Sickle Cell Anemia Diagnosis Using Microscopic Images

Sukhada Aloni[1*], Pravin Adivarekar[2], Mayuri Jain[3], Sachin Takmare[4], Rahul Ambekar[5]

Abstract

Sickle cell anemia is a red blood cell disorder that causes a shortage of oxy- gen due to the sickling of red blood cells in the body. Sickle cell disease is a genetic disorder and typically remains undetected during a human life- time. We have developed a novel image processing-based system that can detect sickle cell anemia from brightfield microscopic images. An input image is passed through three pipelines that perform a variety of distance image processing operations such as top-hat filtering, local adaptive thresholding, morphological erosion, Canny edge detection, contour extraction, etc. The outputs of the three pipelines are then combined using an OR operation and passed through an area-based filtering operation to identify sickle cells and healthy cells. We have trained and evaluated our model on a dataset sourced from Kaggle, public datasets, and images of RBCs collected by us. We were able to achieve accuracy and precision of 96.46 % and 94.82 %, respectively, on the dataset. We could classify sickle cell anemia disease as well as trait patients. We expect that our proposed algorithm can be used in remote locations where a fast diagnosis is needed in resource-limited settings.

Keywords: microscopic images, sickle cell anemia, image processing

DOI Number: 10.14704/Nq.2022.20.17.Nq88024 **Neuroquantology 2022; 20(17):178-184**

1. Introduction

Sickle cell anemia is a disorder caused due to mutations in the β globin gene Nongbri et al. (2017). It results due to a mutation in haemoglobin where normal haemoglobin (HbA) is mutated with HbS. When a person is subjected to stress or higher altitude, the RBCs in the body become sickle-shaped and cause a clot in the vascular system Gladwin (2016). These Vasco occlusions cause joint pains, bulges on the skin, and sometimes even damage the spleen causing extra bleeding in the liver and could be fatal. The red blood cells

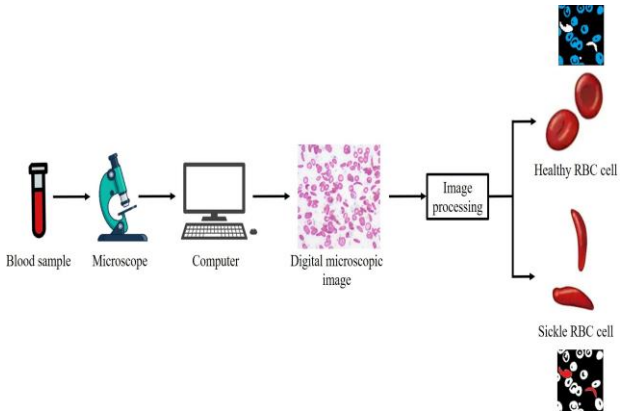

Figure 1: Concept diagram of the proposed sickle cell detection method. The blood sample of a patient is viewed under a microscope. The images of the RBCs visible through the microscope are captured. These digital microscopic images are offered as input to our method. The input image is processed through three different pipelines comprising of distinct image processing operations. The output of all three pipelines is combined to detect sickle cells in a given image. Our proposed system classifies each RBC as either a healthy RBC or a sickle cell.

***Corresponding Author:-** Sukhada Aloni
Address: [1*,2,3,4,5] Computer Engineering A. P. Shah Institute of Technology, Thane, 400615, Maharashtra, India
Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

which are sickled in shape cannot carry oxygen in the body Sahu et al. (2015). This lack of oxygen causes weakness in the person and sometimes may lead to multiple organ failures. Sickle cell is seen when a person goes to places where there is a lack of oxygen such as mountains or closed rooms without ventilation, etc Cymerman and Rock (1994). Normal healthy red blood cells are very flexible and can easily pass through the vessels. When sickled the cell becomes rigid and less deformable and gets stuck into the blood vessels. There is no cure for sickle cell anemia, except bone marrow transplant de la Fuente et al. (2019). But early knowledge about having sickle cell anemia can avoid its transfer via genes so that this disorder is not passed on hereditarily. Signs of sickle cell anemia are generally found only in grown-up children. Typical cell lives around 120 days but sickle cell can only last up to 10 to 20 days Neumayr et al. (2019). Hence patient might feel a shortage of red blood cells in the body, causing anemia. Sickle cell anemic patient generally suffers from pain crisis, and this crisis can last from a few hours to a few days Tawfic et al. (2014). Sickle cell anemia can cause swelling to the hands and feet and even frequent infection. Anemia can cause delayed growth and

