Received: 21 July 2021 | Revised: 10 January 2022 | Accepted: 22 January 2022

IET Image Processing

DOI: 10.1049/ipr2.12439

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH PAPER

An algorithm to detect overlapping red blood cells for sickle cell disease diagnosis

Mabirizi Vicent[1] | **Kawuma Simon**[2] | **Safari Yonasi**[3]

[1] Department of Information Technology and Computer Science, Kabale University, Kabale, Uganda

[2] Department of Software Engineering, Mbarara University of Science and Technology, Mbarara, Uganda

[3] Department Information Technology, Mbarara University of Science and Technology, Mbarara, Uganda

Correspondence
Mabirizi Vicent, Department of Information Technology and Computer Science, Kabale University, Kabale, Uganda.
Email: vmabirizi@kab.ac.ug

Funding information
International Conference on Computational Intelligence: Modeling, Techniques and Applications, Grant/Award Number: CIMTA-2013

Abstract

In Africa, Uganda is among the countries with a high number of babies (20,000 babies) born with sickle cell, contributing between 6.8% of the children born with sickle cell every year worldwide and approximately 4.5% of the children born with hemoglobinopathies worldwide. It is estimated that by 2050, sickle cell cases will increase by 30% if no intervention is put in place. To facilitate early detection of sickle cell anaemia, medical experts employ machine learning algorithms to detect sickle cell abnormality. Previous research revealed that algorithms for recognizing shape of a sickle cell from blood smear by fractional dimension, cannot detect sickle cells if applied on blood samples containing overlapping red blood cells. In this research, the authors developed an algorithm to detect overlapping red blood cells for sickle cell disease diagnosis. The algorithm uses canny edge and double threshold machine learning techniques and takes overlapping red blood cells images as inputs to detect if these cells are sickle cell anaemic. These images have a scale magnification of (200×, 400×, 650×) pixel taken using a microscope. The algorithm was tested on a total of 1000 digital images and the overall accuracy, sensitivity and specificity were 98.18%, 98.29% and 97.98% respectively.

1 | INTRODUCTION

Blood is an important fluid in the human body, with the red blood cell (RBC) being the prime component that contains the haemoglobin responsible for gaseous exchange. The normal shape of a RBC is biconcave, its size is between 6.8 and 7.8 μm in diameter and between 2 and 2.5 μm in thickness [1]. However, due to certain disorders such as sickle cell, this shape can be deformed into an ovalocyte (C shape) thus resulting in sickle cell anaemia [2].

Sickle cell is a haemoglobinopathy that is hereditary and characterized by the presence of structurally abnormal haemoglobin [3]. It is estimated that if both parents are sickle cell disease carriers, there exist 25% chances of the baby they produce inheriting the trait from both parents [4]. As a result, thousands of children are born with sickle cell anaemia worldwide.

Research conducted in 2016 revealed that Uganda is ranked among the countries with the highest number of babies born with sickle cell anaemia with an estimation of 15,000 children which contributes 6.8% of the 220,000 children born with sickle cell every year worldwide and approximately 4.5% of children born with hemoglobinopathies [5]. These statistics are higher compared to the 2014/2015 survey where 12,979 cases of sickle cell among babies were recorded in Uganda [6]. Furthermore, studies estimated a 30% increase of sickle cell anaemia by 2050 worldwide if no intervention is done to reduce the spread of sickle cell anaemia [2].

As a result of these statistics, there has been a global concern drawing more attention towards reducing sickle cell cases through a number of strategies. Clinically, sickle cell anaemia can be treated by born marrow transplantation and curative therapies such as gene therapy, gene editing. However, these treatments are very expensive thus, majority of individuals in low developed countries like Uganda cannot afford this [7].

In Uganda, interventions like the Kabaka run was launched in 2017, aiming at raising funds to support both children and adults born with sickle cell anaemia [8]. Other approaches such as the sickle cells screening campaigns are enrolled across the country. However, these interventions are not effective. For instance, the Kabaka run supports those with the infection but does not

prevent its manipulation. Screening is hard to access because a haemoglobin electrophoresis machine is too expensive to be enrolled in every health facility across the country.

Recently there has been an impressive effort in Uganda to reduce sickle cell cases through newborn screening of sickle cell disease by a joint effort of the HIV/sickle programs [9].

The conventional method is to test a patient's blood sample for sickle cell anaemia however, this process requires a lot of time and medical expertise to conduct the diagnosis, interpret the diagnosis reports, and sometimes it is problematic to identify or observe the cells. As a solution to mitigate this problem, researchers have adopted the use of machine learning algorithms to diagnose sickle cell. For example, image segmentation for feature extraction was utilized by Manoj Kumar Sahu et al. [10] to develop a method to recognize the shape of sickle cells from blood smear images by fractional dimension. Furthermore, in a study by R. Desai and H.G. Virani [11], a circular Hough transform technique was used to develop a method for the detection of RBC from blood smears.

From existing literature [12–14], it was identified that these algorithms have a common limitation of not being able to detect sickle cell anaemia in overlapping red blood cells, that is, they give inaccurate results if applied on images containing overlapping red blood cells.

In order to address this limitation, this research developed an algorithm for diagnosing sickle cell anaemia in overlapping red blood cells.

1.1 | The red blood cells

Red blood cells are also known as erythrocytes which are fluids found in human blood responsible for gaseous exchange [15]. These cells are developed in the born marrow and circulate for about 100–120 days in the body before their components are recycled, with each circulation taking about 60 s [16]. It is estimated that 40–45% of the total volume of blood in the human body is red blood cells and constitutes 84% of the cells [15, 17].

