CME LIVE OPEN ACCESS

Engineering and Applied Sciences Journal

Research Article

An Image-Based Sickle Cell Detection Method

Florence B Tushabe[1*], Samuel Mwesige[2], Vicent Kasule[3], Emily Nsiimire[4], Sarah C Musani[5], David Areu[6] and Emmanuel Othieno[7]

[1]*Works as an Associate Professor in the School of Engineering and Technology of Soroti University.*

[2]*Works in the Bio Chemistry Department of Soroti University.*

[3]*Student of Electronics and Computer Engineering at Soroti University.*

[4]*Bachelors of Nursing Sciences student at Soroti University.*

[5]*Works in the Bio Chemistry department of Soroti University.*

[6]*Finance Department of Soroti University.*

[7]*Associate Professor of Pathology at Soroti University.*

Corresponding Author: Florence B Tushabe, works as an Associate Professor in the School of Engineering and Technology of Soroti University.

Received: 2025 Jan 29 **Accepted:** 2025 Feb 18 **Published:** 2025 Feb 28

Abstract

This article presents a method of sickle cell detection from microscopic images. We extract five attribute values from the connected components of an image, and train machine learning classifiers to recognize the sickle cells. Four classifiers were experimented with and the vest one was the K-Nearest neighbor classifier with 97.3% accuracy. The other classifiers are the Neutral network, Decision tree and Naïve Bayesian classifiers which resulted in accuracy rates of between 89-96.3%. This method is applicable for use in low cost computers since it is computationally cheap. The findings of this research can be considered as a screening method for diagnosing sickle cell aneamia.

Keywords: Sickle Cell, Sickle Cell Detection, Anaemia Diagnosis, AI for Sickle Cell, Machine Learning in Health, SCD App

1. Introduction

Sickle Cell disease (SCD) is a public health disorder affecting millions of people across the globe. It is a genetic condition characterized by production of abnormal red blood cells which take on the shape of a sickle, as opposed to the normal round ones. SCD causes several complications including pain, swelling, anaemia, organ damage, blindness, stroke and premature death. Current estimates suggest that there are 6 million Africans living with sickle cell disease and over 50 - 80% of these patients die before adulthood [1]. The World Health Organization asserts that 70% of sickle cell anemia deaths in Africa are preventable if some interventions such as early screening or identification of sickle cells is carried out [2]. Most patients in low resourced countries like in Africa, delay to test for SCD because of difficulties in accessing the testing services caused by scarcity of medical personnel, expensive costs, long distance travel to access the services or delays of as long as days/weeks to receive feedback. All this makes routine screening for SCD largely impractical and thereby delay timely intervention.

2. Literature Review

In ImPatho is introduced to the world [3]. It is a tool that detects sickle cells using only one attribute – eccentricity. Unfortunately, the paper does not discuss the accuracy rates of the detection algorithm. In the roundness of a component is used as the distinguishing feature. However, only 4 images were used in the study, which is a very tiny and insignificant dataset to make conclusions from [4]. In carried out a comprehensive review on automated methods to detect sickle cells [5]. The findings revealed that detection accuracy ranged from 80.6%, 88%, 91% to 95% [6-9]. These accuracy rates can still be improved upon.

3. Methodology

This work proposes an image-based Sickle cell detection method that has been implemented in MATLAB [10].

3.1. It Consists of Four Steps that are Highlighted Below

3.1.1 Image Acquisition

A 560-image dataset that is publicly available on Kaggle was used in this study. We used an image dataset that we developed and published in [11]. Each image consists of red blood cells which are wither all normal (negative) or containing atleast one sickle cell (positive). The images were captured from blood samples collected from Eastern Uganda where sickle cells disease is prevalent [12]. The same researchers of this work collected and processed the blood samples in [12]. The dataset has already labelled images with the sickle cells surrounded by a bounding box around them for identification [11]. The dataset contains four hundred and twenty-two (422) positive images, with

43% (180) of them having been captured with field stains and 57% (242) with Leishman's stains. The dataset contains one hundred forty-seven (147) negative images/slides with 62% (91) having been captured by field stains and 38% (56) by Leishman stains.

