Acta Scientific Microbiology (ISSN: 2581-3226)

Volume 7 Issue 12 December 2024

Research Article

A Dataset of Microscopic Images of Sickle and Normal Red Blood Cells

Florence Tushabe[1]*, Samuel Mwesige[2], Kasule Vicent[1], Emmanuel Othieno[2], Emily Nsiimire[2], Philip Mutabazi[3], Sarah Musani[2] and David Areu[1]

[1]*School of Engineering and Technology, Soroti University, Soroti, Uganda*

[2]*School of Health Sciences, Soroti University, Soroti, Uganda*

[3]*Kismut Office Solutions, Uganda*

***Corresponding Author:** Florence Tushabe, School of Engineering and Technology, Soroti University, Soroti, Uganda.

Received: September 24, 2024

Published: November 14, 2024

Abstract

The World Health Organization (WHO) estimates that about 5% of the world's population carries genes responsible for sickle cell trait and each year about 300 000 infants are born with sickle cell disease (SCD). SCD is a lifelong condition where sickled red blood cells block perfusion, leading to several complications such as anaemia, pain, swelling, organ damage, blindness, stroke and premature death. Artificial intelligence techniques could lead to more insights to understanding SCD including improved testing and treatment options. This necessitates the use of as much data about SCD as possible, including image data. However, there are few available image datasets that depict microscopic images of sickled and normal red blood cells. There are even fewer datasets whose images were captured using non-professional cameras like those in mobile phones. In this work, we present a 569-image dataset of sickled and normal red blood cells, well-labelled and publicly available freely. The images were captured using mobile phone cameras which are widely used even in resource constrained places like in Africa. The blood samples were collected from hospitals in Soroti and Kumi districts in Uganda, Africa. They were analysed using Leichman and field stains and labelled by qualified Laboratory scientists. This dataset is useful for mobile applications that apply computer vision, deep learning, data science, artificial intelligence or machine learning techniques for medical diagnosis, health research, pharmaceutics and blood banks.

Keywords: Red Blood Cells; Sickle Cell Disease; Sickle Cell Anaemia Images; Sickle Cells Database; Microscopic Images

Introduction

The World Health Organization (WHO) estimates that about 5% of the world's population carries genes responsible for sickle cell trait and each year about 300,000 infants are born with sickle cell disease (SCD) [1]. Sickle cell disease (SCD) is a public health disorder affecting millions of people across the globe.

It is a genetic condition characterised by the production of abnormal red blood cells which take on the shape of a sickle or a crescent, as opposed to the normal round ones. An Adenine-to-Thymine point mutation in the β- globin gene produces abnormal haemoglobin S (Hb S), which polymerizes in the deoxygenated state, resulting in the physical deformation or sickling of erythrocytes [2]. The sickled cells become sticky and rigid and slow blood flow, leading to several complications including anaemia, pain, swelling, organ damage, blindness, stroke and premature death.

Machine learning techniques can be used to manipulate sickle cell data to aid in understanding SCD better and enhance improved disease management options. However, there are almost no publicly available SCD image datasets that researchers can use to conduct further analysis of the disease. Although many resource-constrained places in Africa and South America have access to smartphones and mobile applications, they still lack microscopes with built-in cameras. This application gap creates a big opportunity for us to combine mobile applications with microscopic data.

Upon reviewing SCD interventions, we have found only three tiny SCD microscopic image datasets, whose data is too small to significantly train machine learning models. In addition, images were captured by professional people who used good cameras or custom-built accessories. This means that the quality of such images does not reflect the scenarios where non-professional photos are to be input or processed. It is therefore crucial that a more substantial SCD image dataset with a divergent image scope is created in order to harness the opportunities of artificial intelligence within SCD medical research, including mobile applications.

Literature Review

Gonzales., *et al*. [3] created an image dataset of 211 sickled images and 202 normal cells. However, they used three types of images: artificial images, which were automatically generated in a random manner using a computer code; real images from peripheral blood smear sample images that contained normal and elongated erythrocytes; and synthetic images generated from real isolated cells. The dataset is good but since it also contains many artificial images, it is not a realistic dataset to use for some research. In 2020, The GitHub platform which is probably the most popular AI open-source platform published a small dataset of 12 images [4]. On the Kaggle Data science platform, there is one SCD dataset with four images [5]. All these datasets are too tiny to be significant for data science projects and this was the motivation of compiling this dataset.