These cells are flexible, deformable and are able to adhere to other cells, unlike other vertebrates, RBCs in mammals have no nucleus instead they have a cell membrane [18, 19] which is responsible for a number of roles depending on its composition. The cell membrane is composed of three layers including the glycocalyx, lipid bilayer and membrane skeleton [20].

A normal red blood cell is assessed by considering five major factors including the size, shape, color or hemoglobin content and presence of intracellular. Often there are biological terms used to describe different sizes of red cells including normocyte (normal - cell less than 8.2 μm) and microcyte (cell size less 7.2 μm) [12].

It is biologically proved that one blood cell can abstract another. Such cells are referred to as overlapping cells as shown in Figure 1a.

1.2 | Sickle cell anaemia

Sickle cell anaemia is one of the commonest sickle cell diseases which is a group of blood disorders that affects red blood cells

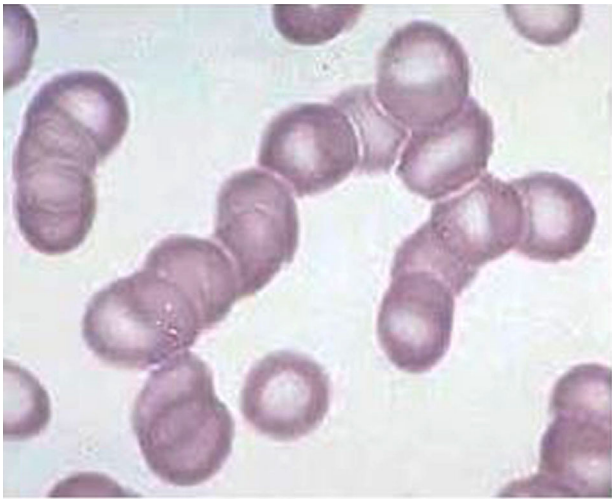

FIGURE 1 Overlapping red blood cells adopted from [20]

by deforming their shape into a sickle shape [5] thus affecting oxygen transformation in the body system [21].

1.3 | Sickle cell screening and diagnosis

Sickle cell screening is an important aspect of health care especially before choosing a marriage partner since it helps to detect the presence of sickle cell traits in the blood. In a situation where both partners are diagnosed with the trait, they are advised not to intermarry so as to reduce the chances of sickle cell manipulation. Thus, sickle cell screening is very vital [4].

Sickle cell screening has been proven to reduce the number of sickle cells diagnosed cases reported every year among new born. However, screening results are not confirmatory tests since they only detect the presence of sickle cell traits in the red blood cells [22]. Therefore, if test results are positive, a second test known as haemoglobin electrophoresis is carried out to determine the blood condition. The presence of two abnormal haemoglobin genes will call for sickle cell diagnosis, this test is carried out from the laboratory by a trained technician with the aid of a haemoglobin electrophoresis machine using blood samples. This study used image processing techniques to analyse overlapping red blood cells for sickle cell detection, unlike the existing algorithms which focus on non-overlapping red blood cells.

1.4 | Research objective and questions

The goal of this study was to detect whether overlapping red blood cells are sickle cell anaemia. With this in mind, researchers formulated the following research questions to guide this study as follows:

RQ1: How can sickle cell diagnosis bio markers be extracted from overlapping red blood cell digital images? A total of 1000 blood smear images were obtained from the haematology atlas digital library [23].

These images had a scale magnification of ×200, ×400, ×650 pixels taken using a photo microscope and a Nikon camera [24].

These images had too much noise including; other blood cells like white blood cells, platelets and other unwanted fluids therefore, water threshold data preprocessing technique was used to remove the unwanted blood components and later canny edge detection—a machine learning image segmentation technique [25] was applied to extract biomarkers used to classify results.

RQ2: How can an algorithm for detecting presence of sickle cell anaemia in overlapping red blood cells be developed? An algorithm was developed using MatLab programming language. The algorithm uses digital images of overlapping red blood cells as inputs to detect presence of sickle cell anaemia.

RQ3: How can we evaluate the accuracy of an algorithm for detecting overlapping red blood cells? Algorithm evaluation was based on predefined (know) properties of a normal red blood cell including; cell diameter (6.8–7.8 μm), cell form factor (0.8 μm) and shape by assessing its accuracy, sensitivity and specificity.

2 | RELATED WORK

In this section, a symposium on how the current work relates to the previous work has been discussed. The previous study by Chintawar et al. [12] developed an image processing algorithm to automate the diagnosis of sickle-cells present in thin blood smear. The algorithm uses sample images acquired using a charge-coupled device camera connected to a light microscope. The obtained images are processed using clustering-based segmentation technique to identify red blood cells and sickle cells present on microscopic slides. Besides its limitation on overlapping red blood cells, the process of acquiring images is expensive in a certain environment where light microscopes cannot be accessed.

Saima Bala et al. [10] implemented an automatic detection technique to observe sickle cells in red blood cells through image splitting up. In their technique, Watershed segmentation technique was used for extracting red blood cells. This technique focused on non-overlapping red cells thereby neglecting overlapping red blood cells.