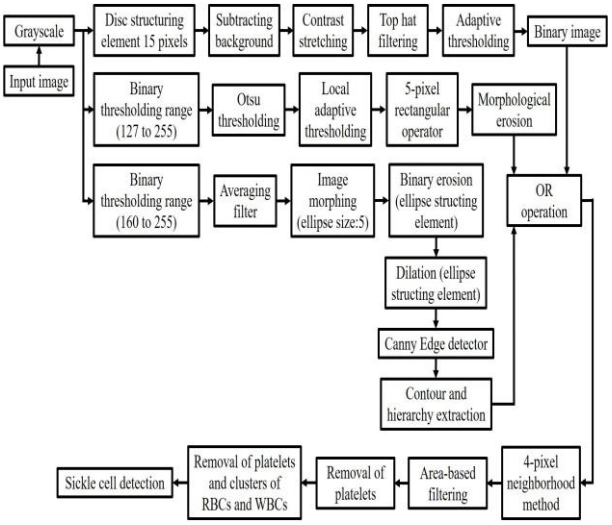

Figure 2: Block diagram detailing each operation involved in detecting sickle cell from a given digital microscope image. After the input image is converted to grayscale, it is passed through three different pipelines. The first pipeline uses a disc structuring element to remove the background. Then contrast stretching, top-hat filtering, and adaptive thresholding are used to generate a binary image. In the second pipeline, binary, Otsu, and local adaptive thresholding are consecutively performed on the input grayscale image. Finally, a binary image is generated after using a rectangular operator and morphological operator. In the third pipeline, binary thresholding, averaging filter, image morphing, binary erosion, dilation, Canny edge detection, and contour extraction is used to generate a binary image. The binary image outputs of all three pipelines are ORed and area-based filtering is applied to detect sickle cells.

eyesight problems. Hence detection of sickle cell anemia is very important and the earlier the detection the better it is.

There are generally two types of sickle cell disorders, Sickle Cell Disease (SCD) and Sickle Cell Trait (SCT) Creary et al. (2007). The bloodstream of a person suffering from SCD consists of sickle cells along with regular healthy red blood cells. Such patients suffer from all the conventional effects of sickle cell anemia. Patients whose genes carry the trait of sickle cell are said to be suffering from SCT Alvarez et al. (2015). Such patients do not suffer from the typical effects of sickle cell anemia, but when they are in low oxygen environments sickle cells are found to be in their bloodstream. In

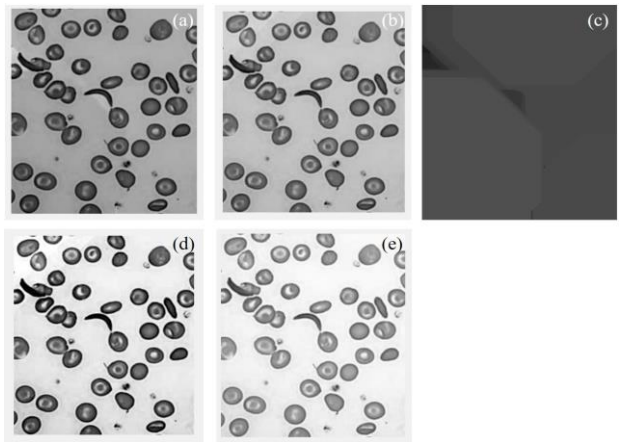

Figure 3: (a) Input image. (b) Output of averaging filter. (c) Output after image morphing. (d) Output after contrast stretching. (e) Output after the local adaptive threshold.

regular environments with sufficient oxygen, sickle cells were not found in the bloodstream of patients suffering from SCT. If two people suffering from SCT have an offspring, then there is a high chance that the child would suffer from SCD Gallo et al. (2010). If a person suffering from SCT has a child with a person not suffering from SCT then there is a high chance that the child would

suffer from either SCT or SCD.

