

RESEARCH

Open Access



# CAM: a novel aid system to analyse the coloration quality of thick blood smears using image processing and machine learning techniques

W. M. Fong Amaris<sup>1,2\*</sup>, Daniel R. Suárez<sup>3</sup>, Liliana J. Cortés-Cortés<sup>4</sup> and Carol Martinez<sup>5\*</sup>

## Abstract

**Background** Battling malaria's morbidity and mortality rates demands innovative methods related to malaria diagnosis. Thick blood smears (TBS) are the gold standard for diagnosing malaria, but their coloration quality is dependent on supplies and adherence to standard protocols. Machine learning has been proposed to automate diagnosis, but the impact of smear coloration on parasite detection has not yet been fully explored.

**Methods** To develop Coloration Analysis in Malaria (CAM), an image database containing 600 images was created. The database was randomly divided into training (70%), validation (15%), and test (15%) sets. Nineteen feature vectors were studied based on variances, correlation coefficients, and histograms (specific variables from histograms, full histograms, and principal components from the histograms). The Machine Learning Matlab Toolbox was used to select the best candidate feature vectors and machine learning classifiers. The candidate classifiers were then tuned for validation and tested to ultimately select the best one.

**Results** This work introduces CAM, a machine learning system designed for automatic TBS image quality analysis. The results demonstrated that the cubic SVM classifier outperformed others in classifying coloration quality in TBS, achieving a true negative rate of 95% and a true positive rate of 97%.

**Conclusions** An image-based approach was developed to automatically evaluate the coloration quality of TBS. This finding highlights the potential of image-based analysis to assess TBS coloration quality. CAM is intended to function as a supportive tool for analyzing the coloration quality of thick blood smears.

**Keywords** Thick blood smears, Coloration quality, Image processing, Machine learning, Malaria diagnosis

\*Correspondence:

W. M. Fong Amaris  
we\_fong@javeriana.edu.co; wfong110@gmail.com

Carol Martinez  
carol.martinezluna@uni.lu

<sup>1</sup> Pontificia Universidad Javeriana, Faculty of Engineering, Bogotá, Colombia

<sup>2</sup> Universidade Federal do Pará, Institute of Biological Sciences, Belém, Brazil

<sup>3</sup> Facultad de Ingeniería, Pontificia Universidad Javeriana, Bogotá, Colombia

<sup>4</sup> Laboratory of Parasitology, National Health Institute of Colombia, Bogotá, Colombia

<sup>5</sup> Space Robotics (SpaceR) Research Group, Interdisciplinary Centre for Security, Reliability, and Trust (SnT), University of Luxembourg, Luxembourg, Luxembourg



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Malaria is an infectious disease and one of the most critical public health problems worldwide. According to the World Health Organization (WHO), in 2022, there were 249 million malaria cases, an increase of 5 million cases compared with 2021 [1]. In Colombia, where malaria outbreaks are common in rural areas, health posts and patients must deal with geographical limitations, violence-related problems, and severe climate conditions, among other factors. These conditions affect the dynamics of rural health posts, which must streamline the process, dealing quickly with all the cases (conducting fast sample preparation and quick and accurate diagnosis) [2, 3].

The microscopic diagnostic procedure is considered the “gold standard” method worldwide, and thick blood smears (TBS) is the reference method chosen as the first option for malaria diagnosis [4–8]. TBS allows quantitative results to monitor patients by analysing the parasitaemia before, during, and after the treatment [6, 9].

Studies have been conducted to automate malaria diagnosis using images from thin and thick smears. However, regularly, those studies have limitations, such as the exclusive use of high-quality images that do not include common imperfections like the presence of image artifacts, staining troubles, and factors related to the non-standardization of the image acquisition process, for example, illumination and the type of functions used from the microscope where they acquired the photographs [10–20].

Based on the literature review, the primary focus of published works is on devising strategies to facilitate malaria diagnosis, aiming for quick and parasite detection. Many papers highlight the issues associated with inconsistencies in the staining of thin and thick smears [10–14, 17–22]. These inconsistencies are particularly prevalent in TBS, as noted by multiple studies [13, 14, 21, 23, 24]. However, an unaddressed area in the research is the analysis of the quality of smear coloration. The variations in smear coloration could potentially impact malaria diagnosis, indicating a need for further investigation into this factor.