Materials and Methods

Ethical statement

This research was approved by the Clarke International University Research and Ethics Committee in Uganda with number MRRH-2023-263. Written informed consent for the publication of details relating to a participant was obtained from them (or their parent or legal guardian in the case of children under 18). The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved

Materials

We used two microscopes to examine the stained thin blood smears under high-power magnification of x100 objective. The microscopes are: Optika model B-292PLi, SN 543199 and Olympus Magnus model MX21iLEDFS11 SN 21D0034.

Four mobile phones were used to shoot the images of the blood slides: Techno Spark 7 model KF75, Techno Camon 19 model C16, RedMi 9A model m2006C3LG and Samsung Galaxy A03s. Figure 1 shows one of the team members capturing an image using these materials.

Other materials used include; Glass slides, EDTA tubes, electric balance, absolute methanol, disodium hydrogen phosphate and potassium dihydrogen phosphate, Field stains A and B, Leishman stain and Immersion oils.

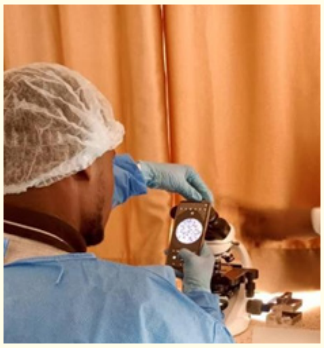


Figure 1: Mr. Kasule capturing microscopic images using the Techno Spark 8. **Source:** Primary data, 2024

Methodology

Below is the methodology that was used during the study.

Sample collection

Blood samples were collected within a period of six months from September 2023 - February 2024. Four mills(4mls) of blood was drawn from the participant by the hospital phlebotomist and placed in EDTA tubes, well packaged and transported to Soroti University teaching biochemistry laboratory for further processing.

One hundred forty samples were randomly collected from the sickle cell clinics of Soroti Regional Referral Hospital and Kumi Hospital as well as from volunteers from Soroti University student community. All the volunteers that were sourced from outside the hospitals were negative for SCD although four of them were carriers as confirmed by the Central Public Health Laboratory in Kampala. Of the 140 blood samples collected,105 were from sicklers and 35 from non-sicklers. Written consent was obtained from all participants above 18 years of age with the ones below 12 years being consented for by their parents/guardians. The ones between 12and 18 years of age signed letters of assent in addition to the letters of consent signed by their guardians/parents.

Inclusion criteria

The following factors formed the inclusion criteria for the study. The participant:

- Voluntarily accepted, as evidenced by a signed consent form.
- Was a resident of Kumi or Soroti districts
- Was aware of their sickle cell disease status (sickler, carrier or normal)
- is aged at least one year old and above
- Fluent in either English or Ateso
- is not sick but on routine checkup

Exclusion criteria

The following factors form the exclusion criteria for the study. The participant was not:

- Pregnant or less than 6 weeks postnatal
- Aged below one year old
- A blood transfusion recipient less than 2 months prior
- Diagnosed with heamophilia
- Under cancer chemotherapy or hydroxyurea treatment recently (4 weeks prior)
- Suffering from mental illness.

Reagent preparation

A thin blood smear was prepared, dried and stained with Romanowsky stains. The basic parts of blood cells pick up acidic stains (eosin) while the acidic parts of blood cells pick up basic stains (methylene blue). Under high-power examination with a microscope, sickle cells are seen at the tail end of the smear where red blood cells don't overlap.

The reagents were prepared by following the summarised process below:

- 6.8mls of buffered water was prepared by mixing Disodium hydrogen phosphate and potassium dihydrogen phosphate
- A stock solution of Leishman stain was prepared by mixing 0.6g of the stain
- The working field stains A and B were prepared from the A and B stock stains respectively.

Sample analysis

After receiving the blood samples, the Laboratory Scientists on the team analysed them to identify the sickle ornormal cells as seen on the microscope. For quality control and comparison purposes, we examined using two smears which are Leishmans tain and field stains. The thin smears were prepared following the Standard Operation Procedures on Thin Blood films. Figure 2 shows some of the processed samples on slides.