In the current manuscript, we are proposing a novel image processing algorithm to detect sickle cell anemia from bright-field microscopic images. This method will help diagnose the sickle cell disease or trait at a point of care in a remote location using portable microscopes and single-board computers. We make use of a combination of three parallel image processing pipelines that leads us to robust object detection for further geometric shape classification, as shown in Fig. 1. The sickle cells are classified based on their

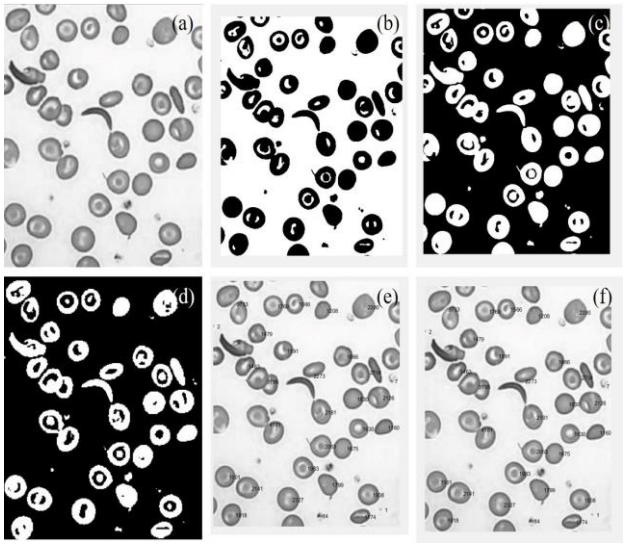

Figure 4: (a) Output after top hat filtering. (b) Output after binarization. (c) Inverse image after binarization. (d) Output after removal of all cells touching the edges of the image. (e) Output after area computation. (f) Final output after performing OR operation on the binary image outputs of all three algorithms.

roundness. For normal patients, the roundness distribution is single-peaked whereas for sickle cell anemia patients the roundness distribution is bimodal.

2. Literature Review

Watershed segmentation has been heavily used for analyzing overlapping red blood cells (RBC). But its main drawback is the occurrence of over-segmentation. To address this issue, Algailani and Hamad (2018) has pro- posed an improved watershed segmentation technique for the detection of sickle cell disease. They have found that the effectiveness of the watershed algorithm can be increased by performing it on an image that has been previously filtered by non-local means of denoising. They were able to achieve an accuracy of 93.21 %. To assess the current stage of sickle cell anemia, it is necessary to accurately count the number of sickle cells in the patient's blood-stream. Fadhel et al. (2017) has proposed two methods for enumerating sickle cells using Circular Hough Transform (CHT) and Watershed Segmentation (WS). Their system classifies an image of RBCs in two categories: abnormal and normal based on extracted shape characteristics. According to their experiments, CHT was able to provide more accurate and faster results for sickle cell detection than WS.

Many deep learning-based models have been proposed to detect sickle cell anemia from a microscopic image of RBCs. Such models gene-rally classify an input image into three categories, namely, normal cells, sickle cells, and other components of the bloodstream. To effectively train deep learning models, a large dataset is required. Such datasets are not available for training a deep learning model to accurately detect sickle cells in any given image of RBCs. Alzubaidi et al. (2020) has addressed the lack of training data by employing transfer learning to utilize a trained deep learning model for sickle cell detection. Transfer learning uses a deep learning model that was efficiently trained on a large data-set of some domain and then training that model again on a smaller dataset of an unrelated domain. When the source domain of a trained model is very different than the target domain, the perfor-mance of the model is degraded at times. Hence, Alzubaidi et al. (2020) have used employed trans-fer learning for a deep learning model that had medical images as the source and target domain. Data augmentation was used to increase the size of their dataset. They were able to achieve an accuracy of 99.54 % on the erythrocytesIDB dataset, along with 98.87 % accuracy on a dataset that they had compiled. Delgado-Font et al. (2020) have developed an sys- tem for counting red blood cells (RBCs) from peripheral blood smear images. Chan-Vese active contour model was used to high-light the objects of inter- est. Circular shape factor (CSF) and elliptical shape factor (ESF) were used to classify an RBC into three categories: healthy,

