East African Scholars Journal of Engineering and Computer Sciences
Abbreviated Key Title: East African Scholars J Eng Comput Sci
ISSN: 2617-4480 (Print) & ISSN: 2663-0346 (Online)
Published By East African Scholars Publisher, Kenya
Volume-8 | Issue-5 | Nov-Dec-2025 | DOI: https://doi.org/10.36349/easjecs.2025.v08i05.001

OPEN ACCESS

Original Research Article

A SMOTE-Based Deep Learning Approach for Sickle Cell Detection in Low-Resolution, Class-Imbalanced Microscopic Images (SDL-SCD)

Kisakye Martha Janepher[1], Abubakhari Sserwadda[2]*, Ali Najib[3], Tushabe Florence[2]

[1]College of Computing and Information Sciences, Makerere University, Kampala, Uganda
[2]School of Engineering and Technology, Soroti University, Soroti, Uganda
[3]Faculty of Science and Technology, Victoria University, Kampala Uganda

Article History
Received: 17.04.2025
Accepted: 23.05.2025
Published: 12.11.2025

Journal homepage:
https://www.easpublisher.com

Quick Response Code

Abstract: sickle cell disease is a genetic condition characterized by abnormal red blood cell morphologies. It can be quite challenging to identify and monitor its response to treatment. Although deep learning-based models exhibit great potential in medical image processing, existing approaches often fail to cope with variability in sickle cell morphology. Additionally, publicly available sickle cell datasets tend to have a few samples with imbalanced classes. To mitigate the above challenges, we propose using the synthetic minority sampling technique (SMOTE) mechanism to handle class imbalances and a deep CNN architecture that aims to capture complex patterns and descriptive features in a newly created low-resolution sickle cell dataset from hospitals in eastern Uganda. This could help improve the efficiency of the diagnosis and classification of the disease. We performed experiments and examined several algorithms in the literature for related tasks. Based on the evaluation results, the proposed SMOTE-based DL-SCD outperforms the best baseline, its variant without the SMOTE component, with a 2.06% increase in classification accuracy. SDL-SCD could help to conveniently and early detect sickle cell anemia, especially in low-developed settings where medical services are constrained. Our code is accessible at https://github.com/MarthaKJ/sickle-cell-detection-using-nvidia.

Keywords: Convolutional Neural Networks (CNN), Sickle Cell, Synthetic Minority Over-sampling technique (SMOTE).
Categories: G.2.2, I.2.6, I.5.1, I.5.4.

Copyright © 2025 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution **4.0 International License (CC BY-NC 4.0)** which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use provided the original author and source are credited.

INTRODUCTION

Sickle cell disease (SCD) is the most common inherited hematologic disorder worldwide and a public health priority [Kang *et al*., 2024]. The majority of the world's burden of SCD is in sub-Saharan Africa, affecting millions of people of all ages. It is estimated that 200, 000 to 300,000 children are born with SCD every year in Africa alone [Modell and Darlison, 2008, Kavanagh *et al*., 2022]. The prevalence of the disease varies between countries, being approximately 20% in Cameroon, Ghana, and Nigeria and even increasing to 45% in some parts of Uganda [Egesa *et al*., 2022]. Because sickle cells have aberrant hemoglobin, they have a sickle shape and are difficult to pass through blood vessels [Azar and Wong, 2017]. These stiff and sticky sickle cells tend to group together and become lodged in the blood vessel [Li *et al*., 2017, Nader *et al*., 2020]. The sickle cells can only live for 10 to 20 days, after which they die [Umar *et al*., 2023], depriving several parts of the body of oxygen. Thus, ischemia, pain, and other subsequent complications. The symptoms of severe anemia include yellowing of the skin, delayed growth and development, difficulty in breathing, and enlargement or inflammation of the hands and legs [Ware *et al*., 2017]. Severe and debilitating pain episodes can occur in the chest, abdomen, joints, and bones. In addition, patients may also have an increased risk of organ damage, infections, exhaustion, and visual impairment [Dirksen *et al*., 2011, GREEN and CONLEY, 1951]. The conventional method is to test a patient's blood sample for sickle cell anemia. Often it is challenging to observe. sickle cells [Sharma *et al*., 2016]. Various methods have been developed for the screening and diagnosis of Sickle Cell Disease, including laboratory-based methods such as high-performance liquid chromatography and genetic tests. Other effective measures of sickle cell detection include deep learning-

