

# AUTOMATIC CLASSIFICATION OF SICKLE CELL ANEMIA USING RANDOM FOREST CLASSIFIER

S. Ranjana<sup>1</sup>, R. Manimegala<sup>2</sup>, K. Priya<sup>3</sup>

<sup>1,2,3</sup>Assistant Professor, Faculties of Humanities Science, Meenakshi Academy of Higher Education Research Chennai, TamilNadu, India

E-mail: ranjanas@maherfhs.ac.in<sup>1</sup>, manimegalar@maherfhs.ac.in<sup>2</sup>, priyak@maherfhs.ac.in<sup>3</sup>

## Abstract:

*SCA is a genetic category of red diseases of the blood cells. People in their red blood cells contain an abnormal protein. Part of a group of SCA diseases is sickle cell anaemia (SCA). Sickle cell anaemia is a red blood cell condition that has not been inherited in the body with ample red cells to hold oxygen.. It is dangerous because it can cause extreme pain, anemia and other symptoms. The early diagnosis is required for sickle cell anemia. In this study, the automatic classification of SCA system is discussed. Initially, the input images are given to median filter for pre-processing. Then the Gray Level Co-occurrence Matrix (GLCM) and Haralick features are extracted. Finally, Random Forest (RF) classifier is used for Prediction. The performance of SCA system produces the classification accuracy of 95%using RF classifier.*

*Keywords: Sickle cell anemia, Gray level co-occurrence matrix, haralick features, Random Forest*

## INTRODUCTION

Typically, the flexible red blood cells move blood vessels quickly. The red blood is like crescent SCA sickles or lunas. Via narrow blood vessels, these rigid, sticky cells can slow down the blood flow or block body oxygen. Many people in the United States with SCA are of African descent, but this disorder is also prevalent in Hispanics. SCA early signs and symptoms include swelling of hands and feet; anemia, including weakness or extreme fatigue; jaundice. SCA can eventually lead to complications, including inflammation, slow growth and so-called pain crises. Most kids with cell sickness are painless during emergencies, but teens and adults also suffer from persistent chronic pain.

Human blood samples detection in diagnosing the sickle cell anemia is discussed in [1] using Gaussian filter. The input images smoothed and preprocessed by using Gaussian filter. Then the second derivative is examined. The pixels are marked as edge pixels .Brain Cancer Back propagation Network Identification Program Implementation and MRI [2]The research

includes the extraction by discrete transformation of the wavelet texture from the given MRI brain sample and the subsequent neuro classification of the procedure.

Classification of brain MRI image with data mining algorithms [3]. An important task is to identify cancer in MRI of the brain. Specific methods for classifying brain cancer are used in the present study. This accessible image processing technique is used to successfully identify and remove brain cancer from MRI images. Classification and segmentation of brain tumors in MRI images using PNN [4]. Among the numerous cancers, the extent of the brain tumor is very severe to save a lifetime of prompt diagnosis and appropriate care. Detection of these cells is a concern because of tumor cell formation.

Classification of brain image with SVM study of various waves [5]. The transforming wavelet is used for the extraction of functions. Most techniques of transformation generate the same dimensional value coefficient as the original image. Classification of brain tumors based on Naïve Bayes [6]. The preprocessors are used for the brain MRI data base, morphological operations, subtracting pixels, the maximum entropy threshold, the extraction of statistics and the prediction algorithm based on Naïve Bayes classification [7-9].

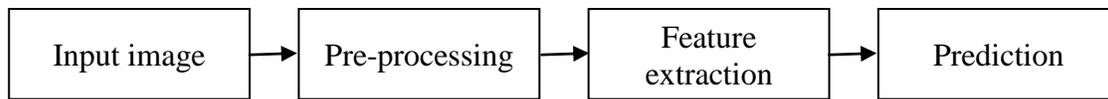
This study presents the SCA classification. The remainder of the paper is structured as follows: Section 2 addresses the techniques and materials used to define SCA. Section 3 discusses the experimental observations and discussions. The SCA closes the last part.

## **Methods and Materials**

Each algorithm has a price in the field of data analysis. However, given the general situation, the market issue as much as possible needs to be classified. The existence of the data makes it very difficult to know intuitively what to follow. We would address one of the most trusted methods of the top classifier, the Random Forest Classifier, by data experts. Random Forest also has a methodology of regression algorithm that is discussed here.