elongated, or some other disorder. They have utilized the erythrocytesIDB dataset for training and evaluating their system. They were also able to achieve a precision value of 0.96, 0.98, and 0.71 for normal, elongated, and other disorder-affected RBCs, respectively.

Barpanda (2013), have developed an image processing algorithm to automatically detect sickle-cell present in thin blood smears. They have used a charge-couples device camera to acquire images from a light microscope and used priori knowledge of the classification problem. They collected the red blood cell microscopic slide samples from IG Hospital, Rourkela to implement a clustering-based segmentation technique for identifying sickle cells and red blood cells. They were able to achieve an execution time of around 10 seconds for ten Gray-Level Co-Occurrence Matrices (GLCM) inputs. As traditional methods of analyzing RBCs are error-prone and inefficient, Fadhel et al. (2020) has proposed a neural network model that classifies the input image of a digital microscope in three categories: normal, abnormal, and other blood content. They have improved the efficiency of their system by executing it on an FPGA board (Altera DE2 Cyclone II), enabling them to use the parallel processing capabilities of the board. They were able to achieve an accuracy of 87.15 % on a dataset of 202 images sourced from public websites. Many systems for automated sickle cell detection have been proposed, but very few have claimed to detect thalassemia. Hence, Sharma et al. (2016) has proposed a system that can detect both the disorders. Their system performs a variety of image processing-related operations on an in- put image of RBCs, such as applying a median filter, marker-controlled water segmentation, morphological feature extraction, etc. The features extracted after all these operations are then used to train a K-nearest neighbor (K- NN) model. The model classified an input feature vector in three classes: sickle cells, dacrocytes, and elliptocytes. These cells are mainly responsible for sickle cell anemia and thalassemia. They were able to achieve accuracy and sensitivity of 80.6 % and 87.6 %, respectively.

Xu et al. (2017) has proposed an effective method for RBC shape quantification and classification. They have used automatic hierarchical RBC extraction to identify regions of interest (ROI) such as RBCs. Then touching RBCs are separated by employing an improved random walk method that is based on automatic seed generation. RBC patch-size normalization was used to ensure that all clusters of RBCs are of uniform and similar sizes. Finally, a deep convolutional neural network (DCNN) was used to classify RBCs into three categories: discocytes, elongated, and oval-shaped RBCs. They have used 700 images of single-cell RBCs obtained from eight patients for training and testing their system. They were able to achieve a high-recall rate of 93.8 %. Yeruva et al. (2021) has proposed a Multi-Layer Perceptron (MLP) to detect sickle cells effectively with high precision. Red Blood Cells (RBC) are classified into three classes, normal, sickle cell, and Thalassemia. They have utilized a dataset obtainned from the Thalassemia and Sickle Cell Society located in India. The dataset consisted of 1387 images of digital microscopic RBCs with appropriate labels. They were able to achieve an accuracy of 96.04 %. They had also trained and tested various machine learning algorithms such as Support Vector Machine (SVM), K-Nearest Neighbour algorithm (KNN), logistic regression, etc.

Knowlton et al. (2015) have proposed an innovative attachment for commercial smartphones through which sickle cells could be detected in a sample with less than 1 micro-liter of the blood sample. The attachment was 3D printed and it consists of an LED to illuminate the sample, a lens to view the RBCs, along two permanent magnets that would levitate the RBCs in the sample. They have based their sickle cell detection system on the levitation patterns of healthy and sickle cell RBCs. They found out that with time, healthy RBCs levitate at the same height whereas sickle cells continue to sink lower.