***Corresponding Author:** Abubakhari Sserwadda
School of Engineering and Technology, Soroti University, Soroti, Uganda

based convolutional networks. Owing to the fact that convolutional neural networks (CNNs) are efficient in spotting patterns in images [Liu *et al*., 2018], yet, sickle cells have unique shapes and patterns that these CNNs can learn to recognize, this research utilises CNNs to detect sickle cell anemia. The investigation of image processing for sickle cell anaemia detection has received a lot of attention lately.

Related Work

In this section, the related methods and studies presented in the literature are discussed. The review is organized into three main categories: Multistage approaches, feature-based techniques, and deep learning-based methods for sickle cell anemia detection

Traditional Image Processing-Based Methods

Often, related studies use a multi-stage approach that includes several independent phases like image pre-processing, segmenting cells, extracting features, and classifying individual cells using different machine learning models for instance, [Savitt and Goldberg, 1989] proposed a technique to identify sickle cell morphology from blood smear pictures using fractional dimensions. They used a variety of segmentation strategies, including cluster-based segmentation and edge detection. Note that CNNs are capable of automatically extracting pertinent features from blood smear images, yet the fractional dimension technique requires manual feature engineering [Kumar *et al*., 2021]. In addition, clustering-based segmentation and manual feature extraction from thin blood smear images were presented by [Barpanda, 2013]. In [Chy and Rahaman, 2018], another method is proposed that uses traditional image processing techniques. The approach involves collecting blood images, preprocessing them using grayscale conversion and filtering, isolating red blood cells using threshold segmentation, and extracting information such as aspect ratio, entropy, and statistical measurements. A support vector machine classifier is then trained on these attributes to provide diagnosis. Additionally, [Elsalamony, 2017] utilized the Circular Hough Transform (CHT) for cell differentiation, and Elsalamony implemented a two-step process comprising shape signature methodology followed by neural network classification; our CNN model provides a more comprehensive and likely more effective method. While beneficial for many situations, this multi-stage procedure frequently accumulates errors in its independent sub-tasks, reducing the accuracy of the final findings.

Feature-based Techniques

Several studies have utilized machine learning techniques alongside feature engineering for classification tasks. Work in [Gual-Arnau *et al*., 2015], relies on established feature sets such as Fourier coefficients and form attributes. In addition, Clustering-based segmentation and manual feature extraction from thin blood smear images were presented by [Barpanda, 2013]. Furthermore, [Elsalamony, 2017] utilized the Circular Hough Transform (CHT) for cell differentiation, and Elsalamony implemented a two-step process comprising shape signature methodology followed by neural network classification; our CNN model provides a more comprehensive and likely more effective method. These methods involve sophisticated feature extraction phases, which can be time-consuming and error-prone, especially when used to massive data sets.

Deep learning-based graph embedding methods

Deep learning approaches are growing as an appealing option for automating feature extraction and categorization. [Xu *et al*., 2017] proposed a two-stage technique that uses AlexNet to isolate categorization from region-of-interest (ROI) extraction. To detect odd sickle cell presence, the ROI was first extracted from red blood cells and then classified using the AlexNet model. This method separates feature extraction from classification, which can still result in inconsistencies regardless of whether it is successful. CNNs are capable of automatically extracting pertinent features from blood smear images, yet the fractional dimension technique necessitates manual feature engineering [Kumar *et al*., 2021]. Although these methods work well, particularly in settings with limited resources, the majority of them rely on preset characteristics or require separate subtasks. The accuracy and caliber of the final outcomes can be significantly impacted by the accumulation of errors in these subtasks.

