Data in Brief 57 (2024) 110950

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Dataset for a novel AI-powered diagnostic tool for *Plasmodium* parasite detection authors

Olumide T. Adeleke[a,b], Halleluyah O. Aworinde[c,g], Mary Oboh[d], Oladipo Oladosu[e], Alaba B. Ayenigba[a], Bukola Atobatele[d], Oludamola V. Adeleke[a], Tunde S. Oladipo[b], Segun Adebayo[f,g,1,*]

[a] *Directorate of Health Services, Bowen University, Iwo, Nigeria*
[b] *Family Medicine Department, Bowen University Teaching Hospital, Ogbomoso, Nigeria*
[c] *Computer Science Department, College of Computing and Communication Studies, Bowen University, Iwo, Nigeria*
[d] *Microbiology, Department, College of Agriculture, Engineering and Science, Bowen University, Iwo, Nigeria*
[e] *Pure and Applied Biology Department, College of Agriculture, Engineering and Science, Bowen University, Iwo, Nigeria*
[f] *Mechatronics Engineering Programme, College of Agriculture, Engineering and Science, Bowen University, Iwo, Nigeria*
[g] *Computing and Analytics Research Laboratory, College of Agriculture, Engineering and Science, Bowen University, Iwo, Nigeria*

ARTICLE INFO

Article history:
Received 20 August 2024
Revised 4 September 2024
Accepted 12 September 2024
Available online 17 September 2024

Dataset link: Blood Smear Images Dataset for AI-based Malaria Parasite Prediction (Original data)

Keywords:
Malaria
Dataset
Signals
Machine learning
Disease

ABSTRACT

Malaria remains a serious public health problem in many developing countries, particularly in Sub-Saharan Africa. Early detection and treatment of malaria are crucial in the fight against malaria in order to reduce morbidity and mortality, especially in the endemic regions. We set out to develop a simple, accurate, and efficient innovative diagnostic tool for malaria parasite identification that uses automated image processing to provide shorter diagnosis times while improving accuracy, efficiency, and standardization. Our primary goal in this study is to collect, curate, annotate and achieve blood smear images containing Plasmodium species for effective malaria diagnosis using Artificial Intelligent based

* Corresponding author at: Mechatronics Engineering Programme, College of Agriculture, Engineering and Science, Bowen University, Iwo, Nigeria.
E-mail address: segun.adebayo@bowen.edu.ng (S. Adebayo).
Social media: @tomolumighty (O.T. Adeleke)
[1] https://x.com/SegunAd55985685?t=rfM7Q4Q5vGZ7FaIUFddNCQ&s=08.

https://doi.org/10.1016/j.dib.2024.110950

system. The study curated 881 blood smear images which are categorized as positive and negative images.

Specifications Table

Subject	Applied Machine Learning.
Specific subject area	Image Processing.
Type of data	Image, Annotated Image
Data collection	We collected from each individual after taking informed consent, about 1 mL of blood and placed it in an EDTA bottle for a malaria test. Malaria blood films were prepared using 20 µL of the EDTA blood. For each participant, approximately 2 µL and 6 µL of blood were used to make thin and thick smears on the same slide, respectively. Two slides were prepared per participant. The blood films were properly air-dried before being stained. Prior to staining the malaria blood slides, absolute methanol was used to fix the thin films by dipping it into absolute methanol to avoid contact with thick film. The slides were then dried and placed on the staining rack. Following that, the slides were stained for 45 min using 3 % Giemsa stain. The stained smears were gently rinsed off with clean water and the slides were carefully dried, kept in a slide box before being examined under a microscope, A digital screen light microscope was used to examine stained thick and thin films with an objective lens of × 100 (immersion oil). Slides were reported negative after scanning through 100 high power fields (HPF) without a malaria parasite. Every slide was reviewed by a second malaria microscopist. Malaria parasites were identified using the World Health Organization bench aids [1].
Data source location	Ogbomoso, Iwo and Port Harcourt in Nigeria
Data accessibility	Repository name: Mendeley Data identification number: DOI: 10.17632/6zpxnhjxzz.1 Direct URL to data: https://data.mendeley.com/datasets/6zpxnhjxzz/1
Related research article	None.