3.1.2. Pre-Processing

All images were pre-processed by conversion to binary, inversion for a black background, extraction of their connected components and filtering out of noise. Fig. 1 illustrates how image 20 was transformed when the region of interest was the sickled cells.

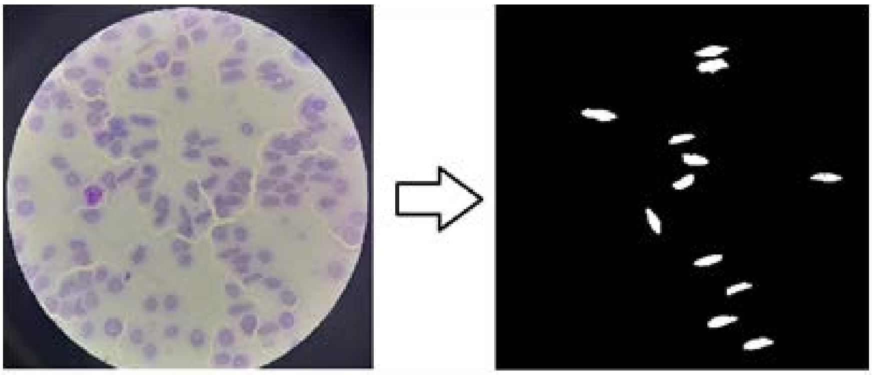

Figure 1: A Transformed Image

After preprocessing all the 569 images, the total number of connected components extracted were 21,645 with the sickled cells being only 1,870 in number while the negative ones were 19,775. Clearly, we still had a small positive set compared to the negative ones but that is the nature of each image because the negative cells far exceed the positive ones for all the images.

3.1.3. Feature Extraction

Five features were then extracted from the connected components (objects) which are: Major Axis Length, Minor Axis Length, Eccentricity, Convex Area and Circularity [10]. The Major Axis Length is the length (in pixels) of the major axis of the ellipse that has the same normalized second central moments as the region. The Minor Axis Length is the length (in pixels) of the minor axis of the ellipse that has the same normalized second central moments as the region. The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length. The Convex Area is the number of pixels in convex image.

The Circularity (C), is the roundness of the object and computed as in (1)

$$C = \frac{4\pi A}{p^2} * \left(1 - \frac{0.5}{r}\right)^2 \qquad (1)$$

where p is the perimeter and r the radius defined as

$$r = \frac{p}{2\pi} + 0.5 \qquad (2)$$

3.1.4. Training

The data was then divided into two parts, with one part (80%) as training data, the second one as testing data (20%). We applied four machine learning algorithms and trained them to recognize the Sickle cells. They are.

- K-Nearest Neighbor (KNN) with K = 2 and the City Block distance metric.
- Decision Tree (DT) with a maximum number of splits = 100.
- Naïve Bayesian (NB) using Gaussian model.
- Neural Networks (NN).

4. Result

4.1. Performance of NB

Naïve Bayesian registered an accuracy of 88.7% with a prediction speed of 1,200,000 observations per second. The classifier correctly categorized 24.0% of positive cases as positive (448 of the 1,870 positive samples) and 94.9% of the negatives cases as negative (18,758 of 19,775 negative samples).

4.2. Performance of DT

A better performance was registered by the fine Decision tree. This was 96.3% accurate and a prediction speed of 1,700,000 observations per second. The classifier correctly categorized 67.7% of positive cases as positive (1,266 of 1,870 that were positive) and 99.0% of the negatives cases as negative (19,576 of 19,775 negatives).

4.3. Performance of NN

The second-best performance was by a Neural Network which has 96.9% accuracy using a prediction speed of 1,400,000 observations per second. The classifier correctly categorized 74.2% (1,387 out of 1,870) of positive cases as positive and 99.0% (19,576 out of 19,775) of the negatives cases as negative.