While some of the above approaches are effective, especially in resource-constrained environments, the majority of these strategies are based on predetermined features. Moreover, the errors in the independent sub-tasks characterizing the above strategies may result in large accumulated errors that severely affect the accuracy and quality of the end results. To overcome these challenges, we propose end-to-end deep learning that can identify intricate patterns in sickle cell shape that would be difficult for traditional machine learning approaches to detect. The proposed SDL-SCD is capable of automatically learning complex low- and high-level features, and hierarchical representations and does away with the requirement for distinct phases by combining feature extraction and classification into a single end-to-end learning process. This strategy has the potential to result in more robust categorization and can be enhanced by leveraging transfer learning and GPU acceleration. Below we summarise the major contributions of our proposed SDL-SCD model.

First, we prepare the relatively large raw sickle cell dataset, extracted from local hospitals in Eastern Uganda to be fit for learning models. To the best of our knowledge, no publication has been made on this dataset yet.

Second, the study solves the class imbalance in the dataset, by effectively adapting SMOTE to oversample the minority class(negative).

Lastly, we thoroughly evaluate the proposed SDL-SCD model and demonstrate its efficiency in sickle cell detection as compared to the baseline models.

Proposed method

The processed SDL-SCD architecture consists of image preprocessing, class balancing with data argumentation using SMOTE, and deep learning-based categorization.

Data collection

Our dataset comprises medical images depicting blood smear samples, which are divided into two classes: Negative: These images depict normal red blood cells absent of sickle cell characteristics. Positive: These images show the typical ”sickle-shaped” cells associated with sickle cell disease.

The dataset that was used includes real-world samples from the referral hospitals in Kumi and Soroti. 140 patients submitted data samples, which were processed by applying the Leichman Strains and field strains techniques. The initial dataset is highly imbalanced with far more positives than negative samples.

Preprocessing

To ensure that each image has a uniform input size for a deep learning algorithm, these steps are taken: Grayscale Conversion: Color distinctions in medical imaging often provide insufficient information for classification. We decrease the complexity of computation by transforming the images to grayscale, allowing the model to focus on the structural properties of the cells (such as shape, density, and texture), which are more important for sickle cell detection. Image Resizing: Each image is scaled to a consistent size of 100x100 pixels. This standardization guarantees that the model processes all inputs uniformly, hence increasing training efficiency and accuracy. The scaling is done carefully to preserve crucial properties such as the form of sickle cells, which is essential for accurate diagnosis. Normalization: Image pixels are normalized to ensure that all photos have the same scale, which improves the model’s generalization.

Addressing Class Imbalance using SMOTE

The primary issue in this study is the visible variance in the number of negative (nonsickle cell) and positive (sickle cell) samples. This mismatch may bias the model toward forecasting the majority class, limiting its effectiveness in discovering negative cases.

To address this issue, we use the Synthetic Minority Oversampling Technique (SMOTE). SMOTE is an increasingly common technique for oversampling the minority class using generated samples. Instead of just copying existing images, SMOTE generates new synthetic samples by applying interpolation to existing negative samples. This strategy captures differences in sickle cell form, giving the model more diverse examples and allowing it to better distinguish between normal and sickle cells. This is how SMOTE works; The underrepresented class (negative) is identified by the algorithm. SMOTE then creates synthetic samples by adding additional data points between the line segments that connect existing samples, starting with the minority sample and working its way up to its k-nearest neighbors. Through this approach, the dataset gains depth and variety without experiencing overfitting.