1. Value of the Data

- These datasets would be used to develop a fast, accurate, simple, affordable, efficient, and sensitive novel diagnostic tool for malaria parasite detection that can be made easily available for laboratories located in malaria-endemic regions worldwide.
- Malaria remains a serious public health challenge particularly in Nigeria and many other countries in sub-Saharan Africa, despite being treatable and preventable [2]. Early detection and treatment of malaria are essential in the fight against malaria diseases to lower rates of morbidity and mortality, particularly in developing nations where malaria is endemic [3].
- There is paucity of publicly available datasets on malaria parasite images. Making this annotated dataset obtained from this project aligns perfectly well with the principles of promoting findability, accessibility, interoperability, and reproducibility (FAIR) of data [4].
- The dataset will help to optimize the reuse of these data in the development of computer vision-based models for further segmentation, and classification.
- The dataset will give more insights beyond human physical observation for accurate and precise detection of malaria parasites. It will help to engage laboratory technicians in a more effective training and practice.

2. Background

With millions of malaria cases and thousands of malaria-related deaths occurring globally each year, the female Anopheles mosquito, transmitting malaria parasites, is the insect that has

caused the greatest amount of death and suffering in human history. Timely and accurate diagnosis is crucial in the malaria control and elimination efforts. However, in many low and middle-income countries, early and accurate malaria diagnosis is being hampered by the demand of microscopic detection of plasmodium parasites for technical expertise, time and labor-intensiveness. Malaria Rapid Diagnostic Tests (mRDTs) offer easier access but have their well-recognized limitations.

Our overarching research objective is to annotate blood smear images, extract features from the data, develop a model in the diagnosis of malaria and compare the model with expert microscopy employed in diagnosing malaria from infected individuals.

The gold standard for diagnosing malaria, microscopy, is a laborious process that relies on human experience/technique and power supply, both of which are somewhat not readily available in many malaria endemic regions [5]. The malaria rapid diagnostic test (mRDT) may not be very sensitive as a result of low parasite density [6] and the prozone phenomenon [7].

To address these shortcomings, we are developing a novel AI-powered diagnostic tool for malaria parasite detection. This fast, accurate, simple, and affordable method has the potential to be readily available in malaria-endemic regions worldwide, a critical step towards combating this devastating infectious disease. Our goal is to revolutionize malaria diagnostics by using automated image analysis that offers faster diagnosis times along with improved accuracy, efficiency, and standardization. The system will aid the detection of malaria parasites in blood samples with a high precision rate compared to the existing manual intervention. Further to this, we hope to reduce to the barest minimum the risk of human errors associated with manual microscopy.

3. Data Description

The data used in this study were sourced from three malaria endemic states in Nigeria. Patients with confirmed *Plasmodium species* infections and those whose blood had no parasite in them were enrolled into this study.

Samples from Bowen University Hospital Iwo and Bowen University Teaching Hospital Ogbomoso were collected from February to June 2024. While those from Cross River were collected between August 2023 and May 2024, and the study participants were children under the age of 10 years.

The dataset can be grouped into two categories: The first category is the raw dataset which has a total of 881 images in the dataset. This category contains two folders: slides_with_negative_cells and slides_with_positive_cells. There were 550 positive images in the positive folder and 331 images in the negative folder as depicted in Table 1. The other category is the annotated dataset which contain blood smear images with their equivalent annotation file in the label folder. The annotation file contains the coordinates of the annotations on the image and the object categories as shown in Fig. 1.

Positive Images refers to images containing any species of the malaria parasites- Plasmodium malarie, Plasmodium ovale, and Plasmodium falciparum as shown in Fig. 1(a), (b) and

Table 1
Dataset statistics.

Data source	Number of participants	Positive images	Negative images
Bowen University Hospital Iwo, Osun State, Nigeria.	354	297	57
Cross River, Nigeria, Nigeria.	227	100	127
Bowen University Teaching Hospital Ogbomoso, Oyo State, Nigeria.	300	153	147