Bharam et al. [26] proposed a method for classifying healthy red blood cells and sickle cells using spatiotemporal analysis with a compact and low-cost 3D printed shearing interferometer. The method aimed at studying morphology and mechanical properties of red blood cells to identify sickle cells. In their prototype, a laser source, a microscope objective, a glass plate, and an imaging sensor were used to extract the cell membrane fluctuation by capturing its lateral and axial direction as features used in classification. A case-control approach was applied. In this case, eight consented sickle cell patients and six healthy control volunteers were recruited. The developed prototype was exposed to the imaged blood samples captured with compact DHM setup and a video containing hologram frames was recorded.

Results obtained after the analysis were presented in three different categories based on different cases including spatiotemporal based feature which yielded 78.00% accuracy, with a specificity of 81.33% and sensitivity of 74.67%. In case two, the morphology-based feature was considered. Results obtained in this case yielded 92.67% accuracy with a specificity of 96.00% and a sensitivity of 89.33%. Classification results obtained using both the spatiotemporal and the morphological-based features were 93.33% accurate, 100% specificity of 100% and 86.67% sensitivity. Considering results obtained in both cases, a combined test yielded better results compared to those where single features were considered.

Despite the less expertise needed in this method compared to the traditional haemoglobin electrophoresis, there is a noticeable limitation requirement of 2 h to process the data, costs for acquiring the machine, limited accessibility especially in developing countries with limited resources, and above all, it does not cater for overlapping red blood cells.

Begum et al. [27] proposed a method to identify sickle cell anaemia by sugar sequencing. The method was based on molecular analysis to extract genomic DNA from the patient body using the Qiagen genomic DNA extraction kit. Molecular analysis was done by sequencing of the entire region of Exon, part of the intron, promoter region and splice junction site of the beta globing gene, then polymerase chain reaction (PCR) method was used to amplify the amniotic fluid. Analysis was performed by automated capillary electrophoresis in the 3500-genetic analyser. The sequences obtained from the capillary electrophoresis were then aligned using the reference sequence of the beta-globin gene in the Seqscape sequence alignment software version 2.5.

The method was able to detect abnormal behaviour of sickle cell anaemia including lack of other complications apart from general body weakness. However, this process involves many complex steps which require specialists to interpret the finding as well as acquire expensive devices such as the Qiagen genomic DNA extraction kit thus making it inappropriate for an environment without the required apparatuses to carry out the analysis.

Sen et al. [28] proposed a machine learning technique to diagnose and classify sickle cell anaemia in human red blood cells. In their method, microscopic images were used to analyse the shape of a red blood cell. Later, Otsu thresholding was applied to perform segmentation while random forest, logistic regression, naïve Bayes, and support vector machine were applied to classify cell shapes into circular, elongated, and others. However, limiting the scope to only cell shape may affect the accuracy of results obtained since other determinant factors such as cell form factor and cell thickness are neglected in this experiment.

Laith et al. [29] proposed deep learning models for classifying red blood cells in microscopy images to aid in sickle cell anaemia diagnosis. The study aimed at addressing the issue of training data through optimizing model performance using transfer learning techniques.

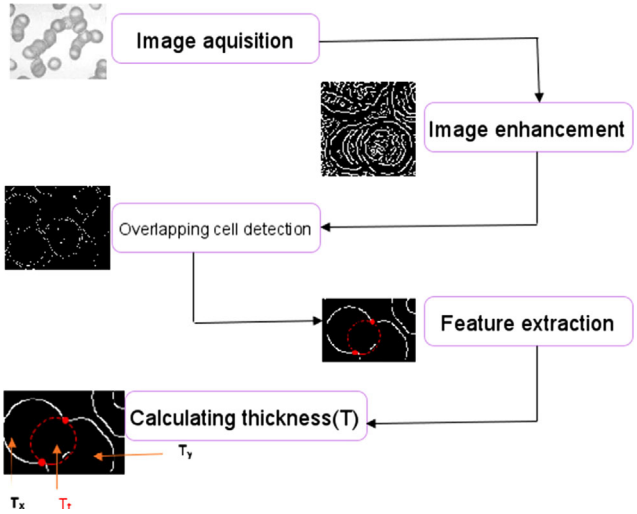

FIGURE 2 Research methodology block diagram

Seravanan et al. [30] utilized adaptive thresholding diagnostic techniques in their proposed method to detect sickle cell anaemia using microscopic blood images. The study mainly focused on classifying blood smear images. However, in their scope overlapping red blood cells were never catered for thus limiting diagnosis of sickle cell anaemia using the proposed method.

3 | METHODOLOGY

This section describes the methods used to achieve research objectives. In this research, researchers followed image processing techniques namely; image acquisition image enhancement overlapping cell detection feature extraction classification as illustrated in Figure 2b.

1. Image acquisition: This stage involved acquiring data sets used for this research. Red blood smear images were obtained from the haematology atlas library [25]. This data set contained a total of 1000 blood sample images with a scale magnification of ×200, ×400, ×650 pixels taken using a photo microscope and a Nikon camera [26].
2. Image enhancement: The obtained data set had too much noise including other blood cells like white blood cells, platelets and other fluids which were unwanted. Therefore, researchers performed data pre-processing techniques; grey scale conversion and data threshold to eliminate all the unwanted cells. As a result, only overlapping red blood cells were reserved cell as shown in Figure 3c.
3. Overlapping cells detection: The cleaned images from step 2 contained both overlapping and non-overlapping red blood cells thus, eliminating non overlapping cells was necessary. This was achieved by applying Canny edge detection mechanism. Canny edge detection is an edge detection operator that uses a multi-stage algorithm to detect a wide range of edges in images [31]. Canny edge follows the following step as illustrated in Figure 4d.