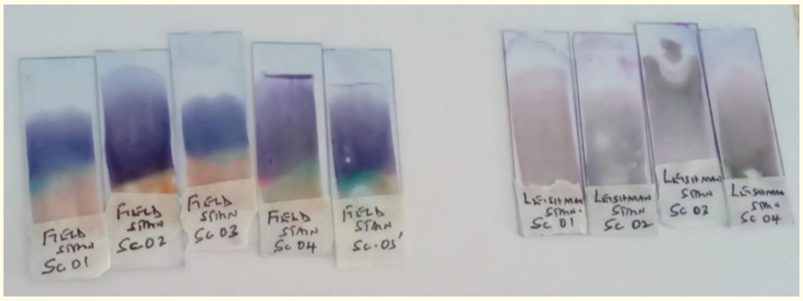


Figure 2: Field stained thin smears (left) and Leishman stained thin smears (right).

Photo shooting

Once the proper microscopic view was identified by the medical team, the other members of the team used themobile phones to capture the microscopic images of the blood slide. The images were captured after placing themobile phone before the eyepiece of the Microscope. No additional accessory was used. Refer to Figure 1 whichshows one of the team members capturing the microscopic images.

For each smear, three images were shot (one per person per phone). We used two men and one woman to capture the images, all of whom were amateur camerapersons. Each phone camera setting was varied to captureboth good and poor images which reflect the different scenarios within a typical working environment.

We arranged the images by grouping the positive and negative ones together. Each image was captioned as either sickle cells present or sickle cells absent depending on whether the sickle cell was seen.

Image labelling

The images of the microscopic slides (as seen through the microscopic lens) were captured as elaborated in Section 3.3.4. They turned out to be circular in nature as opposed to those captured from inbuilt microscope cameras which are rectangular.

The Laboratory Scientist on the team then identified the sickle cells in the image, which were labelled by drawing a rectangular bounding box around each. An image has been labelled positive when at least one sicklecell is identified.

Results

Demographic summary

Fifty six percent (56%) of the respondents were males and 44% female. Also, 54% were below 15 years of age and 46% were atleast 15 years of age. Forty two (42%) of the participants (59 respondents) were sourced fromKumi district and 58% from Soroti district (35 participants sourced from Soroti University and 46 from Soroti Regional Referral hospital). Figure 3 illustrates these findings.

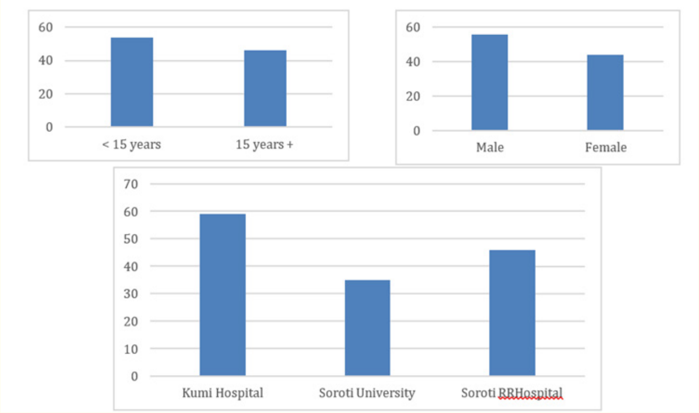


Figure 3: Demographic summary by Age, Gender and Location.

Sickle cell status

Thirty-one respondents (22%) did not know their sickle cell status before participation in the study and one hundred and nine respondents (78%) knew their sickle cell status as positive since they were already attendingthe sickle cell clinics. By the end of the research, all participants knew their Sickle Cell Status. We tested the blood samples using the two different screening methods. All the thirty one participants who did not know theirsickle cell status turned out to be negative. We had to send eight samples which exhibited suspicious cells to the Central Public Health Laboratory in Kampala for confirmation and four turned out as sickle cell carriers. The carriers were counselled when giving them their results.

The image dataset

A 569-image color dataset has been produced from the study. All the images have been resized to have a horizontal width of 1000 pixels.

