Cell classification in microscopic images for anemia detection

Priyadharsini RAVISANKAR, Beulah ARUL, Srivarsha ELANGO

Department of Computer Science and Engineering

Sri Sivasubramaniya Nadar College of Engineering, Chennai, Tamil Nadu, India

priyadharsinir@ssn.edu.in, beulaharul@ssn.edu.in, srivarshaelango@gmail.com

Abstract: Sickle cell anemia is an inherited disorder of the red blood cells in which there is an insufficient number of healthy red blood cells to transport oxygen effectively throughout the body. When observed under a microscope, the blood cells of an individual with sickle cell anemia exhibit a crescent or sickle-like shape. Image segmentation and classification techniques are necessary for detecting sickle cell anemia from microscopic images. Segmentation of the images is performed to distinguish between healthy and sickle (unhealthy) red blood cells. A mask of the microscopic image is first generated, which aids in the classification of cells. The mask generated is then further classified into circular, elongated and other cells, with the help of image processing. The erythrocytesIDB dataset is used for the detection of cells, which was provided by Universitat de les Illes Balears and available at http://erythrocytesidb.uib.es/. Upon the successful classification of cells into their respective types, it is concluded that circular cells are representative of healthy red blood cells, whereas elongated cells are characteristic of sickle cells. However, cells classified as "others" are indeterminate, and their association with sickle cells is uncertain, hence presenting a degree of ambiguity. In terms of performance evaluation, the proposed method achieved an accuracy of 94.907% in cell classification.

Keywords: Classification, ErythrocytesIDB Dataset, Segmentation, Sickle Cell Detection.

1. Introduction

Sickle cell anemia is an inherited blood disorder in which the red blood cells become abnormally shaped and can cause blockages in blood vessels, leading to various complications. One of the distinctive traits of sickle cell anemia is the tendency of the blood cells to undergo sickling, resulting in their elongation, stiffness and reduced flexibility, which in turn leads to blockages within small blood vessels. The blood sample of an anemic patient, when viewed using High-Performance Liquid Chromatography, helps in identifying the presence of sickle cells.

Detection of sickle cell anemia from microscopic images using image processing, involves segmentation of the images. Segmenting the images is a crucial step in the classification of images into healthy and anemic cells. A major challenge faced in processing microscopic images is their quality and resolution which highly differ from one another. Additional processing techniques must be employed to overcome the same, lest they result in a reduction of the accuracy of detection.

The dataset used for testing and training purposes has been provided by Universitat de les Illes Balears and is available at “http://erythrocytesidb.uib.es/”. The dataset contains microscopic images of varying quality and color tones. Sample images from the dataset are shown in Figure 1. These images have to be processed accordingly so as to enable the generation of an accurate mask. The mask thus generated is then classified into circular, elongated and other cells with the help of an appropriately defined threshold.

The threshold is in essence a roundness factor which helps to separate the cells above a certain value to be healthy RBCs (circular cells) and those below another value to be sickle cells (elongated cells). The cells which lie between these two values are classified as other cells which have an equal probability of being healthy or unhealthy. The classification of other cells requires expert medical knowledge which is outside the context of our work.

The efficiency of our proposed method is determined by the ratio of the cells being identified by our algorithm to the actual number present, expressed in percentage.