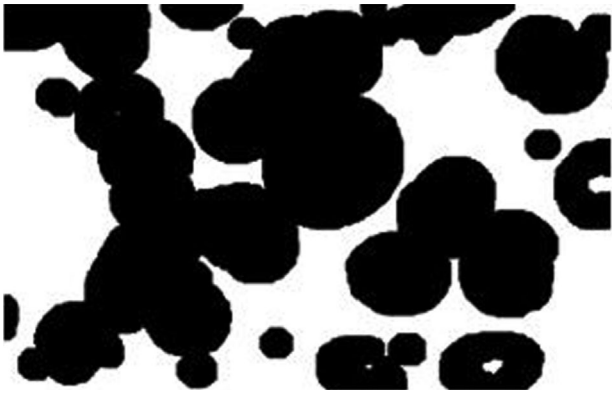

FIGURE 3 Processed image

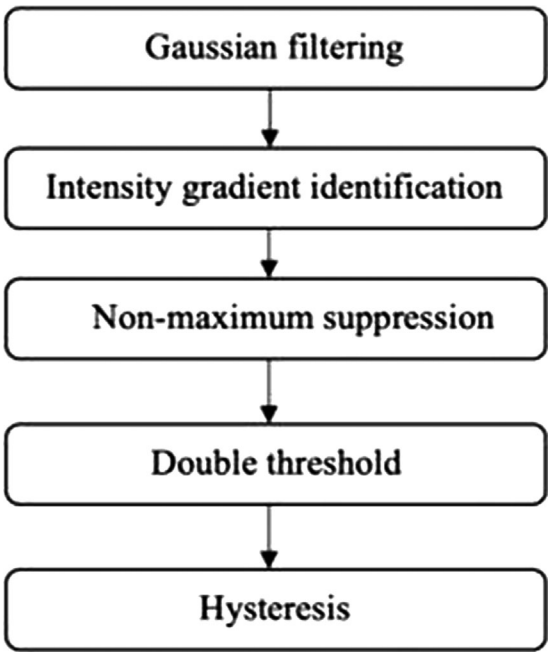

FIGURE 4 Canny edge detection operator steps

(i) Gaussian filtering: During data cleaning, some noise remained and this could affect the edge detection process by creating false edges thus, Gaussian filtering was applied to eliminate these false edges. Figure 5e shows smoothed data image after applying Gaussian filtering.
(ii) Intensity gradients identification: Image intensity was identified using four filters to detect horizontal, vertical and diagonal edges in the blurred image. The operator returned a value for the first derivative in the horizontal direction (Gx) and the vertical direction (Gy). From this, the edge gradient and direction were calculated using Equation (1).

$$G = \sqrt{Gx^2 + Gy^2} \quad (1)$$

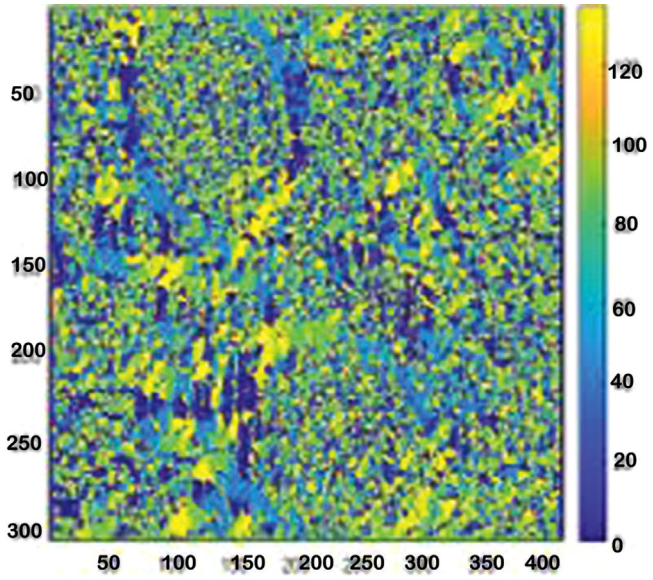

FIGURE 5 Noise-free image filtered by Gaussian filter

(iii) Non-maximum suppression: An edge thinning technique was applied to remove false response to edge detection, this was archived by finding pixel points with the sharpest change of intensity value. Each pixel of the image was; (i) compared in terms of the edge strength of the current pixel with the edge strength of the pixel in the positive and negative gradient directions. (ii) Checked if the edge strength of the current pixel is the largest compared to the other pixels in the mask with the same direction. For example, a pixel pointing in the *x*-direction would be compared to the pixel on the left and right in the horizontal axis, then the value would be preserved. Otherwise, the value is suppressed.

(iv) *Double threshold*: Application of non-maximum suppression provided a more accurate representation of real edges in an image. However, in some scenarios some unwanted pixels remained due to noise and colour variation. To cater for these false responses, filtering was carried out to remove pixels with weak gradient. This was achieved by adjusting the threshold high (*H*) and low (*L*) scale values to 0.175 and 0.075 respectively. In this case edge pixel whose value is higher than *H* value was marked as a strong edge pixel while that whose value is lower than *H* value and higher than *L* value was labelled as weak edge pixel and those whose value was lower than L value were suppressed. This stage provided a clear image ready for extraction of bio markers as shown in Figure 6f.

(v) Hysteresis: An edge tracking mechanism for strengthening edges with weak gradient to improve detection accuracy. During non-maximum suppression some of the edge pixels were falsely represented as strong ones, the application of double threshold exposed these edges thus, hysteresis was applied to strengthen these week edges to improve overlapping cells detection accuracy.