Proposed SDL-SCD model

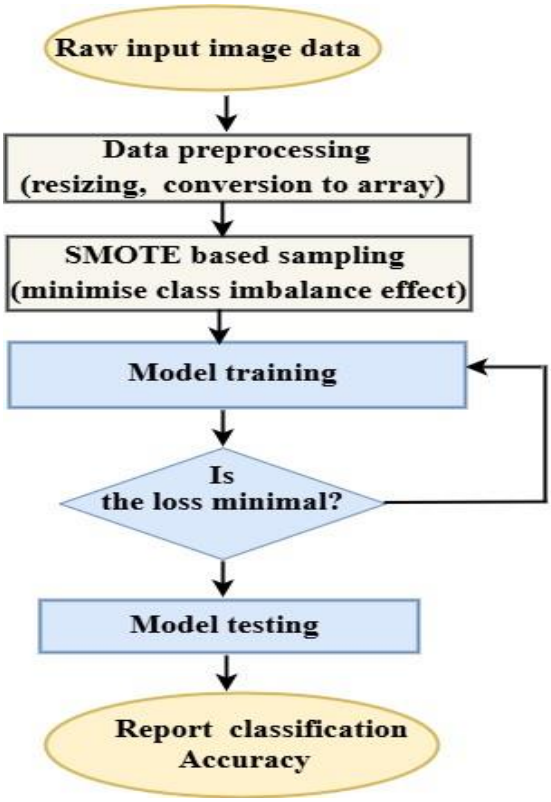

Figure 1: The process starts with acquiring input that is in image format, then a preprocessing phase which includes data cleaning, Images are converted into an array using cv2 and then resized to 150 pixels. The SMOTE-based sampling approach is then performed to oversample the negative images. A CNN-based classification model comprising of the convolution, pooling, and fully connected neurons is trained to extract and learn features for detecting the presence or absence of sickle cells in the image

Figure 6 summarizes the process that starts with acquiring input that is in image format, then a preprocessing phase which includes data cleaning, images are converted into an array using cv2 and then resized to 150 pixels. The SMOTE-based sampling approach is then performed to over-sample the negative

images. A CNN-based classification model composed of fully connected and convolution-constrained neurons is trained to extract and learn characteristics to detect the presence or absence of sickle cells in the image.

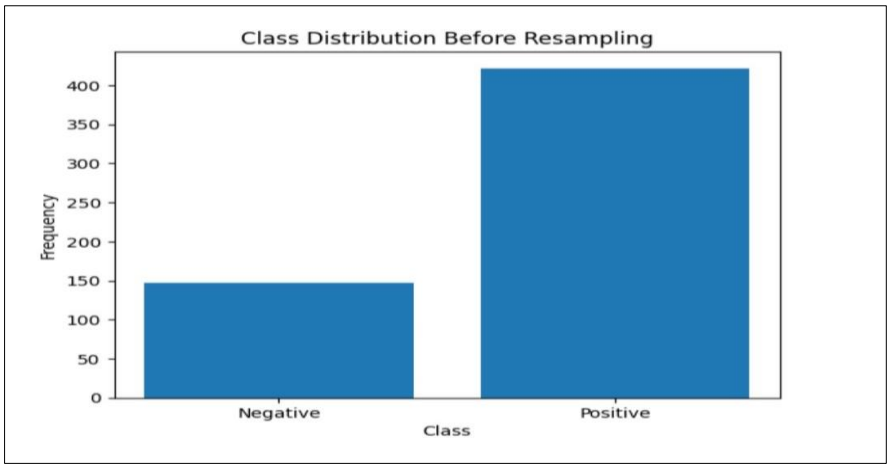

Figure 1: Before sampling

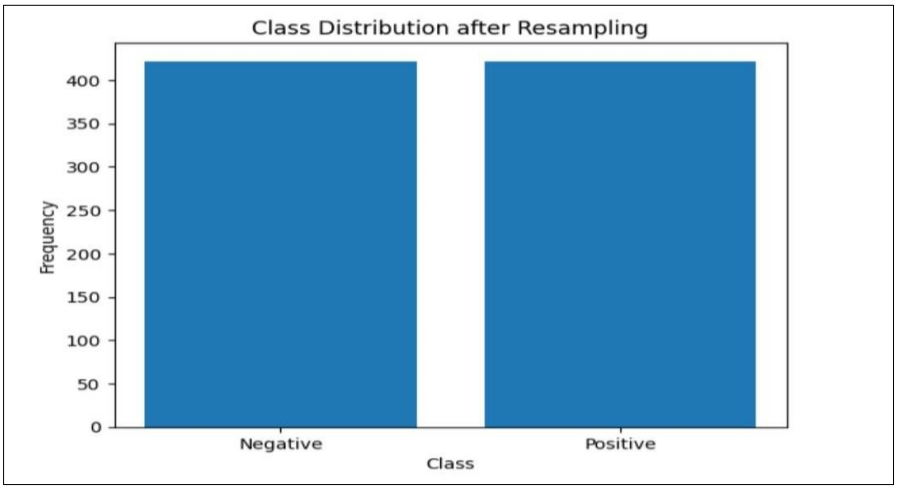

Figure 2: Dealing with imbalances, our dataset before and after applying SMOTE to over-sample the negative images

Thanks to SMOTE based sampling approach, the class imbalanced effect in the dataset where the negative class had fewer samples as compared to the positive one was minimized, This consequently improved the model performance by reducing the model's bias against the negative class.