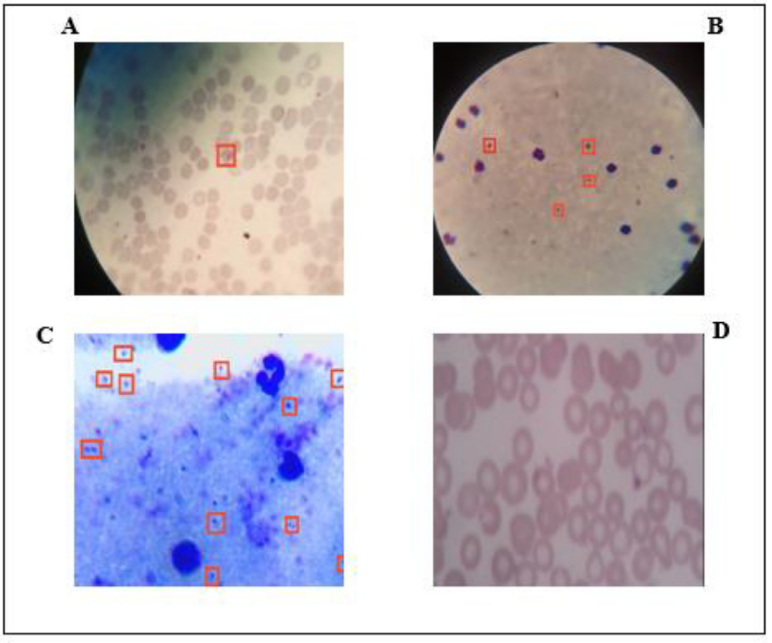

Fig. 1. Sampled Annotated Dataset (a) Plasmodium malarie, (b) Plasmodium ovalae, (c) Plasmodium falciparum (d) Negative image (no malaria parasite present).

(c) respectively while a negative image refers to an image without any of the species of plasmodium as shown in Fig. 1(d).

4. Experimental Design, Materials and Methods

Consented patients from Osun State, Oyo State and Cross River had their blood samples collected for malaria testing. Both thick and thin blood films were prepared from each participant for Plasmodium falciparum characterization. The blood films were allowed to air-dry, and stained with 3 % Giemsa stain for 45 min, after which they were allowed to dry and viewed under the 100× magnification. Microscopic images from the 881 participants recruited were captured in the dataset. Blood films slides were collected from 3 different locations as shown in Fig. 1. The slides were inspected by the domain experts and grouped into two categories to create a dataset: cells with malaria parasite (Positive Images) and cells without malaria parasite (Negative Images). The dataset was then uploaded to a Computer Vision Annotation Tool (CVAT) platform as shown in Fig. 3 for the annotation of regions with object of interests (Plasmodium malarie, Plasmodium ovalae, and/or Plasmodium falciparum) (Fig. 2).

The CVAT is an open-source project by Sekachev et al. [1]. It prioritizes efficiency in image annotation tasks by incorporating deep learning functionalities. These functionalities include features like automatic object detection and semi-automatic segmentation. CVAT functions as a web application, with docker being the primary installation method. While it operates as a web app, CVAT's core objective isn't crowdsourced annotation. Instead, it emphasizes reproducible builds through containerization and user-friendly access through web browsers. When exporting the annotated images, they are accompanied with annotations stored in JavaScript object notation (JSON) file format. These annotations follow the You Only Look Once (YOLO) format [8]. With this format, the coordinates of the annotations on the image and the object categories are stored for each image.

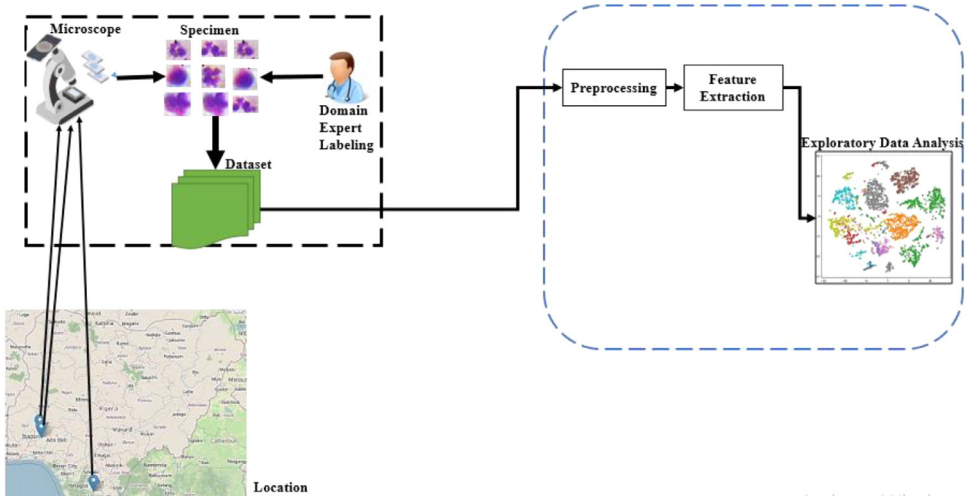

Fig. 2. Data Acquisition experimental setup.