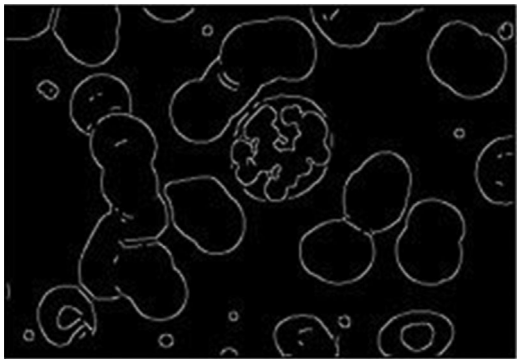

FIGURE 6 Dataset with detected edges of overlapping red blood cells

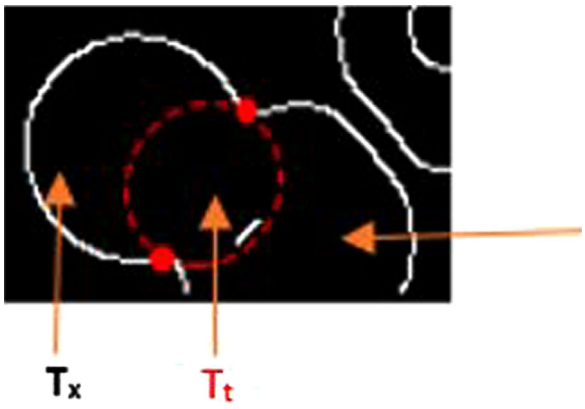

FIGURE 7 Block diagram showing a touching section of overlapping cells

4. Feature extraction: At this stage the algorithm identified and marked points where the two overlapping cells intersect. This was done by tracking the image brightness discontinuities through monitoring it's depth, surface orientation, changes in material properties (pixels: bit depth and dimension) and scene illumination. The identified points were then marked with red dots as shown in Figure 7g

T_x - cell x, T_t - intersection of cells x and y, T_y - cell y

5. Calculating area of regions T_x, T_t and T_y.

Having identified and marked points of intersection, area corresponding to regions A_x, A_t and A_y were calculated using the Matlab regionprops function. A_t value was used to obtain other bio-markers (thickness and volume) while cell form factor was calculated from the values of minor and major axis of overlapping region of the two cells as explained bellow.

3.1 | Calculating thickness of the touching cells

From the equation of sphere [32, 33];

$$T_t = \text{Volume} \div \text{Area} \quad (2)$$

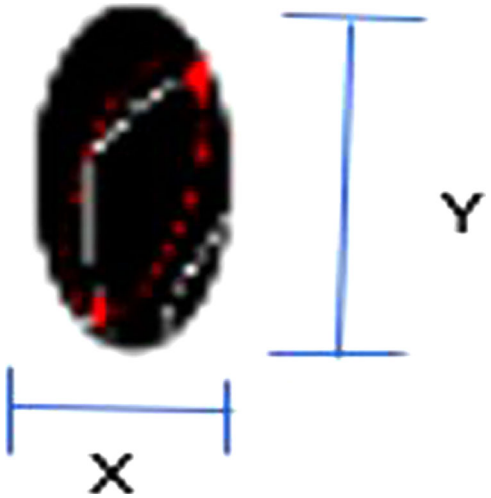

FIGURE 8 The major and minor axis of a red blood cell

Where:

$$\text{Area } (A) = 4\pi r^2 \tag{3}$$

[33].

$$\text{Volume (V)} = 4/3(\pi r^3) \tag{4}$$

Substituting Equations (3) and (4) in (2), cell thickness was calculated using:

$$Tt = 4\pi \left(A \div 4\pi\right) / 3\pi^2 \tag{5}$$

The process was repeated for the corresponding T_x and T_y values

A_t was then converted from pixel to micrometre, where $p_x = 264.58\,\mu$m.

A comparison of results obtained at T_t, T_x and T_y was done to draw conclusion whether the touching cells are sicklier or non-sicklier. This comparison was based on assumptions:

(i) $T_t = T_x + T_y$ (6)
(ii) $T_x \approx T_t - T_y$ (7)
(iii) $T_y \approx T_t - T_x$ (8)

3.2 | Calculating the form factor (*F*)

Cell form factor was calculated using Equation (6)

$$F = \frac{y}{x} \tag{6}$$

where x is the minor axis and y is the major axis as illustrated in Figure 8h.

4 | PROPOSED WORK

This section presents how results for RQ1, RQ2, and QR3 were obtained.

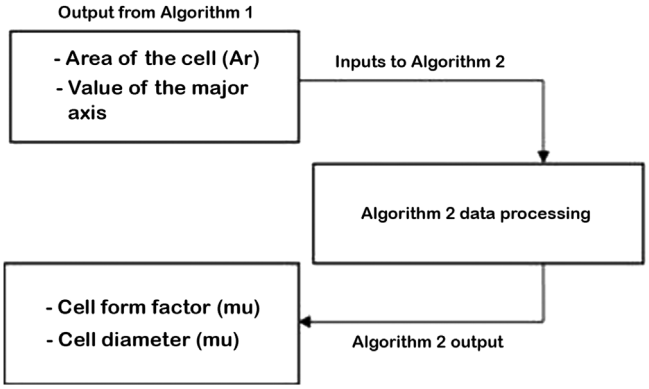

FIGURE 9 Outputs from Algorithms 1 and 2

4.1 | An algorithm for extracting sickle cell diagnosis bio-markers from red blood cell digital image

An array A containing images $x_1 : x_n$ was created and each image in the array was converted to grey scale then stored in directory D.