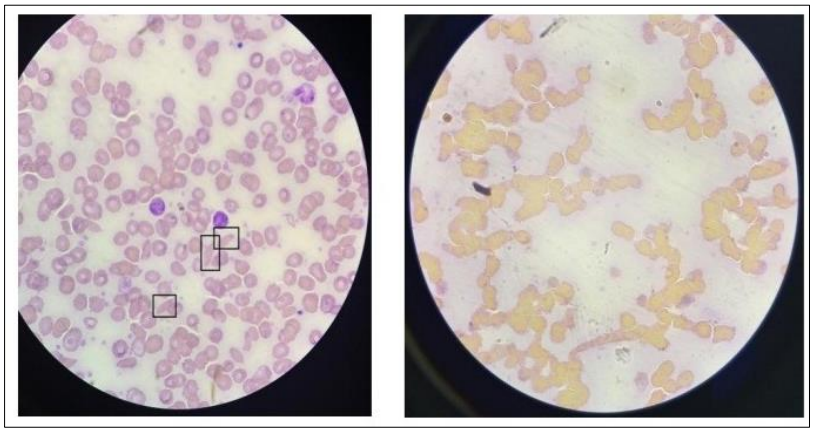

Figure 3: Shows sample of the dataset images. on right is a sample of a positive image and the right is a sample of a negative image justified by the macro package automatically

Training the Model

Our model was trained using CNN. A batch size of 32 was used during the training process to ensure the effective use of computational resources and model convergence. We selected the Adam optimizer because dynamically modifies learning rates to guarantee steady training. A learning rate of 0.0001, which provides a balance between stability and divergence speed, enabling the model to progressively lower the error without overfitting. Categorical cross-entropy was used as the loss function, which is suitable for multi-class classification problems such as differentiating sickle cell victims from healthy individuals. We then trained our model with 50 epochs and generalized the model's performance by tracking its evaluation on the validation data at the end of each epoch. Thereafter, metrics like accuracy and loss over the epochs were plotted aiding in identifying over-fitting. 80% of the dataset was utilized as training data, and 20% was used for validation.

Table 1: Summary of classification results including Precision, Recall, and F1-score.

Algorithm	Accuracy	Precision	Recall	F1-Score
Decision Trees	0.91	0.91	0.91	0.91
Random Forest	0.97	0.97	0.97	0.97
Support Vector Machine	0.97	0.97	0.97	0.97
CNN without SMOTE	0.84	0.78	0.90	0.84
CNN with SMOTE	0.99	0.99	0.99	0.99