3. Methodology

The original microscopic images are converted from RGB color planes to greyscale to detect the cells. In all there are three parallel processes were followed to ensure that none of the sickle cells in the visible field of view is missed during the

analysis, as shown in Fig. 2. In the first method, a structuring element of size 15 pixels disc is used to remove the background using image opening operation. The background removal is important because there were some cases where the light source was non-uniformly illuminating the image and hence subtracting a background image from the original image allowed us to achieve a clear output. Once the background is removed then the image is then contrast stretched within the saturation limits of 1 % and 99 %. This allowed us to use the full dynamic range of the image.

$$T_w(f) = f - f \circ b \quad (1)$$

where,
f represents an array of pixel values on which filtering has to be performed, b represents the structuring element.

We also tried top-hat filtering with a disc size of 15 pixels to achieve a similar result as that of background removal and contrast stretching. The formula used for top-hat filtering is mentioned in equation 1. To remove any further noise first the grayscale image is converted to binary using adaptive thresholding. All the smaller objects less than 50 pixels were removed from the binary image. In the second method, binary thresholding was performed after conversion of the image in grayscale. Thresholding within the range of 127 to 255 was performed on the image to binarize it. But in the case of non-uniform illumination, this thresholding fails. In such cases, we use Otsu thresholding. In some cases, because of the cells being in different planes, some of the cells were getting chopped off due to the Otsu's thresholding, in such cases, local adaptive thresholding was used. Here the thresh binary values were set automatically and roughly 10 % of the image dimension were considered as individual smaller image regions. Once the non-uniform illumination and non-planar cells issues were solved further morphological erosion was carried out on the image with 5-pixel rectangular operator. In the third method, the image is initially converted to binary using a threshold value of 160 and 255. Further, an averaging filter was applied with a kernel of size 5 and a value of 1/9 to each element in the kernel with a depth of - 1, filtering is applied onto the image. The entire image is then morphed using an ellipse shape of size 3 pixels. The image is then passed through a binary erosion operation with the ellipse structuring element followed by the dilation of the same structuring element. A Canny edge detector is then applied to the dilated image. Finally, all the contours and their hierarchy were extracted using the same approximation and erosion. Once the final image is obtained via all three methods all the images were ORed and then a combined binary image was used to find the objects using the 4-pixel neighborhood method. Once all the objects are determined area-based filtering is applied to remove small platelets (below 50 pixels) and all clusters of RBCs and WBCs (greater than 300 pixels). Typical RBC lies between 50 to 300 pixels at 40x magnification with a resolution of 512 X 512 pixels. In case resolution or magnification is changed then the values for platelet filtering (¡50 pixels) and overlapping RBC filtering or WBC filtering (¿300 pixels) need to be adjusted accordingly. Finally to determine if the cell is sickle or not we consider the circularity and aspect ratio of the measured object. The formula of circularity is mentioned in equation 2. For regular cells, the aspect ratio lies between 1 to 1.1 and circularity lies between 0.9 to 1.

$$Circularity = \frac{4 * \pi * Area}{Perimeter^2} \quad (2)$$

where,

Area represents the area of the object,
Perimeter represents the perimeter of the object.

4. Results and discussions

We have evaluated our model with a dataset that was sourced from Kaggle and The Cell Image Library, along with images captured by us using a digital microscope. The dataset consisted of 3000 images out of which 2100 images were used to train the system. 900 images of the dataset were earmarked for evaluating the model. For an example input image of RBCs, the outputs at various stages of processing are shown in Fig. 3 and Fig. 4. The sample input image is visible in Fig. 3 (a),

while the output of the averaging filter is shown in Fig. 3 (b). The outputs after image morphing, contrast stretching, and local adaptive thresholding are shown in Fig. 3 (c), Fig. 3 (d), and Fig. 3 (e), respectively. Fig. 4 (a), Fig. 4 (b), and Fig. 4 (c), display the results obtained after top-hat filtering, binarization, and inverse image after binarization. After removing all cells touching the edges of the image, the processed image is shown in Fig. 4 (d). After area computation is applied, the area of each RBC is calculated as shown in Fig. 4 (e). The final output after combining the outputs of all three pipelines is shown in Fig. 4 (e).