4.4. Performance of KNN

The best performance was by KNN which registered an accuracy of 97.3% and a prediction speed of 470,000 observations per second. The KNN confusion matrix is presented in Fig. 2 where its observed that the classifier correctly categorized 76.2% of positive cases as positive (1,424 of the 1,870 positive samples) and 99.3% of the negatives cases as negative (19,641 of 19,775 negative samples). This is an impressive performance especially for the negative components.

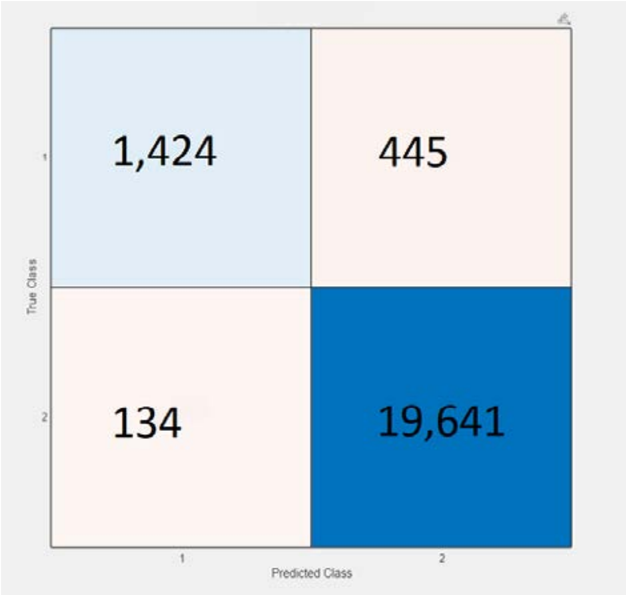

Figure 2: Confusion Matrix for KNN

5. Discussion

A summary of the performance of the four classifiers is presented in Table I where it is observed that the accuracy of the four classifiers ranged between range of 89 - 97%. This is quite impressive indeed. However, the best positive (sickle cell) rate was only 76% as opposed to 99% for the normal cells. This is explained by the large negative training set as compared with the meagre positive set which comprised only 9% of the training set. The best training algorithm recorded a true negative rate of 99.3%. This means that 100% of what the classifier categorizes as negative is negative indeed. But only 76% of what is said to be positive is positive indeed. These percentages are indeed in the range of what the other detection or screening methods available on the market produce. A smartphone App for blood pressure measurement recorded accuracy rates of only 85% [13]. In eight Apps were analysed by seven medical doctors and the top three Apps provided accuracy rates between 95.1 – 97.8%. With our App registering a rate of 97.3%, this shows that it performs within the range of well performing diagnosis Apps [14].

It is very interesting to note that when applied Convolution neural networks (CNN) on the same dataset, an accuracy of 97.7% was obtained. This performance is similar to what KNN registered here (97.3%) [15]. Therefore, it is safe to say that there is no significant difference in performance between the CNN and KNN classifiers. Additionally, this work revealed that the fastest classifiers are NB and DT but unfortunately, they also provide the lowest accuracy rates. The slowest classifier is KNN but registered the highest accuracy rate of 97.3%. Therefore, KNN is the recommended classifier of this research due to high accuracy rates and despite the slower performance.

ATTRIBUTE	NB	DT	NN	KNN
Speed	1,200,000	1,700,0000	1,400,0000	470,000
Training Time (Sec)	2.38	1.87	25.6	56.5
Accuracy	88.7%	96.3%	96.9%	97.3%
True Positive Rate	24.0	67.7	74.2	76.2
True Negative Rate	94.9	99.0	99.0	99.3

Table I: Summary of the 4 Classifiers

6. Conclusion

This work presents a method for detecting sickle cells from digital images. We present five attributes that can be extracted from the image components and performance results of four classifiers: Naïve Bayes, Decision Trees, K-Nearest Neighbor and Neural networks. Our experiments reveal that KNN is superior at recognizing sickle cells than the other three, since 100% of what it says is negative is negative indeed and 76% of what it says is positive is positive indeed. The true positive rate can further be improved by collecting more images with sickle cells and hence training on a wider set of positive images. The performance of this method is already in the range of most good detection or screening methods available on the market. The methods proposed don't require use of sophiscated processors like GPUs but can comfortably work on ordinary CPU Processors. This method is therefore suitable for use in even resource constrained countries like in Africa and South America. This method can be implemented onto smartphones which are common gadget available to health workers in Africa, including those in rural settings and therefore, using such a readily available item for SCD screening, may go a long way in enhancing access to screening interventions.