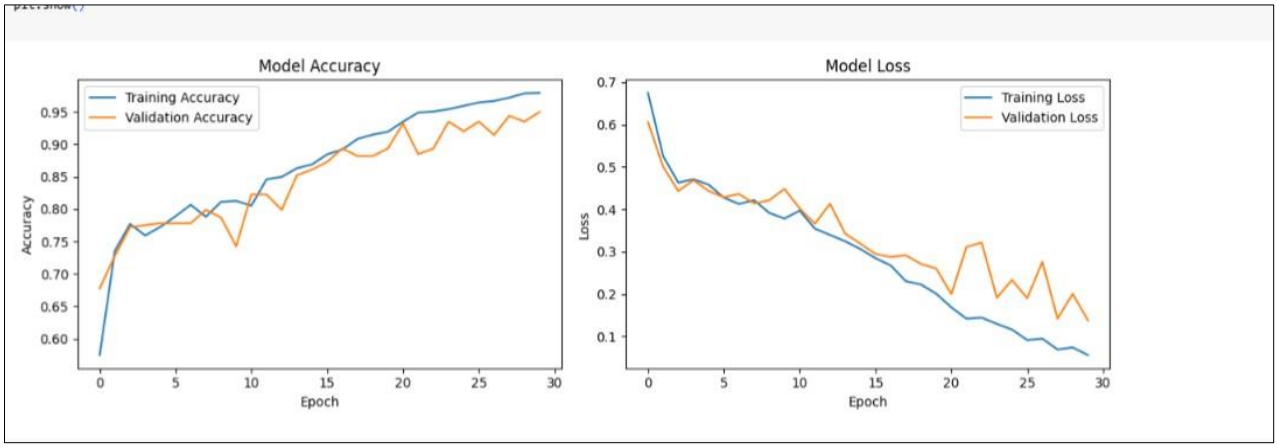

Figure 4: Training and validation and loss curves for a model that was trained with *SMOTE*

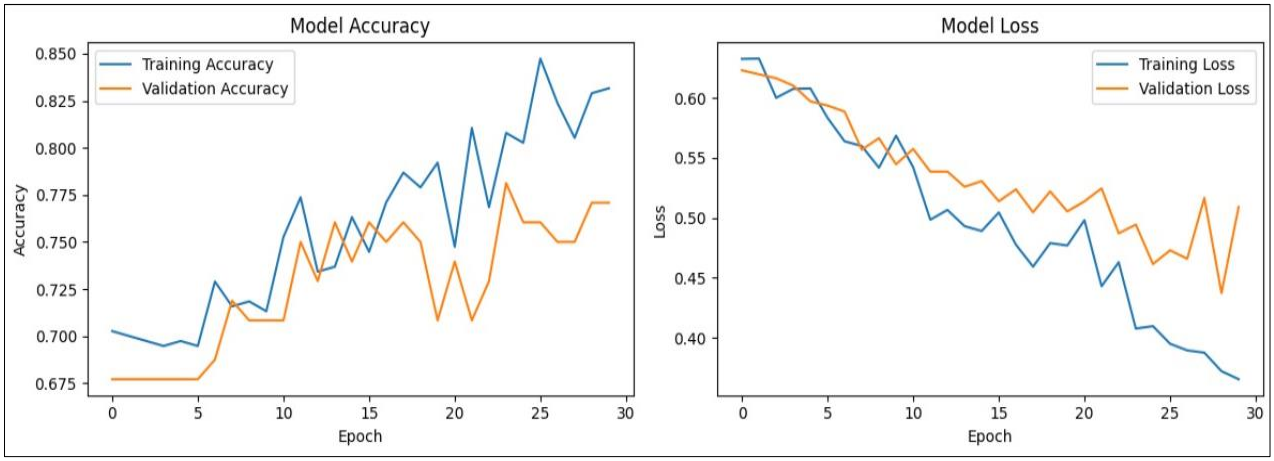

Figure 5: Training and validation loss curves for a model that was trained before applying SMOTE

Comparison Methods

We then compared the results of the model with other algorithms, including, decision trees and random forest Support Vector Machine and we got lower accuracy compared to CNN**??**. We also used images outside the dataset to test our model and gave us accurate results.