https://doi.org/10.33436/v34i1y202401

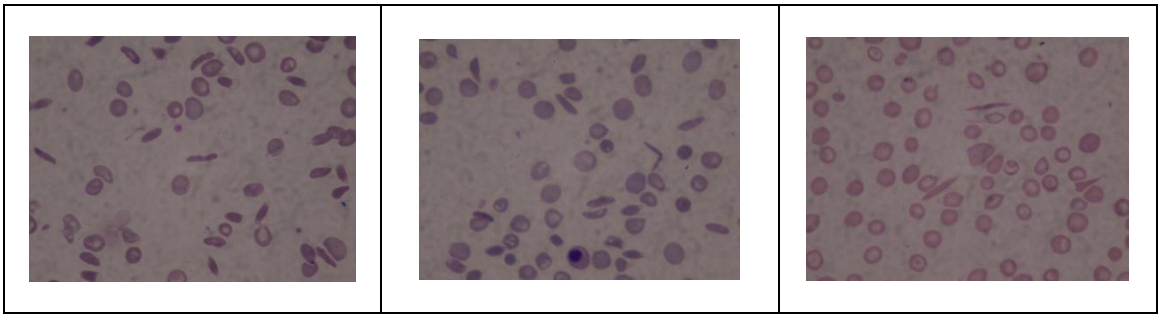

Figure 1. Sample images from dataset

The proposed work aims at classifying the cells in microscopic images which would be helpful for treatment of anemia. The objective of the work is to separate the cells based on their shape as Circular (healthy cells), Elongated (sickle cells) and Others (either healthy or unhealthy cells). If the microscopic image of a patient has a greater number of circular cells then he/ she is said to be healthy. On the other hand, if the elongated cells are large in the samples, then the situation is categorized as anemia which requires immediate attention with medication. The proposed method using segmentation technique has been proved to classify cells with an accuracy rate of 94.907%.

2. Literature survey

A method for the analysis of the shape of erythrocytes in peripheral blood samples has been proposed by many researchers (Gonzalez-Hidalgo et al., 2015; Hegde et al., 2018; Kiruthika et al., 2021). The proposed technique makes use of ellipse adjustments and a new algorithm in order to detect notable points. It also applies a set of constraints to eliminate numerous images processing steps, significantly reducing the execution time. A review of the current and emerging techniques for sickle cell disease detection has been conducted (Arishi et al., 2021). This study highlights the different potential methods that could be applied for the early detection and diagnosis of the SCD. The various available technologies were tabulated and compared based on a number of performance measures including sensitivity, specificity and accuracy. An efficient technique using a decision system has been proposed (Acharya & Prakassha, 2019).

The decision system is built by deriving rules from a decision tree. Granulometric analysis is carried out to separate the RBCs, and a modified Watershed transform algorithm is used to isolate cluttered cells. It defines a systematic approach to categorize blood cells with/without central pallor. An approach to select the best classification method and features has been explored (Petrović et al., 2020).

Microscopic images are pre-processed and segmented for ensuring high feature quality. The best parameters for every classifier are found using Randomized & Grid search and the classification methods are compared with the state-of-the-art, to obtain better results. A machine learning algorithm to detect overlapping RBCs for sickle cell diagnosis has been developed (Vicent et al., 2022).

The algorithm employs canny edge and double threshold techniques to detect if overlapping red blood cells are sickle cell anemic or not. The identification of aberration in normal parameters of RBCs in an anemic blood sample has been explored (Rakshit & Bhowwmik, 2013).

Preprocessing is done using Weiner filter Sobel edge detection methods to find the boundary of the corpuscles. Region properties are then used to formulate a metric to determine the abnormal shape of corpuscles to diagnose the disease. An image processing technique to detect abnormal blood cells has been proposed by researchers (Bhatt & Prabha, 2015).

Edge detection technique is first performed to calculate the area and perimeter to determine the form factor. The count of abnormal cells is then calculated. A review of the current developments in computerized PBS analysis has been conducted (Navya et al., 2022). Existing automated image processing methods for identification, segmentation, feature extraction and classification of RBCs have been summarized in the paper. Various approaches for the classification of anemia have also been included as part of this review.

3. Proposed work

The sample microscopic images from the dataset are initially segmented to simplify the process of classification that follows. Various state of the art techniques are available for segmentation including edge detection, thresholding and clustering. Clustering techniques are further divided into hard clustering, k-means clustering and fuzzy clustering techniques.

The technique adopted in our proposed work is the thresholding-based segmentation technique which recognizes various regions in the image based on a threshold value. Pixels with intensity above a particular threshold are classified as the foreground region and are assigned a value 1, while those below the threshold are assigned a value of zero (background), thereby generating a binary mask of the image.

A major challenge faced in the segmentation process is the difference in the image qualities. The dataset for the work includes microscopic images of blood samples providing no guarantee regarding the resolution and clarity of the image. However, the algorithm is successful in overcoming this issue by increasing the sharpness of every image prior to segmentation. This improves the resolution of the image, thereby making it easier to distinguish the background from the foreground.