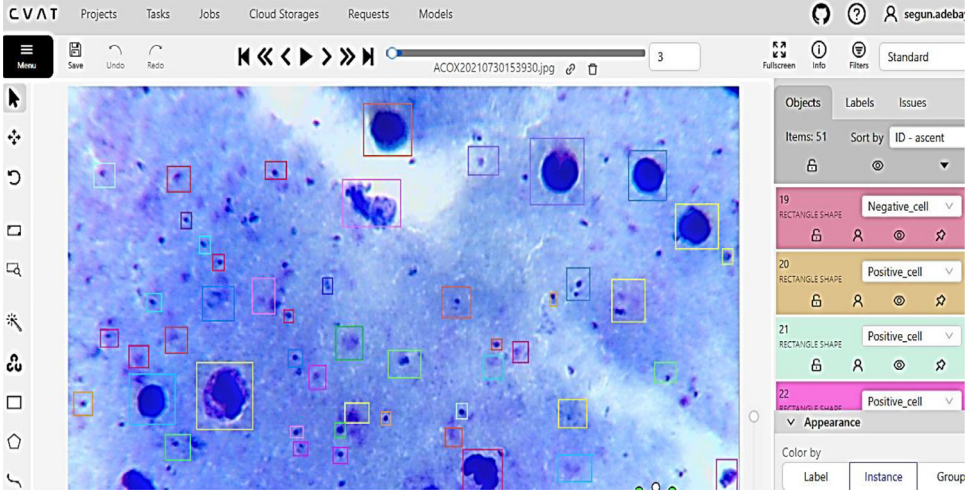

Fig. 3. Annotation of the datasets using CVAT tool.

Limitations

Not applicable.

Ethics Statement

Ethical approvals for this study were obtained from Bowen University Teaching Hospital Research Ethical Board and the Cross River State Ministry of Health Research Ethics Committee. Informed consents were obtained from all participants prior enrolment after detailed explanation of the study.

Ethics Governing Body name: Bowen University Teaching Hospital Research Ethics Committee. Protocol Number: BUTH/REC-1101.

Informed consent was obtained from all the study participants.

CRediT Author Statement

Alaba B. Ayenigba, Olumide T. Adeleke, Mary Oboh, Oladipo Oladosu and **Tunde S. Oladipo** participated in data collection and labelling of the image dataset collected from Bowen University Hospital, Bowen University Teaching Hospital (BUTH), and Port-Harcourt

Segun Adebayo, Halleluyah Aworinde, Mary Oboh, Oladipo Oladosu segmented and annotated the images.

Olumide T. Adeleke, Halleluyah O. Aworinde, Mary Oboh, Alaba B. Ayenigba, Oladipo Oladosu, Bukola Atobatele, Oludamola V. Adeleke and Segun Adebayo participated in writing and proofreading the manuscript.

All of the authors gave their approval to the final version after offering constructive criticism and helping to develop the research, analysis, and manuscript.

Data Availability

Blood Smear Images Dataset for AI-based Malaria Parasite Prediction (Original data) (Mendeley Data).

Acknowledgments

We gratefully appreciate the management of Bowen University, Iwo (BURG/2024/03) for funding this study. We also acknowledge and appreciate all our patients who participated in this study. Many thanks to all the staff of the hospitals where data collection procedures were carried out.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2024.110950.

References

[1] WHO, Bench Aids for Malaria Microscopy, WorldHealthOrganization, Geneva, 2009 2009.
[2] O.T. Adeleke, A. Oyenuga, T.M. Slusher, D.A. Gbadero, Cluster-randomized controlled trial of intermittent preventive treatment in infancy using sulfadoxine–pyrimethamine (SP-IPTi): a pilot study in Nigeria, J. Trop. Pediatr. 69 (1) (2023) fmad001.
[3] World Health Organization, World Malaria Report, Geneva: World Health Organization, 2022.
[4] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.W. Boiten, Silva da, L.B. Santos, P.E. Bourne, J. Bouwman, The FAIR guiding principles for scientific data management and stewardship, Sci. Data 3 (1) (2016) 1–9.

[5] World Health Organization, Global Technical Strategy for Malaria 2016–2030, World Health Organization, 2015.
[6] J. Maltha, P. Gillet, J. Jacobs, Malaria rapid diagnostic tests in endemic settings, Clin. Microbiol. Infect. 19 (5) (2013) 399–407.
[7] K.E. Poti, D.J. Sullivan, A.M. Dondorp, C.J. Woodrow, HRP2: transforming malariadiagnosis, but with caveats, Trends Parasitol. 36 (2) (2020) 112–126.
[8] L. Rahma, H. Syaputra, A.H. Mirza, S.D. Purnamasari, Objek Deteksi Makanan Khas Palembang Menggunakan Algoritma YOLO (You Only Look Once), J. Nasional Ilmu Komput. 2 (3) (2021) 213–232.