On the testing dataset, our method was able to achieve accuracy, sensitivity, specificity, and precision of 96.46 %, 94.18 %, 95.07 %, and 94.82 %, respectively. We have also been able to achieve a false-positive rate, a false- negative rate, and an F1 score of 0.0493, 0.0582, and 94.50 %, respectively. The proposed method has been compared with reported literature in Table 1. 3000 images used by us to train and test our model were the highest number of images used in comparison with the reported literature. Yeruva et al. (2021) has used the second largest dataset consisting of 1387 images. The highest accuracy of 99.54 % was achieved by Alzubaidi et al. (2020), followed by the second-highest accuracy of 94.64 %, achieved by our proposed method. The large difference in the accuracy obtained by us and by Alzubaidi et al. (2020) can be attributed to the number of images used. While Alzubaidi et al. (2020) used 629 images, we have employed 3000 images. Hence, even though our model didn't achieve the highest accuracy it can accurately detect more diverse types of sickle cells than the reported literature. Most of the proposed algorithms classify an RBC in three classes, normal, sickle cell, and other blood content such as platelets. Our method detects and removes

Table 1: Comparison of reported literature with the proposed algorithm. We have used the highest number of images while evaluating our model. Our model achieved an accuracy of 94.64 % which was the second-highest with Alzubaidi et al. (2020) achieving the highest accuracy of 99.54 %.

Method	Number of images	Methodology Watershed	Precision	Accuracy (in %)	Recall (in %)	Number of classes
Algailani *et al.*	-	segmentation Non-local filter	-	93.21	-	2
Alzubaidi *et al.*	629	Transfer Learning Chan-Vese active contour model	-	99.54	-	3
Delgado *et al.*	629	Circular and elliptical shape factor	98	-	-	3
Fadhel *et al.*	202	Neural network Parallel processing FPGA	-	87.15	-	3
Sharma *et al.*	-	Marker-controlled water segmentation Morphological feature extraction	-	80.6	-	3
Xu *et al.*	700	Automatic hierarchical RBC extraction	-	-	93.8	3
Yeruva *et al.*	1387	Multi-Layer Perceptron	-	96.04	-	3
Proposed Method	3000	Area-based and Shape-based filtering	94.82	94.64	-	2

platelets and other components present in the bloodstream in the processing stage. Hence when the outputs of three pipelines are analyzed, only RBCs are visible in the processed input image. By eliminating the other blood content our accuracy of sickle cell detection has increased.

5. Conclusions

We have proposed an algorithm for automated detection of sickle cells using the images of RBCs taken from digital microscopes. The images of the RBC were taken using bright field imaging. We have utilized area and shape- based filtering in combination with image processing techniques such as Otsu thresholding, top-hat filtering, bottom-hat filtering, etc. We have tested our proposed algorithm on a dataset of 3000 images, achieving accuracy and precision of 94.64 % and 94.82 %, respectively. Reported literature consists of detecting sickle cells through a single method, while we are combining the outputs of three pipelines for accurate and efficient detection. The accuracy of our method can be increased by applying staining to the images.

References

Algailani, H., Hamad, M.E.S., 2018. Detection of sickle cell disease based on an improved watershed segmentation, in: 2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), IEEE. pp. 1–4.

Alvarez, O., Rodriguez, M.M., Jordan, L., Sarnaik, S., 2015. Renal medullary carcinoma and sickle cell trait: a systematic review. Pediatric blood & cancer 62, 1694–1699.

Alzubaidi, L., Fadhel, M.A., Al-Shamma, O., Zhang, J., Duan, Y., 2020. Deep learning models for classification of red blood cells in microscopy images to aid in sickle cell anemia diagnosis. Electronics 9, 427.

Barpanda, S.S., 2013. Use of Image Processing Techniques to Automatically Diagnose Sickle-Cell Anemia Present in Red Blood Cells Smear. Ph.D. thesis.