The dataset contains four hundred and twenty two (422) positive images, with 43% (180) of them having beencaptured 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. Figure 4 illustrates a summary of the dataset by sickle cell status of the images. Figure 5 displays a sample of the images in

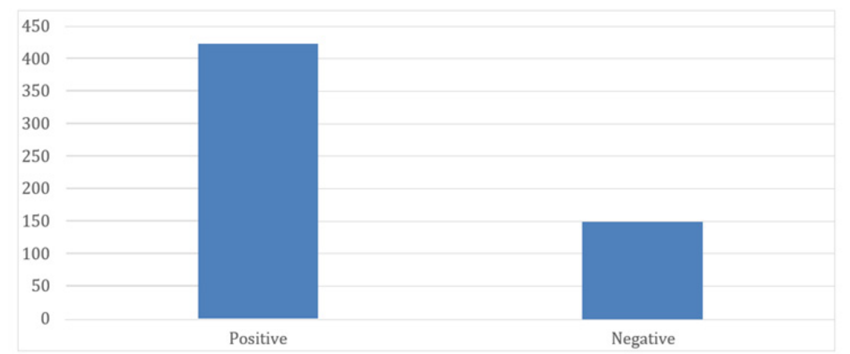


Figure 4: Summary of the dataset.

the dataset. Figure 5(a) shows a Leishman stained image while Fig 5(b) shows a Field stained one and 5(c) shows a birds view of the dataset. Please note that the Leishman stained images have a pink object color while the field stains have a purple object coloring.

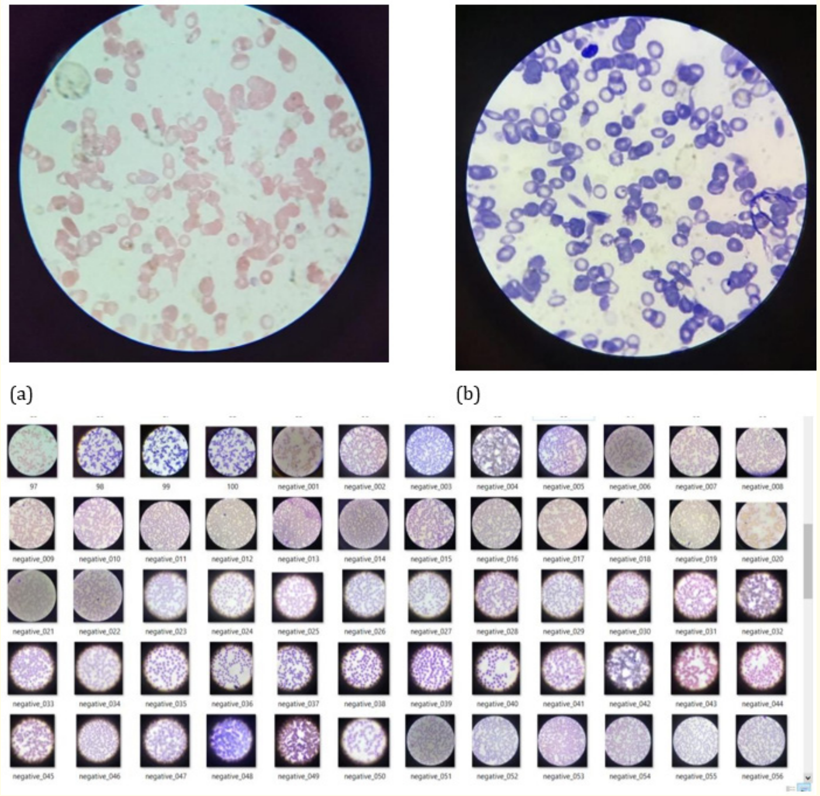


Figure 5: Images from (a) Leishman stained film, (b) Field stained film (c) birds view of a mixture of both.

The dataset has been uploaded onto Kaggle.com [6] and is freely available for research purposes. It contains four folders which are: Positive with two sub-folders labelled and unlabelled. The labelled folder contains 422positive images well labelled with boundary boxes around the sickle cells while the unlabelled folder containsthe 422 positive image with no boundary boxes around the sickle cells. Figure 7 shows six of the labelled imagesin the dataset.