This was done in Step 1 of Algorithm A

4.2 | Sickle cell disease detection algorithm

Output from Phase 2 of Algorithm 1 are inputs for Algorithm 2 as shown in Figure 9h.

The algorithm uses Area T_t and Major axis M_x values as inputs, which are assigned to arrays A and B where y and m are values of A_t and M_x respectively. The output of this computation is the value of T_t and F_f calculated using equations *ix* and *x*.

5 | RESULTS AND DISCUSSION

5.1 | Classification of results

Results obtained were categorized based on whether an image has been diagnosed with presence of sickle cell anaemia or not. To achieve this, Support Vector Machine (SVM), Naïve bayes and Logistic regression machine learning classifiers were used.

These classifiers were used one at a time in a sequence of Naïve bayes, Logistic regression and Support vector machine. As result of applying the three classifiers, SVM yielded more accurate results compared to Naïve Bayes and Logistic regression.

Using SVM, out of 290 normal red blood cells (non-Sickler), 282 were correctly classified as normal cells while eight were falsely classified as abnormal cells by the algorithm. Of the 599 positive red blood cells (Sickler), 592 were correctly classified as abnormal cells and seven were incorrectly classified as normal cells as shown in Table 1.

ALGORITHM 1

Phase 1

$A = Array\ [x_0: x_n]$ where x is an image and n is size of the array

Input: A

Directory: D

Output: G_i (Gray scale image)

1. Create Image array A
2. For x in A read x
3. Convert x to gray scale
4. Output G_i
5. Else
6. "Warning: image file does not exist"
7. End for loop
8. Save G_i in D
9. Exit

Output G_i from phase 1 was used as input at level 2.

The algorithm inputs were filtered using $Filtered_X$ and $Filtered_y$ in Step 1 and 2 to obtain a clear view for edge detection. Then the output is assigned to an array A with indices labelled from 0 to n where n is the size of the array.

From steps 5–10 the algorithm filters detect an image in both x and y planes for both negative and positive values.

From steps 11 to 13 the algorithm applies canny edge detection techniques and the outputs were the values of Area A_t and major axis M_x of the touching section of the overlapping red blood cells as illustrated below.

Phase 2

High threshold value (T-High) = 0.175

Lower threshold value (T-Low) = 0.075

Input: G_i (Gray scale image from Algorithm 1)

Directory: D_p (Storing images with extracted features)

Output: A_i (Area of the touching section of the cells).

Filter for horizontal: $KG_x = [-1, 0, 1; -2, 0, 2; -1, 0, 1]$

Filter for vertical: $KG_y = [1, 2, 1; 0, 0, 0; -1, -2, -1]$ Converted value of A = Value of x converted by Gaussian filter:

1. $Filteredx$ = convert (A, KGx, T-Low, T-High)
2. $Filteredy$ = convert (A, KGy, T-Low, T-High)
3. filteredImage = atan2 ($Filteredx$, $Filteredy$)
4. b = filteredImage × $(180^0/\pi)$
5. A = array $[b_o: b_n]$
6. For $i = 1: A$
7. if $b(i) < 0$

   $b(i) = 360 + b(i)$
8. End if
9. End for loop
10. Non-Maximum Suppression of b
11. Hysteresis Thresholding of b
12. Marking points of touching objects in b
13. Extract Area A_t and Major axis M_x of the touching section of the cell b
14. Store A_t and M_x in D_p
15. Exit

ALGORITHM 2

Input: A_t

Input: M_x - Major axis

Directory: D_r - Storage for output. Output: F_f - Form factor

Output T_t - Thickness of the touching section of the cells

1. A = Array $[y_0: y_n] - y$ = Surface area of the A_t
2. B = Array $[m_0: m_n] - m$ = Value of major axis
3. For y in A
4. Read y
5. if y read then
6. $T_t = \pi(A[n] \div 4\pi)/3\pi r^2)$
7. $F_t = B[n]/2(\sqrt{A[n] \div 4\pi})$
8. Print T_t, F_f
9. Else
10. Print "Warning image not found"
11. End for loop
12. Store T_t, F_f in D_r
13. Exit

TABLE 1 Results obtained using support vector machine (SVM)

Laboratory results		Algorithm results using SMV	
Non sicklier cells	290	True negative	282
		False negative	7
Sicklier cells	599	True positive	592
		False positive	8

Results obtained using SVM classifier.

TABLE 2 Results obtained using Naïve bayes

Laboratory results		Algorithm results using naïve bayes	
Non sicklier cells	290	True negative	274
		False negative	16
Sicklier cells	599	True positive	578
		False positive	21

Results obtained using naïve bayes classifier.

Application of Naïve bayes and logistic regression classifier produced less accurate results compared to SVM as shown in Table 2 and 3 respectively.