Acknowledgment

The Authors send sincere gratitude to the subjects who participated in the study especially the researchers, blood donors, service providers and volunteers. More thanks to the staff and students of Soroti University who tested the App and offered logistical support to this research.

References

1. Weatherall, D., Akinyanju, O., Fucharoen, S., Olivieri, N., & Musgrove, P. (2006). Inherited disorders of hemoglobin. *Disease Control Priorities in Developing Countries. 2nd edition.*
2. Makani, J., Cox, S. E., Soka, D., Komba, A. N., Oruo, J., Mwamtemi, H., ... & Newton, C. R. (2011). Mortality in sickle cell anemia in Africa: a prospective cohort study in Tanzania. *PloS one, 6*(2), e14699.
3. Shah, S., Dhameliya, V., & Roy, A. K. (2014, August). ImPatho-an image processing based pathological decision support system for disease identification and a novel tool for overall health governance. In *2014 IEEE Region 10 Humanitarian Technology Conference* (R10 HTC) (pp. 64-69). IEEE.
4. Patgiri, C., & Ganguly, A. (2019, March). Red blood cell and sickle cell detection from microscopic blood images of sickle cell anemic patient. In *2019 international conference on wireless communications signal processing and networking* (WiSPNET) (pp. 474-478). IEEE.
5. Das, P. K., Meher, S., Panda, R., & Abraham, A. (2019). A review of automated methods for the detection of sickle cell disease. *IEEE reviews in biomedical engineering,* 13, 309-324.
6. Sharma, V., Rathore, A., & Vyas, G. (2016, August). 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) (Vol. 3, pp. 1-5). IEEE.
7. Chy, T. S., & Rahaman, M. A. (2019, January). A comparative analysis by KNN, SVM & ELM classification to detect sickle cell anemia. In *2019 International conference on robotics, electrical and signal processing techniques* (ICREST) (pp. 455-459). IEEE.
8. Albayrak, B., Darici, M. B., Kiraci, F., Öğrenci, A. S., Özmen, A., & Ertez, K. (2018, November). Orak hücreli anemi tespiti sickle cell anemia detection. In *2018 Medical Technologies National Congress* (TIPTEKNO) (pp. 1-4). IEEE.
9. Chy, T. S., & Rahaman, M. A. (2018, November). Automatic sickle cell anemia detection using image processing technique. In *2018 International conference on advancement in electrical and electronic engineering* (ICAEEE) (pp. 1-4). IEEE.
10. The MathWorks, Inc. (2023). MATLAB version: 9.13.0 (R2023b). Accessed: November 01, 2024.
11. F. Tushabe et al, 2024, “Sickle cell disease dataset”, distributed by Kaggle.com.
12. Tushabe, F., Mwesige, S., Kasule, V., Nsiimire, E., Musani, S. C., Areu, D., & Othieno, E. (2024). An Image-based Sickle Cell Detection Method. *Authorea Preprints.*
13. Degott, J., Ghajarzadeh-Wurzner, A., Hofmann, G., Proença, M., Bonnier, G., Lemkaddem, A., ... & Schoettker, P. (2021). Smartphone based blood pressure measurement: accuracy of the OptiBP mobile application according to the AAMI/ESH/ISO universal validation protocol. *Blood Pressure Monitoring, 26*(6), 441-448.
14. Gilbert, S., Mehl, A., Baluch, A., Cawley, C., Challiner, J., Fraser, H., ... & Novorol, C. (2020). How accurate are digital symptom assessment apps for suggesting conditions and urgency advice? A clinical vignettes comparison to GPs. *BMJ open, 10*(12), e040269.
15. M. Makola, 2024, “Code for sickle cell CNN classifier (Version 5 of 5)”, distributed by Kaggle.