The folder negative contains one sub-folders: Clear. The Clear subfolder consists of 147 negative images which do not contain any sickle cell. A sample of two negative images is shown in Figure 6. As a side additionaloptional content, we have another set of images saved in folder Not Clear which consists of 122 negative images but which are poor and hence optional to download. Poor images can be useful for testing purposes especially when the application expects input from amateur photographers.

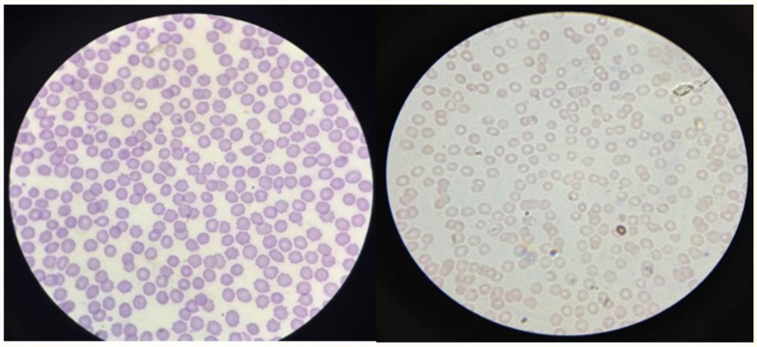


Figure 6: Example of two negative images in the dataset.

A sample of unclear images in the dataset is presented in Figure 8 which shows some images with unexpected artefacts, un-usual colors, badly cropped or stained. Please note that the Unclear Images arenot included in the analysis of the results in this article nor are they uploaded onto Kaggle but can be availed on request.

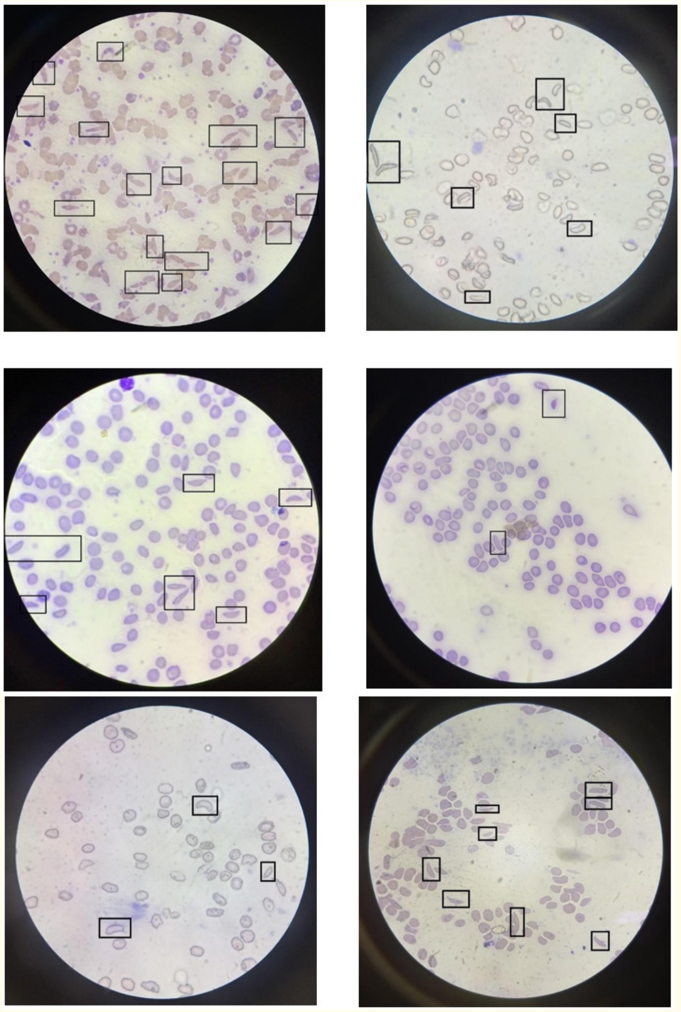
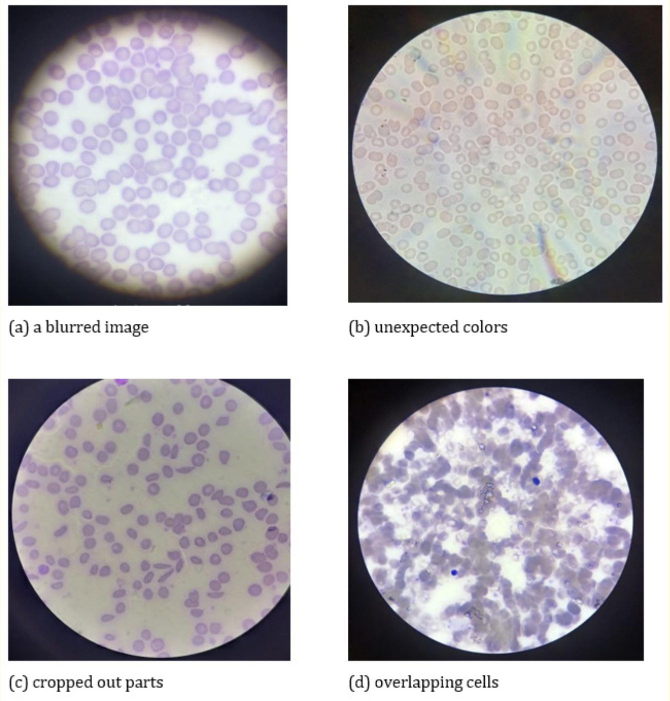