5.2 | Evaluation of the developed algorithm

Evaluation of the algorithm was based on known properties of a normal red blood cell including; cell thickness (2–2.5μm), form factor (0.878 μm) color, and shape of the cell. This was done by comparing the algorithm accuracy, sensitivity and specificity

TABLE 3 Results obtained using logistic regression

Laboratory results		Algorithm results using logistic regression	
Non sicklier cells	290	True negative	279
		False negative	11
Sicklier cells	599	True positive	587
		False positive	12

Results obtained using logistic regression classifier.

with the confirmatory results obtained using haemoglobin electrophoresis machine. In this case;

Accuracy is defined as the quality of being correct and is calculated by the equation:

$$\mathrm{AC} = Tp + Tn / Tp + Tn + Fp + Fn \tag{6}$$

Sensitivity is defined as the measure of the proportion of actual positives that are correctly identified as such. For example, the percentage of sick people who are correctly identified as having the disease x. Sensitivity was calculated by the equation:

$$\mathrm{TP} = Tp / Tn + Fn \tag{7}$$

Specificity is defined as the measure of the proportion of actual negatives that are correctly identified as such for example the percentage of healthy people who are correctly identified as not having the disease x and is given by the equation:

$$\mathrm{TN} = Tn / Tn + Fp \tag{8}$$

where

Tp is True positive, *Tn* is True negative, *Fp* is False positive and *Fn* is False negative.

The overall accuracy, sensitivity and specificity of the classifier was 98.31%, 98.83% and 97.57% respectively.

5.3 | Discussion of results

From the results presented in Table 1, it is observed that the developed algorithm was able to detect overlapping red cells and diagnosed them for sickle cell anaemia. Classification of results clearly shows that the algorithm was in total agreement with the testing dataset up to 98% giving a percentage error of 2%. Furthermore, the results obtained reveals that not all overlapping red cells are sickle cells as stated in the existing algorithms [9–12].

6 | CONCLUSION AND FUTURE WORK

In this study we have assessed whether overlapping red blood cells are sickle cell anaemic and this was achieved through developing an algorithm to detect overlapping red blood cells for sickle cell disease diagnosis. Matrix laboratory (MATLAB) programming language was used to implement functionalities of the algorithm and tested on a total of 1000 red blood cell digital images obtained from haematology digital image library.

The algorithm was developed by utilizing machine learning techniques including canny edge detection to detect edges in digital images and double threshold technique for segmentation while results were classified using SVM, Naïve bayes and logistic regression classifiers. The overall accuracy, sensitivity and specificity of the algorithm verses classifier was 98.18%, 98.29% and 97.98%. respectively.

It was observed that:

- If images are taken from uncontrolled environment where light intensity is not controlled, the accuracy and sensitivity of results reduces due to much noise caused by either too much or low light intensity.
- The algorithm execution time increases with increase in the size of the data set there by increasing the time taken to produce results.

In a follow up study, future scholars can improve the developed algorithm by embedding image filtering features in the algorithm to support the utilization of digital images taken from a non-controlled environment.

It should be noted that the developed algorithm was tested using cell thickness, cell form factor, cell colour and cell shape as bio markers. Future scholars can consider other biomarkers including; gender, ethnicity and age to validate the accuracy of results on those bio-markers.

ACKNOWLEDGEMENTS

Faculty of Computing and Informatics (FCI) of Mbarara University of Science and Technology (MUST), Faculty of Information Technology and Computer Science (FoCLIS) of Kabale University (KAB), and International Conference on Computational Intelligence: Modeling, Techniques and Applications (CIMTA-2013): The authors gratefully acknowledge academic support of the FCI at MUST and FoCLIS at KAB towards this research study.

In a special way authors extend their appreciation to Mr. Mutumba Cosma for his courage, and financial support toward this research study.

CONFLICT OF INTEREST

The authors have no conflicts of interest to disclose.

DATA AVAILABILITY STATEMENT

The dataset used in this study are not publicly available due to privacy and ethical restrictions.

ORCID

Mabirizi Vicent https://orcid.org/0000-0001-8990-4003

REFERENCES

1. Walker, H. K., Hall, W. D., Hurst, J. W.: Peripheral blood smear–clinical methods: The history, physical, and laboratory examinations. Butterworths (1990)