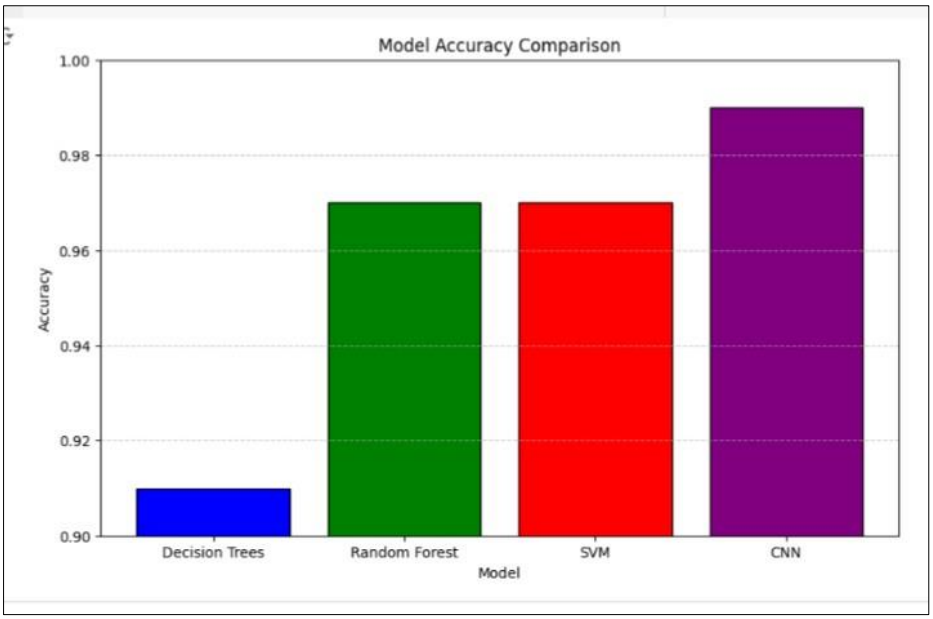

Figure 6: The graph below shows how different models performed

DISCUSSION

To detect sickle cells, we suggested a unique Convolutional Neural Network (CNN) architecture. With the help of convolutional and pooling layers, which apply features to scan the input data and find local patterns, our method makes use of feature extraction, which is a technique that gradually creates more sophisticated feature representations. Through the extraction of these data, the deep CNN is able to learn patterns and combinations that differentiate sickle cells from healthy ones. We were able to decrease the dimensionality of the data through feature extraction, which made it easier to handle and more effective for model training.

Analysis of prediction accuracy measurement with other algorithms

The study evaluated the effectiveness of combining CNN with other algorithms including SVM, Random Forest, and Decision Trees. CNN with SMOTE outperformed SVM with 97% accuracy, random forest with 97%, and decision trees with 91%. However, these models may miss intricate patterns in sickle cell images due to human feature engineering.

Limitations and future work

The small size and imbalanced class distribution of our dataset presented two difficulties.

Although we used SMOTE oversampling to address the class imbalance problem, more advanced approaches could be adopted to boost the classification accuracy. Our Future work aims to enhance the accuracy and robustness of the proposed sickle cell detection model by gathering more data samples from other hospitals and integrating it with real world environments. This will make the model a reliable tool for medical practitioners.

CONCLUSION

We introduce the SDL-SCD model, an automated tool for diagnosing sickle cell anemia in blood samples, offering cost-effectiveness and enhanced productivity. This system reduces manual examination expenses and helps in early attention to the disease. This is a good alternative, especially in low-developed settings where medical services are so contained due to limited finance, technology and trained human expertise. In future work, we hope to integrate the SDL-SCD with transformers and transfer learning approaches to improve sickle cell detection accuracy.

ACKNOWLEDGEMENTS

This research did not receive a grant from any funding agency in the public, commercial or non-profit sectors.

REFERENCES

- [Azar and Wong, 2017] Azar, S. and Wong, T. E. (2017). Sickle cell disease: a brief update. *Medical Clinics*, 101(2):375–393.
- [Barpanda, 2013] Barpanda, S. S. (2013). *Use of image processing techniques to automatically diagnose sickle-cell anemia present in red blood cells smear*. PhD thesis.
- [Chy and Rahaman, 2018] Chy, T. S. and Rahaman, M. A. (2018). Automatic sickle cell anemia detection using image processing technique. In *2018 International conference on advancement in electrical and electronic engineering (ICAEEE)*, pages 1–4. IEEE.