The word "forest" implies that there would be many trees there. The algorithm provides a bundle of decision-making trees for a grouping and it is also a strategy for saving when overfitting a decision-tab model. There is a high variance and low preference to a decision tree, which can yield very volatile results, unlike the often agreed logical regression with a high bias and low variance. This is the only way to save Random Forest. In comparison a random forest methodology is able to rely both on observation and training data variables for the production of individual decision-makers and to vote for classification and the overall average for regression problems. It also uses a bagging technique which selects the columns that are unable to represent significant variables at the root of all decision-making bodies in a random way. In this way, a random forest only generates trees that are interdependent by penalising precision. We have a thumb rules that can be used to pick sub-samples from random forest observations.

Initially, the input sickle cell images are given to median filter for pre-processing. Then haralick and GLCM features are used for feature extraction. Finally, RF classifier is used for prediction. Figure 1 shows the processing of SCA system.



**Figure 1 SCA system processing**

### **Pre-processing using median filter:**

The median filter is a non-optical filter system sometimes used to eliminate image or signal noise. This reduction of noise is a conventional pre-processing method aimed at increasing subsequent processing [10]. Median filtering is important for the processing of digital images because it preserves edges and noise removal under certain conditions, and also has signal processing applications.

### **Haralick and GLCM based feature extraction:**

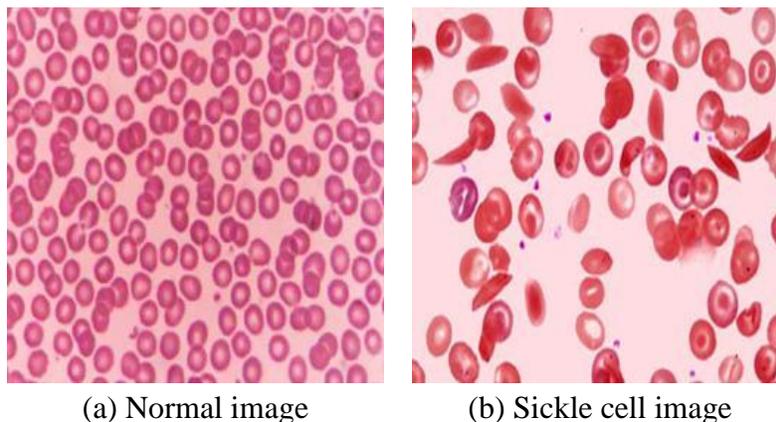
The GLCM functions define the texture of a picture by calculating the number of pixels in a picture, creating GLCM, and then removing statistically determined measures from that matrix in a given spatial relationship [11]. The texture features of Haralick are calculated using a GLCM, a matrix that shows the co-occurrence of adjacent gray levels in the image. The GLCM is a square matrix having the size of the gray level  $N$  in the area of interest.

### **RF based Prediction:**

The algorithm of the random forest comprises of different decision-making trees. When each tree is built, it uses bagging and randomness to try to create a non-relating tree forest whose prediction by comitology is more accurate. A random forest is a meta-estimator that matches a number of decision-tab classificatory for different data samples and uses averages to boost prediction precision and overfit power.

## **Results and Discussion**

The images are given to median filter for pre-processing, then the GLCM and haralick features are used for feature extraction. The sample SCA images and normal images are shown in figure 2.