2. Breakey, V. R., Harris, L., Davis, O., Agarwal, A., Ouellette, C., Akinnawo, E., Stinson, J.: The quality of information about sickle cell disease on the internet for youth. Pediatr. Blood Cancer 64(4), 881–891 (2017). https://doi.org/10.1002/pbc.26309
3. Rakshit, P., Bhowmik, K.: Detection of abnormal findings in human RBC in diagnosing sickle cell anaemia using image processing. Elsevier 10, 28–36 (2013). https://doi.org/10.1016/j.protcy.2013.12.333
4. MoH, Sickle cells disease on the rise in Uganda. http://library.health.go.ug/news/uganda-sickle-cells-disease-rise-uganda (2019). Accessed 09 June 2019
5. Green, N. S., Mathur, S., Kiguli, S., Makani, J., Fashakin, V., LaRussa, P., Lyimo, M., Abrams, E. J., Mulumba, L., Mupere, E.: Family, community, and health system considerations for reducing the burden of pediatric sickle cell disease in uganda through newborn screening. Global Pediatr. Health 3, 2333794X1663776 (2016)
6. Ndeezi, G., Kiyaga, C., Hernandez, A. G., Munube, D., Howard, T. A., Ssewanyana, I., Nsungwa, J., Kiguli, S., Ndugwa, C. M., Ware, R. E., et al.: Burden of sickle cell trait and disease in the Uganda sickle surveillance study (us3): a cross-sectional study. Lancet Global Health 4(3), e195–e200 (2016). https://doi.org/10.1016/S2214-109X(15)00288-0
7. Romero, Z., DeWitt, M., Walters, M. C.: Promise of gene therapy to treat sickle cell disease. Expert Opin. Biol. Ther. 18(11), 1123–1136 (2018)
8. Kabaka injects shs494m to end sickle cell disease, https://www.monitor.co.ug/uganda/news/national/kabaka-injects-shs494m-to-end-sickle-cell-disease-1825910 (2021). Accessed 25 Jan 2021
9. Hernandez, A. G., Kiyaga, C., Howard, T. A., Ssewanyana, I., Ndeezi, G., Aceng, J. R., Ware, R. E.: Operational analysis of the national sickle cell screening programme in the republic of uganda. Afr. J. Lab. Med. 10(1), a1303 (2021)
10. Chy, T. S., Rahaman, M. A.: Automatic sickle cell anemia detection using image processing technique. In: 2018 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE), Gazipur, Bangladesh, pp. 1–4 (2018)
11. Hegde, R. B., Prasad, K., Hebbar, H., Sandhya, I.: Peripheral blood smear analysis using image processing approach for di- agnostic purposes: A review. Biocybern. Biomed. Eng. 38(3), 467–480 (2018)
12. Bala, S., Doegar, A.: Automatic detection of sickle cell in red blood cell using watershed segmentation. Int. J. Adv. Res. Comput. Commun. Eng. 4(6), 488–491 (2015)
13. Bhatt, M., Prabha, S.: Detection of abnormal blood cells using image processing technique. Int. J. Electr. Electron. Eng. 7(01), 89–94 (2015)
14. Chintawar, I., Aishvarya, M., Kuhikar, C.: Detection of sickle cells using image processing. Int. J. Sci. Technol. Eng. 2(9), 335–339 (2016)
15. Turgeon, M. L.: Clinical Hematology: Theory and Procedures, Lippincott Williams & Wilkins, Philadelphia, PA (2005)
16. Blom, J.: Monitoring of Respiration and Circulation, CRC Press, Boca Raton, FL (2003)
17. Sender, R., Fuchs, S., Milo, R.: Revised estimates for the number of human and bacteria cells in the body. PLoS Biol. 14(8), e1002533 (2016)
18. Poole, J: Red cell antigens on band 3 and glycophorin A. Elsevier 14(1), 31–43 (2000). https://doi.org/10.1054/blre.1999.0124
19. Cohen, W.: The cytomorphic system of anucleate non- mammalian erythrocytes. Protoplasma 113(1), 23–32 (1982)
20. Wingstrand, K. G.: Non-nucleated erythrocytes in a teleostean fish maurolicus mu¨lleri (gmelin). Zeitschrift fu¨r Zellforschung und Mikroskopische Anatomie 45(2), 195–200 (1956)
21. Yazdanbakhsh, K., Lomas-Francis, C., Reid, M. E.: Blood groups and diseases associated with inherited abnormalities of the red blood cell membrane. Transfus. Med. Rev. 14(4), 364–374 (2000)
22. M. clinic, Sickle cell anemia. https://www.mayoclinic.org/-/media/kcms/gbs/patient-consumer/images/ (2017). Accessed 15 October 2019
23. Atlas of hematology, http://www.hematologyatlas.com/principalpage.htm. accessed 15 November 2019
24. Lowe, L. H., Bulas, D. I.: Transcranial doppler imaging in children: sickle cell screening and beyond. Pediatr. Radiol. 35(1), 54–65 (2005)
25. David, S. : Neuropsychological functioning of children with sickle cell disease and pica, Ph.D. thesis, Alliant International University (2017)
26. Bao, P., Zhang, L., Wu, X.: Canny edge detection enhancement by scale multiplication. IEEE Trans. Pattern Anal. Mach. Intell. 27(9), 1485–1490 (2005)
27. Javidi, B., Markman, A., Rawat, S., O'Connor, T., Anand, A., Andemariam, B.: Sickle cell disease diagnosis based on spatio- temporal cell dynamics analysis using 3d printed shearing digital holographic microscopy, Opt. Express 26(10), 13614–13627 (2018)
28. Begum, K., Mannan, M., Sanyal, M., Hosen, M., Chakraborty, S., et al.: Molecular diagnostic approach prevails superior over conventional gel-electrophoresis method in detecting sickle cell anemia. J. Mol. Biomark. Diagn. 9(382), 2 (2018)
29. Park, H. S., Rinehart, M. T., Walzer, K. A., Chi, J.-T. A., Wax, A.: Automated detection of p. falciparum using machine learning algorithms with quantitative phase images of unstained cells. PLoS One 11(9), e0163045 (2021)
30. Alzubaidi, L., Fadhel, M. A., Al-Shamma, O., Zhang, J.,Duan, Y.: Deep learning models for classification of red blood cells in microscopy images to aid in sickle cell anemia diagnosis. Electronics 9(3), 427 (2020)
31. Saravanan, D., Rajasekaran, S., et al.: Detection of sickle cell anemia from microscopic blood images using different local adaptive thresholding techniques. Ann. Romanian Soc. Cell Biol. 25(4), 6549–6564 (2021)
32. Wagner, D. B.: Liu hui and tsu keng-chih on the volume of a sphere, Chin. Sci. 3, 59–79 (1978)
33. Gibson, K. D., Scheraga, H. A., Exact calculation of the volume and surface area of fused hard-sphere molecules with unequal atomic radii. Mol. Phys. 62(5), 1247–1265 (1987)

How to cite this article: Vicent, M., Simon, K., Yonasi, S.: An algorithm to detect overlapping red blood cells for sickle cell disease diagnosis. IET Image Process. 16, 1669–1677 (2022). https://doi.org/10.1049/ipr2.12439