The mask is then processed to classify the various cells as circular, elongated and other cells. In order to achieve this, a threshold value known as the rounding factor (referred to as rf in the following sections of this text) is calculated for each red blood cell. All cells which have a rf value greater than 0.734 are classified as circular cells whereas those that have a rf value less than 0.6723 are classified as elongated (sickle) cells. The cells whose rf value lies between these two values may belong to either of them and hence are classified as other cells.

A separate mask for each of the above classified cells has been created and is the output of our proposed work. From these images, the ratio of the number of cells discovered to the actual number of cells is calculated as a percentage. The average ratio determined for the three different types of cells gives the accuracy for each sample image. The final efficiency of the work is calculated as the mean of the efficiency values of all the images present in the dataset.

3.1. Algorithm

a) Load the microscopic image from local file system;

b) Generate an RBC Mask of the microscopic sample:

i. Increase the sharpness of the image in order to improve the clarity and reduce noise;

ii. Extract the green component from the sharpened image;

iii. Label the background and foreground of the image based on an appropriate threshold value;

iv. Negate the background (black) and make the foreground positive (white).

c) Calculate the roundness factor x for each cell in the foreground of the RBC Mask:

i. If $x > 0.734$, the cell is almost circular and hence is labelled as a healthy RBC cell;

ii. If $x < 0.6723$, the cell is regarded as elongated and is classified as a sickle cell;

iii. If x lies between 0.6723 and 0.734, it is irregularly shaped and has an equal probability of being a healthy RBC or sickle cell.

The different types of cells are extracted as mentioned above and the final output images are generated and saved to the file system.

First, the given microscopic image is processed and the green component is extracted from it. Then the background is differentiated from the foreground (cells) based on a user defined threshold value. The background is given a black fill (negative) while the foreground is made white (positive).

After generating the RBC mask, a value known as the roundness factor (x) is calculated for every cell present in it. Cells which have x value > 0.7534 are classified as circular, those which have

$x < 0.6723$ are classified as elongated and the ones whose x value falls between the two extremes are categorized as other cells and the output is generated accordingly.

