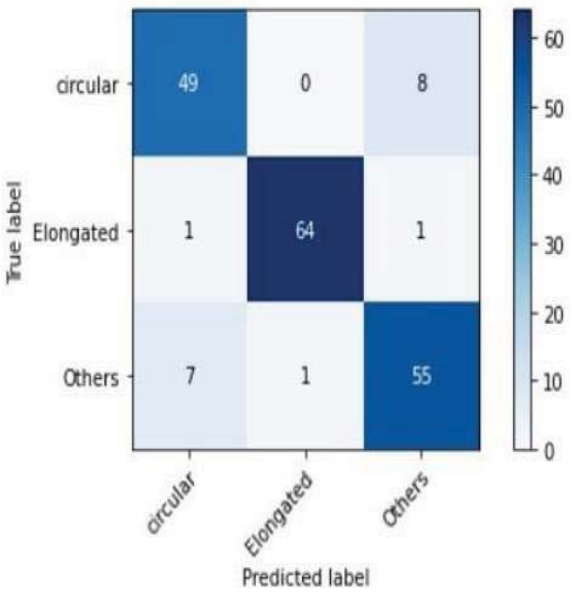


Confusion Matrix of SVM

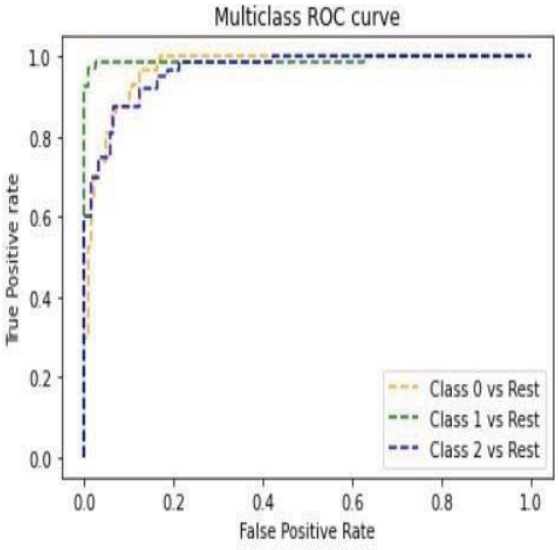


SVM ROC curve

Fig.4. Confusion Matrix& Roc curve of Random forest, Naïve bayes, SVM



Confusion matrix of Logistic Regression



Logistic regression ROC curve

Fig.6. Confusion Matrix& ROC of Logistic regression

V. CONCLUSION

Sickle cell disorder is a situation in which the RBCs are not produced like they should be. In this research Watershed segmentation has used to remove the overlapped cell. features such as geometrical, statistical features and textural are extracted from the cells. we have used four machine learning model to classify circular, elongated and other according to our data set. Naïve Bayes has less precision, recall f-score and accuracy then SVM, Random forest & Logistic regression. Logistic regression & SVM has the same accuracy 90% and is less then random forest which has a accuracy of 92%. Observing the accuracy random forest provide the good result.

VI. REFERENCES

[1] Hanny A Elsalamony "Detection of some anaemia types in human blood smears using neural networks" pp.1-13, May 2016.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Ahmad Sabry Mohamad, Nur Syahirah Abdul Halim, Muhammad Noor Nordin, Roszymah Hamzah Jameela Sathar "Automated Detection of Human RBC in Diagnosing Sickle Cell Anemia with Laplacian of Gaussian Filter" 2018 IEEE Conference on Systems, Process and Control (ICSPC 2018), 14–15 December 2018.K. Elissa, "Title ofpaper if known," unpublished.

[3] Leon L. Wheeless, Roy D. Robinson, Oleg P. Lapets, Christopher Cox, Ana Rubio,Michael Weintraub, and Lennette J. Benjamin "Classification of Red Blood Cells as Normal, Sickle, or Other Abnormal, Using a Single Image Analysis Feature" pp.1-8, 1994.

[4] Hegde, Roopa B., Keerthana Prasad, Harishchandra Hebbar, and I. Sandhya, " Peripheral blood smear analysis using image processing approach for diagnostic purposes: A review," Biocybernetics and Biomedical Engineering, 2018.

[5] Albayrak, B.; Darcie, M.B.; Kiraci, F.; Ö˘grenci, A.S.; Özmen, A; Ertez, K. Orak HücreliAnemi Tespiti "Sickle CellAnemia Detection" In Proceedings of the 2018 IEEE Medical Technologies National Congress (TIPTEKNO), Magusa,Cyprus, 8–10 November 2018; pp.1–4.

[6] Savitt, T.L. and M.F. Goldberg, Herrick's 1910 case report of sickle cell anemia: the rest of the story. Jama, 1989. 261(2): p.266-271.

[7] Chayashree Patgiri and Amrita Ganguly "Red Blood Cell and Sickle Cell Detection from Microscopic Blood Images of Sickle Cell Anemic Patient" 1Department of Electrical Engineering IEEE, pp.1-4, 2019.

[8] M A Fadhel, A J Humaidi, and S R Oleiwi, "Image processing-based diagnosis of sickle cell anemia in erythrocytes," IEEE Annual Conference on New Treand in Information and Communication Technology Applications, pp. 203–207, 2017.

[9] Hajara Abdulkarim Aliyu, Mohd Azhar Abdul Razak, Rubita Sudirman "Segmentation and Detection of Sickle Cell Red Blood Image" pp.1-12, 2017.

[10] Hala Algailani, 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 (ICCCEEE). pp. 1-4 ,2018.

[11] T. S. Chy and M. A. Rahaman, "A Comparative Analysis by KNN, SVM & ELM Classification to Detect Sickle Cell Anemia," 2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), IEEE, pp.455–459, 2019.

[12] Hegde, Roopa B., Keerthana Prasad, Harishchandra Hebbar, and I. Sandya, " Peripheral blood smear analysis using image processing approach for diagnostic purposes: A review," Biocyberntics and Biomedical Engineering, 2018.

[13] 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.

[14] Vishwas Sharma, Adhiraj Rathore, Garima Vyas "Detection of sickle cell anaemia and thalassaemia causing abnormalities in thin smear of human blood sample using image processing" pp. 1-5,2017