Figure 7: Some of the labelled images in the dataset.


(a) a blurred image

(b) unexpected colors

(c) cropped out parts

(d) overlapping cells

Figure 8: Examples of 4 poor images.

We conducted a manual count of the sickle cells per image, and the results show that the majority of images contain 1- 4 sickle cells. This may be so because the sicklers were found in hospital during their routine weekly visits and they were not sick or under a crisis. Figure 9 illustrates the distribution of number of cellsper labelled image within the dataset.

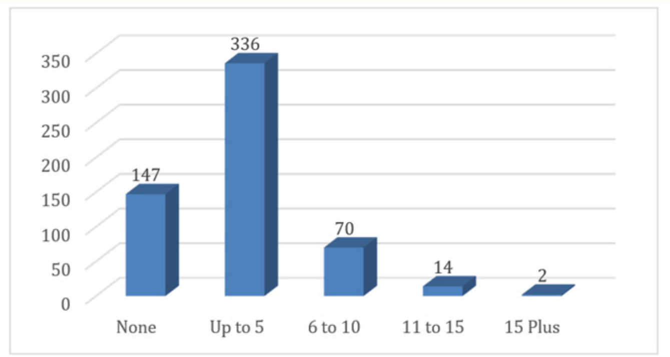


Figure 9: Histogram of the sickle cells per image.

Discussion

This is the largest sickle cell microscopic image dataset that we have come across in the literature. It is the firstone of its kind and can be used as a basis for solutions in resource constrained settings where microscope accessories like cameras are a luxury.

Conclusion

This study has resulted in a dataset of microscopic images containing sickle cells and normal red blood cells. Itcontains 569 microscopic images, of which four hundred and twenty two images contain at least one sickle cellthat is clearly visible while one hundred and forty seven are normal cells that lack any clearly visible sickle cell.This dataset is useful for researchers in medical informatics to utilise to advance knowledge in sickle cell diseasemanagement.

Acknowledgements

The authors would like to thank the Government of Uganda through the Soroti University Research and Innovations Fund SUN-RIF/2022/05 for funding the research; Clarke International University Research and Ethics Committee for the approval, the participants for providing their blood samples and the hospitals and university community for the co-operation during the study.

Bibliography

1. World Health Organisation. “Sickle-cell anaemia Report by the Secretariat, Fifty-ninth World Health Assembly, Provisional agenda item” 11.4 (2006).

2. Samuel M., *et al*. “The Silent Anaemia Epidemic in Children: Theory and Research”. *IJISRT* 7.5 (2020).

3. M González-Hidalgo., *et al*. “Red Blood Cell Cluster Separation From Digital Images for Use in Sickle Cell Disease”. in IEEE Journal of Biomedical and Health Informatics 19.4 (2015): 1514-1525.

4. Weiss (2020).

5. Andrew (2023).

6. F Tushabe., *et al*. “Sickle Cell Disease Dataset”. Kaggle.com.