- [Dirksen *et al*., 2011] Dirksen, S. R., Lewis, S. M., Heitkemper, M. M., and Bucher, L. (2011). Clinical companion to medical-surgical nursing : assessment and management of clinical problems.
- [Egesa *et al*., 2022] Egesa, W. I., Nakalema, G., Waibi, W. M., Turyasiima, M., Amuje, E., Kiconco, G., Odoch, S., Kumbakulu, P. K., Abdirashid, S., and Asiimwe, D. (2022). Sickle cell disease in children and adolescents: a review of the historical, clinical, and public health perspective of sub-saharan africa and beyond. *International Journal of Pediatrics*, 2022(1):3885979.
- [Elsalamony, 2017] Elsalamony, H. A. (2017). Anaemia cells detection based on shape signature using neural networks. *Measurement*, 104:50–59.
- [GREEN and CONLEY, 1951] GREEN, T. W. and CONLEY, C. L. (1951). Occurrence of symptoms of sickle cell disease in the absence of persistent anemia. *Annals of Internal Medicine*, 34(4):849–855.
- [Gual-Arnau *et al*., 2015] Gual-Arnau, X., Herold-García, S., and Simó, A. (2015). Erythrocyte shape classification using integral-geometry-based methods. *Medical & biological engineering & computing*, 53:623–633.
- [Kang *et al*., 2024] Kang, H. A., Wang, B., Barner, J. C., Ataga, K. I., Mignacca, R. C., Chang, A., and Zhang, Y. (2024). Opioid prescribing and outcomes in patients with sickle cell disease post–2016 cdc guideline. *JAMA Internal Medicine*, 184(5):510–518.
- [Kavanagh *et al*., 2022] Kavanagh, P. L., Fasipe, T. A., and Wun, T. (2022). Sickle cell disease: a review. *Jama*, 328(1):57–68.
- [Kumar *et al*., 2021] Kumar, R., Joshi, S., and Dwivedi, A. (2021). Cnn-sspso: a hybrid and optimizedcnnapproachforperipheralbloodcellimage recognitionandclassification. *International Journal of Pattern Recognition and Artificial Intelligence*, 35(05):2157004.
- [Li *et al*., 2017] Li, X., Dao, M., Lykotrafitis, G., and Karniadakis, G. E. (2017). Biomechanics and biorheology of red blood cells in sickle cell anemia. *Journal of biomechanics*, 50:34–41.
- [Liu *et al*., 2018] Liu, X., Liang, D., Yan, S., Chen, D., Qiao, Y., and Yan, J. (2018). Fots: Fast oriented text spotting with a unified network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5676–5685.
- [Modell and Darlison, 2008] Modell, B. and Darlison, M. (2008). Global epidemiology of haemoglobin disorders and derived service indicators. *Bulletin of the World Health Organization*, 86(6):480–487.
- [Nader *et al*., 2020] Nader, E., Romana, M., and Connes, P. (2020). The red blood cell—inflammation vicious circle in sickle cell disease. *Frontiers in immunology*, 11:517556.
- [Savitt and Goldberg, 1989] Savitt, T. L. and Goldberg, M. F. (1989). Herrick's 1910 case report of sickle cell anemia: the rest of the story. *Jama*, 261(2):266–271.
- [Sharma *et al*., 2016] Sharma, V., Rathore, A., and 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)*, volume 3, pages 1–5. IEEE.
- [Umar *et al*., 2023] Umar, M. I., Aliyu, F., Abdullahi, M. I., Aliyu, M. N., Isyaku, I., Aisha, B. B., Sadiq, R. U., Shariff, M. I., and Obeagu, E. I. (2023). Assessment of factors precipitating sickle cell crises among under 5-years children attending sickle cell clinic of murtala muhammad specialist hospital, kano. *Blood*, 11:16.
- [Ware *et al*., 2017] Ware, R. E., de Montalembert, M., Tshilolo, L., and Abboud, M. R. (2017). Sickle cell disease. *The Lancet*, 390(10091):311–323.
- [Xu *et al*., 2017] Xu, M., Papageorgiou, D. P., Abidi, S. Z., Dao, M., Zhao, H., and Karniadakis, G. E. (2017). A deep convolutional neural network for classification of red blood cells in sickle cell anemia. *PLoS computational biology*, 13(10):e1005746.

Cite This Article: Kisakye Martha Janepher, Abubakhari Sserwadda, Ali Najib, Tushabe Florence (2025). A SMOTE-Based Deep Learning Approach for Sickle Cell Detection in Low-Resolution, Class-Imbalanced Microscopic Images (SDL-SCD). *East African Scholars J Eng Comput Sci*, *8*(5), 92-98.