3.2. Experimental results

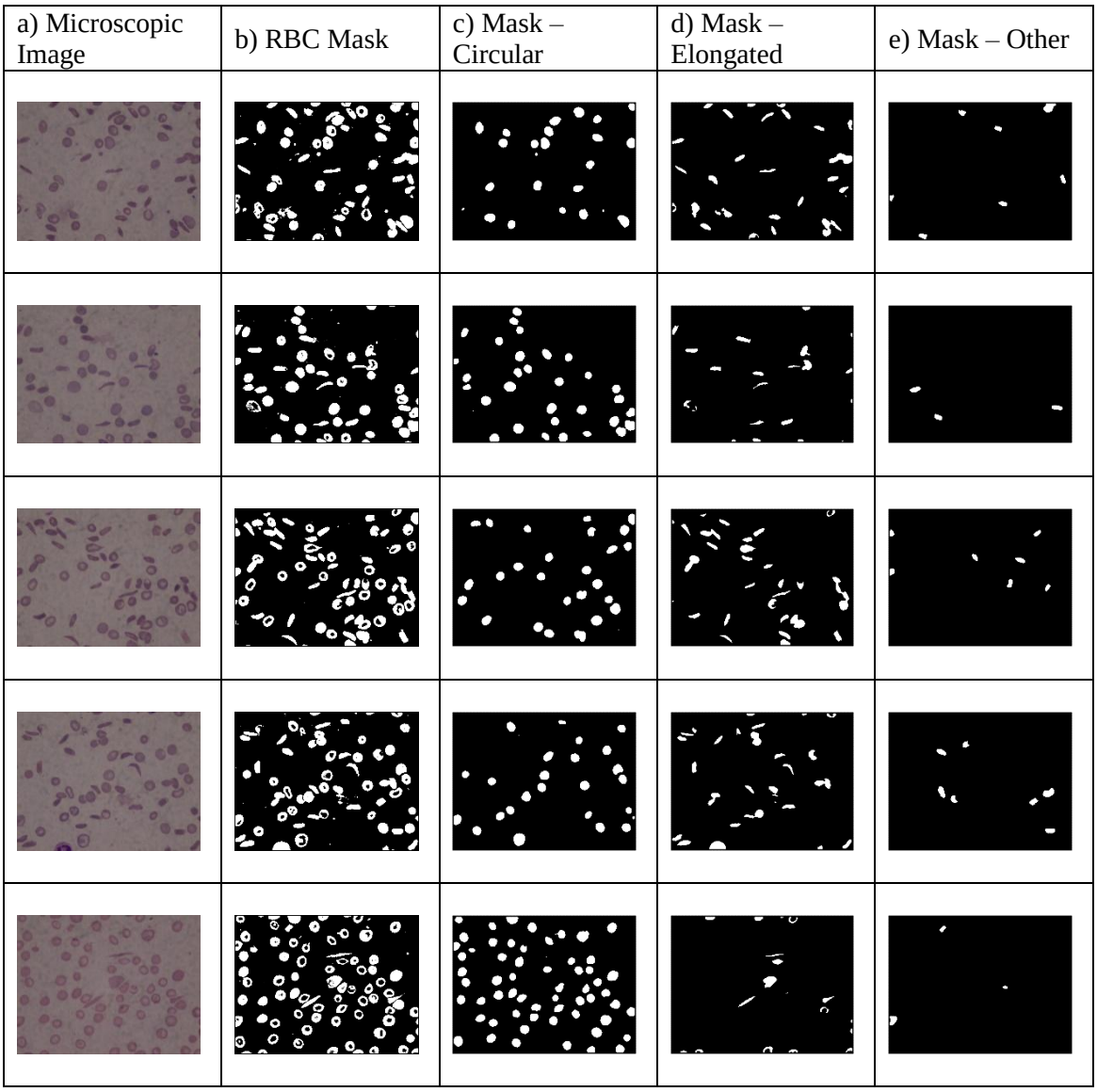

Figure 2. a) Microscopic Image, b) RBC Mask, c) Mask – Circular, d) Mask – Elongated, e) Mask – Other

The images consist of peripheral blood smears samples of patients with sickle cell anemia classified by a specialist from Dr. Juan Bruno Zayas Hospital General in Santiago de Cuba, Cuba. The dataset consists of 50 different samples of patients with sickle cell anemia.

Figure 2 shows the experimental results of detection of sickle cells. Figure 2.a) is the input microscopic image taken for experiment from the dataset. Figure 2.b) represents the RBC mask obtained after sharpening and thresholding. Figure 2.c), 2.d) and 2.e) show the circular, elongated and other shaped cells respectively.

3.3. Performance measure

The performance of the proposed work is measured in terms of its efficiency. Efficiency is calculated as the ratio of the number of cells in each category identified by the algorithm to the number actually present in the microscopic blood sample and is expressed in percentage.

The accuracy of the detected cells is given below:

Circular Cells: 96.414%

Elongated Cells: 94.274%

Other Cells: 91.409%

3.3.1. Comparison of the related work

Works	Accuracy
Feature extraction and labelling [2]	88%
Watershed based segmentation [10]	90%
Proposed work using thresholding-based techniques	**94.907%**

The authors (Aliyu et al., 2019) have used edge-based segmentation techniques followed by labelling which has led to a detection of sickle cells with accuracy 88% and (Sen et al., 2021) have worked with watershed segmentation which has improved the accuracy to 90%. The proposed work using thresholding technique has given a reasonably good accuracy since the threshold was identified using the preprocessing techniques. After identifying the right threshold, the cells were easily classified as circular, elongated and others.

4. Conclusion and future works

Based on the results of the cell classification process, it was determined that circular cells correspond to healthy red blood cells, while elongated cells are indicative of sickle cells. However, there remains a degree of ambiguity regarding the classification of cells designated as "others" and their possible association with sickle cells. In terms of performance evaluation, the proposed method demonstrated a high accuracy rate of 94.907% in correctly classifying the cells, suggesting that it has potential as a reliable tool for identifying sickle cells in blood samples. The future work could involve exploring the potential application of this method in diagnosing sickle cell anemia in clinical settings.

REFERENCES

Acharya, V. & Prakasha, K. (2019) Computer Aided Technique to Separate the Red Blood Cells, Categorize them and Diagnose Sickle Cell Anemia. *Journal of Engineering Science & Technology Review*. 12(2), 67-80. doi:10.25103/jestr.122.10.

Aliyu, H.A., Razak, M.A.A. & Sudirman, R. (2019, November) Segmentation and detection of sickle cell red blood image. In *AIP Conference proceedings, November 2019*. 2173 (1). AIP Publishing.