Creary, M., Williamson, D., Kulkarni, R., 2007. Sickle cell disease: current activities, public health implications, and future directions. Journal of women's health 16, 575–582.

Cymerman, A., Rock, P.B., 1994. Medical problems in high mountain en- vironments. A handbook for medical officers. Technical Report. ARMY RESEARCH INST OF ENVIRONMENTAL MEDICINE NATICK MA.

Delgado-Font, W., Escobedo-Nicot, M., Gonz´alez-Hidalgo, M., Herold- Garcia, S., Jaume-i Cap´o, A., Mir, A., 2020. Diagnosis support of sickle cell anemia by classifying red blood cell shape in peripheral blood images. Medical & biological engineering & computing 58, 1265–1284.

Fadhel, M.A., Al-Shamma, O., Alzubaidi, L., Oleiwi, S.R., 2020. Real-time sickle cell anemia diagnosis based hardware accelerator, in: International Conference on New Trends in Information and Communications Technology Applications, Springer. pp. 189–199.

Fadhel, M.A., Humaidi, A.J., Oleiwi, S.R., 2017. Image processing-based di- agnosis of sickle cell anemia in erythrocytes, in: 2017 Annual Conference on New Trends in Information & Communications Technology Applica- tions (NTICT), IEEE. pp. 203–207.

de la Fuente, J., Dhedin, N., Koyama, T., Bernaudin, F., Kuentz, M., Karnik, L., Soci´e, G., Culos, K.A., Brodsky, R.A., DeBaun, M.R., et al., 2019. Haploidentical bone marrow transplantation with post-transplantation cyclophosphamide plus thiotepa improves donor engraftment in patients with sickle cell anemia: results of an international learning collaborative. Biol- ogy of Blood and Marrow Transplantation 25, 1197–1209.

Gallo, A.M., Wilkie, D., Suarez, M., Labotka, R., Molokie, R., Thompson, A., Hershberger, P., Johnson, B., 2010. Reproductive decisions in people with sickle cell disease or sickle cell trait. Western Journal of Nursing Research 32, 1073–1090.

Gladwin, M.T., 2016. Cardiovascular complications and risk of death in sickle-cell disease. The Lancet 387, 2565–2574.

Knowlton, S., Sencan, I., Aytar, Y., Khoory, J., Heeney, M., Ghiran, I., Tasoglu, S., 2015. Sickle cell detection using a smartphone. Scientific reports 5, 1–11.

Neumayr, L.D., Hoppe, C.C., Brown, C., 2019. Sickle cell disease: current treatment and emerging therapies. Am J Manag Care 25, S335–43.

Nongbri, S.R.L., Verma, H.K., Lakkakula, B.V., Patra, P.K., 2017. Presence of atypical beta globin (hbb) gene cluster haplotypes in sickle cell anemia patients of india. Revista brasileira de hematologia e hemoterapia 39, 180–182.

Sahu, M., Biswas, A.K., Uma, K., 2015. Detection of sickle cell anemia in red blood cell. A. International Journal of Engineering and Applied Sciences (IJEAS) 2.

Sharma, V., Rathore, A., Vyas, G., 2016. Detection of sickle cell anaemia and thalassaemia causing abnormalities in thin smear of human blood sample using image processing, in: 2016 International Conference on Inventive Computation Technologies (ICICT), IEEE. pp. 1–5.

Tawfic, Q.A., Faris, A.S., Kausalya, R., 2014. The role of a low-dose ketamine-midazolam regimen in the management of severe painful crisis in patients with sickle cell disease. Journal of pain and symptom manage- ment 47, 334–340.

Xu, M., Papageorgiou, D.P., Abidi, S.Z., Dao, M., Zhao, H., Karniadakis, G.E., 2017. A deep convolutional neural network for classification of red blood cells in sickle cell anemia. PLoS computational biology 13, e1005746.

Yeruva, S., Varalakshmi, M.S., Gowtham, B.P., Chandana, Y.H., Prasad, P.K., 2021. Identification of sickle cell anemia using deep neural networks. Emerging Science Journal 5, 200–210.