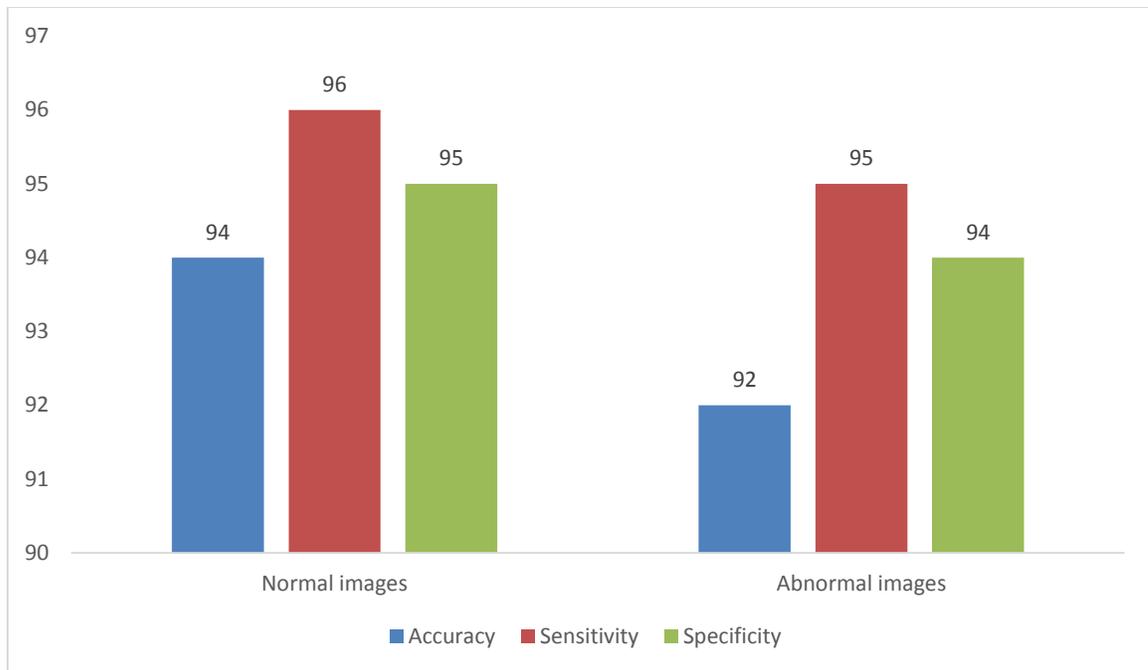


(a) Normal image

(b) Sickle cell image

**Figure 2 Sample SCA images and normal images**

The prediction is made by RF classifier. The classification accuracy, sensitivity and specificity are shown in figure 3.



**Figure 3 Graphical representation of RF prediction**

The figure above indicates that the average accuracy of the classification is 95%. Normal images are grade accurate 94% and their sensitivity and characteristics 96% and 95%. The accuracy of the classification of odd pictures is 92% and the sensitivity and organisms are 95% and 94% using GLCM-based haralick and RF classifier..

### Conclusion:

An automatic classification of sickle cell anemia using random forest classifier is described in this study. Initially the images are preprocessed by median filter. The GLCM and haralick functionality are extracted. For final performance estimation, the RF classifier is used. The total classification accuracy is 95% using statistical characteristics and RF classification. It's sensitivity and specificity are 94 % and 96 %. The accuracy for SCA images are 92 % and also its sensitivity and specificity are 95% and 94%.

### REFERENCE:

- [1] Mohamad AS, Halim NS, Nordin MN, Hamzah R, Sathar J. Automatic detection of RBC human with Laplacian Gaussian Filter in diagnosis of Sickle Cell Anemia. In 2018 IEEE Conference on Systems, Process and Control (ICSPC) 2018 Dec 14 (pp. 214-217). IEEE.
- [2] Chy TS, Rahaman MA. KNN, SVM & ELM Classification for Sickle Cell Anemia A comparative analysis. In 2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST) 2019 Jan 10 (pp. 455-459). IEEE.
- [3] Chy TS, Rahaman MA. Automatic detection of sickle cell anaemia by imagery. In 2018 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE) 2018 Nov 22 (pp. 1-4). IEEE.

- [4] Panneerselvam P. Embedded system application of sickle cell anaemia for a genetic disease. In 2014 International Conference on Advances in Electrical Engineering (ICAEE) 2014 Jan 9 (pp. 1-4). IEEE.
- [5] AbdulraheemFadhel M, Humaidi AJ, RazzaqOleiwi S. Diagnosis of sickle cell anaemia in erythrocytes based on image processing. In 2017 Annual Conference on New Trends in Information & Communications Technology Applications (NTICT) 2017 Mar 7 (pp. 203-207). IEEE.
- [6] Berger SA, Carlson BE. Sickle microcirculation cell blood flow. In The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2004 Sep 1 (Vol. 2, pp. 5057-5060). IEEE.
- [7] Saravanan, N. (2013). Hand geometry identification based on K-media clustering and segmentation algorithm optimised.. International Journal of MC Square Scientific Research, 5(1), 11-14.
- [8] Narayanan, K. L., & Ramesh, G. P. (2017). Discrete wavelet compression of the baseline image with frequency strip removal and improvement of performance. International Journal of MC Square Scientific Research, 9(2), 176-182.
- [9] R, Reka. 2018. "Gray Scale Image Segmentation Directive Thresholding Algorithm", International Journal Of Advances In Signal And Image Sciences, 4 (1), 23-29.
- [10] Vijayabaskar, Kumanan T. (2020) Department of CSE, Meenakshi Academic of Higher Education and Research, Chennai, India, Link Quality and Energy-Aware Metric-Based Routing Strategy in WSNS. In: Solanki V., Hoang M., Lu Z., Pattnaik P. (eds) Intelligent Computing in Engineering. Advances in Intelligent Systems and Computing, vol 1125. Springer, Singapore.
- [11] Sivasankar C., Kumanan T. (2020) Department of CSE, Meenakshi Academic of Higher Education and Research, Chennai, India, Friend List-Based Reliable Routing in Autonomous Mobile Networks. In: Solanki V., Hoang M., Lu Z., Pattnaik P. (eds) Intelligent Computing in Engineering. Advances in Intelligent Systems and Computing, vol 1125. Springer, Singapore.