Arishi, W.A., Alhadrami, H.A. & Zourob, M. (2021) Techniques for the detection of sickle cell disease: a review. *Micromachines*. 12(5), 519. doi: 10.3390/mi12050519.

Bhatt, M. & Prabha, S. (2015) Detection of abnormal blood cells using image processing technique. *International Journal of Electrical and Electronics Engineers (IJEEE)*. 7(1), 89-94.

Gonzalez-Hidalgo, M., Guerrero-Pena, F.A., Herold-García, S., Jaume-i-Capó, A. & Marrero-Fernández, P.D. (2014) Red blood cell cluster separation from digital images for use in sickle cell disease. *IEEE journal of biomedical and health informatics*. 19(4), 1514-1525. doi:10.1109/JBHI.2014.2356402.

Hegde, R.B., Prasad, K., Hebbar, H. & Sandhya, I. (2018) Peripheral blood smear analysis using image processing approach for diagnostic purposes: A review. *Biocybernetics and Biomedical Engineering*. 38(3), 467-480.

Kiruthika, V., Vallikannu, A.L. & Vimalarani, G. (2021, September) Automatic Sickle Cell Anaemia Detection Using Image Processing Technique. In *International Conference on Micro-Electronics and Telecommunication Engineering*. Singapore, Springer Nature Singapore. pp. 281-288.

Navya, K.T., Prasad, K. & Singh, B.M.K. (2022) Analysis of red blood cells from peripheral blood smear images for anemia detection: a methodological review. *Medical & Biological Engineering & Computing*. 60(9), 2445-2462. doi: 10.1007/s11517-022-02614-z.

Petrović, N., Moyà-Alcover, G., Jaume-i-Capó, A. & González-Hidalgo, M. (2020) Sickle-cell disease diagnosis support selecting the most appropriate machine learning method: Towards a general and interpretable approach for cell morphology analysis from microscopy images. *Computers in Biology and Medicine*. 126, 104027. doi: 10.1016/j.compbiomed.2020.10402.

Rakshit, P. & Bhowmik, K. (2013) Detection of abnormal findings in human RBC in diagnosing sickle cell anemia using image processing. *Procedia Technology*. 10(5), 28-36. doi:10.1016/j.protcy.2013.12.333.

Sen, B., Ganesh, A., Bhan, A., Dixit, S. & Goyal, A. (2021) February. Machine learning based Diagnosis and classification of Sickle Cell Anemia in Human RBC. In *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 4-6 February 2021*. IEEE. pp. 753-758.

Vicent, M., Simon, K. & Yonasi, S. (2022) An algorithm to detect overlapping red blood cells for sickle cell disease diagnosis. *IET Image Processing*. 16(6), 1669-1677. doi.org/10.1049/ipr2.12439.

Priyadharsini RAVISANKAR is working as Associate Professor in department of Computer Science and Engineering, Sri Sivasubramaniya Nadar College of Engineering, Tamil Nadu. She received her Ph.D. in Information and Communication Engineering from Anna University, Chennai. She has more than 20 years of teaching experience. Her areas of research interest include underwater acoustic image processing, Computer Vision and Data analytics. She has published and presented her research works in reputed journals and conferences.

Beulah ARUL is an Assistant Professor in the Department of Computer Science and Engineering with over 18 years of experience in teaching. She completed her Ph.D. from Anna University, Chennai, in 2021 under the Faculty of Information and Communication Engineering. Beulah Arul began her career as a lecturer at the University College of Engineering, Thiruvananthapuram in 2002. After completing her M.Tech, she joined SSN in June 2007 as a lecturer in the Department of Computer Science and Engineering. Dr. Beulah Arul's research interests include Image Processing, Computer Vision, Machine Learning and Deep Learning Techniques. She has been successful in guiding several internally funded projects and her work has been recognized through numerous publications in reputed international journals and conferences.

Srivarsha ELANGO is a Software Development Engineer. She obtained her Bachelor's Degree in Computer Science and Engineering from Sri Sivasubramaniya Nadar College of Engineering, Tamil Nadu. She is currently working as a Software Development Engineer at PayPal, Chennai. Her areas of research include Data Analytics, Machine Learning and Web Development.