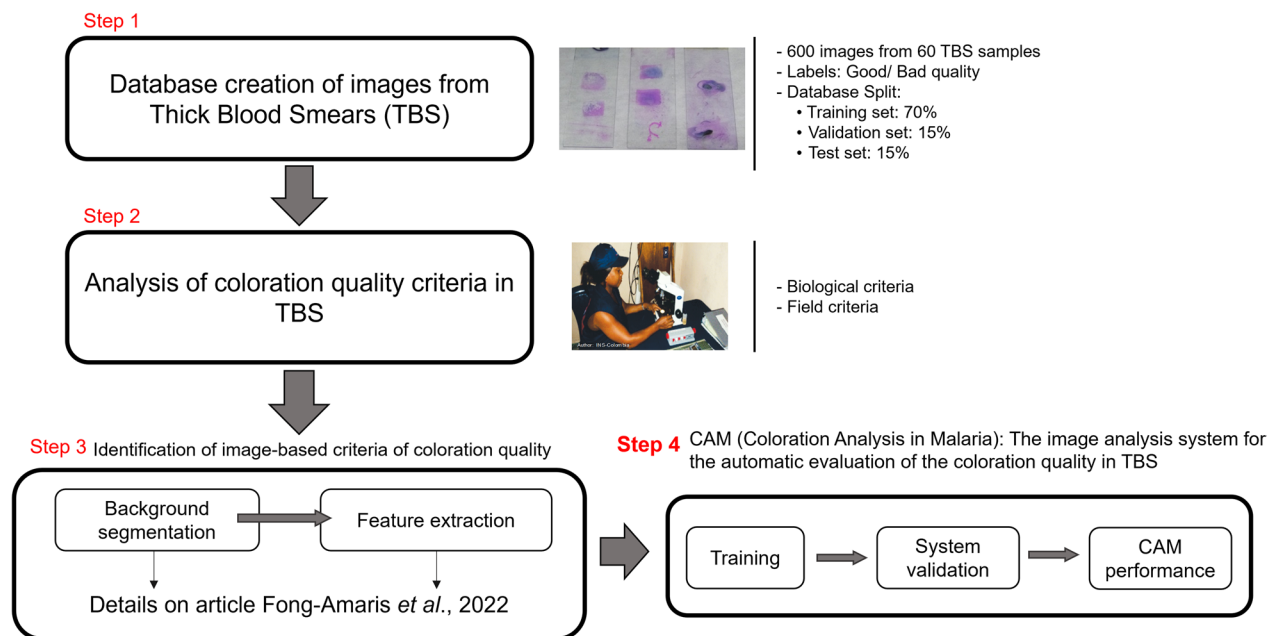
Malaria diagnosis, involving a roughly one-hour sample collection, preparation, and visual examination using TBS, is subject to considerable variability and subjectivity [4, 25]. Influencing factors include the conditions and location of sample preparation, the availability and quality of equipment and supplies like stains, and minor changes in stain concentration or pH levels of the staining solution or washing water [26]. The staining stage is particularly critical as it directly impacts the visibility of parasites and, thus, the accuracy of diagnosis [27]. However, this stage involves a

subjective estimation process where the determination of coloration depends on human judgment and the microscopists' experience, adding another layer of subjectivity [26–29]. The latter underscores the importance of stringent quality control and standardization in malaria diagnosis to minimize these factors' impact. Recognizing the limitations previously discussed, in this work, Coloration analysis in malaria (CAM) is a system that uses image processing and machine learning algorithms. CAM is designed to assess the colour quality of TBS and automatically verify if they meet the essential quality criteria, such as appropriate background colour, thereby improving the process of microscopic examination in malaria diagnosis.

In previous works [28, 29], it was conducted an extensive image processing analysis to determine the best features for evaluating the background coloration quality in TBS. The current paper describes the workflow of CAM and a comprehensive evaluation of machine learning techniques that are the basis for the presented analyses. A database of 600 images of *Plasmodium vivax* TBS with balanced classes (good/bad coloration quality) was created to train, validate, and test machine-learning classifiers. To discriminate the two quality classes (good or bad), this paper presents the tuning of the parameters and the selection process of the best feature vector and classifier.

## Methods

The four-step methodology presented in Fig. 1 was followed to develop the image-based approach to automatically evaluate the coloration quality of TBS (stained with the modified Romanowsky stain [4, 30, 31]). The first step corresponds to creating the database with images from TBS with good and bad coloration quality. The database was used to select image-based criteria of coloration quality and to establish and evaluate the machine learning classifier of CAM. The second step corresponds to the analysis of coloration quality criteria used in malaria diagnosis laboratories to assess the quality of TBS. The third step corresponds to identifying image-based criteria to evaluate the coloration quality [29]. In this step, different image-based processing techniques were tested to select the most suitable features to identify the coloration quality. Finally, the fourth step corresponds to selecting and tuning a classifier that automatically evaluates the coloration quality in thick blood smear images. Steps 1, 2, and 3 were presented and described in [29]. This paper focuses on Step 4, the methodology's feature and classifier selection steps.



**Fig. 1** Methodological steps followed to evaluate the staining of a malaria smear by analysing the coloration quality of TBS

### Image database of TBS

A database containing images and labels was created to design and evaluate different image-based strategies to determine the quality of the staining procedure for TBS stained with modified Romanowsky stain. Most literature on parasite detection in TBS uses staining with Giemsa [20, 32, 33] and Jaswant Singh Battacharya stain [13]. However, to the authors' knowledge, no study has used databases with images from TBS stained with modified Romanowsky stain. Nor are publicly available databases related to the coloration quality of TBS (good/bad).

The images were captured following a protocol to ensure that the acquired images contained the same colour information as that seen through the microscope. Details of the protocol can be found in [29]. The optical microscope (Axio Zeiss Scope.A1, Brand: Carl Zeiss, Country: Germany) was used to create the database. The 100X microscope lens used for the microscopic visualization has the reference: Objective-lens A-Plan 100×/1.25 Oil M27, Carl Zeiss, Germany.

The study population corresponds to malaria cases due to *P. vivax* in 2017 and 2018. *Plasmodium vivax* is the predominant parasite in Central and South America, representing 75% of malaria cases [34]. From a complete set of 60 TBS, 600 photographic records (10 images of different fields per slide) were obtained. The captured areas correspond to fields close to the sample's centre to ensure uniformity in the images (in periphery regions, the thickness of the blood can vary), as depicted in Fig. 2. The images acquired through the optical microscope are

2056×2452 pixels and were saved in Portable Network Graphics (PNG) format.

The 600 images were labelled using the web-based labelling tool Labelbox [35]. Two labels were created for the background coloration quality, which can be good (1) or bad (2). The determination of TBS quality (good or bad) was defined based on the background colour of the sample as it was described in previous work [29] and according to the established by the INS-Colombia [25], WHO [4] and other works [36, 37]. If the background showed a pink or light blue colour with pink areas in the microscopic field, it was considered bad quality. Otherwise, the image was classified as having good coloration quality (Table 1). Finally, the database (600 images) was randomly split into training (70%), validation (15%), and test (15%) sets, as shown in Table 1.

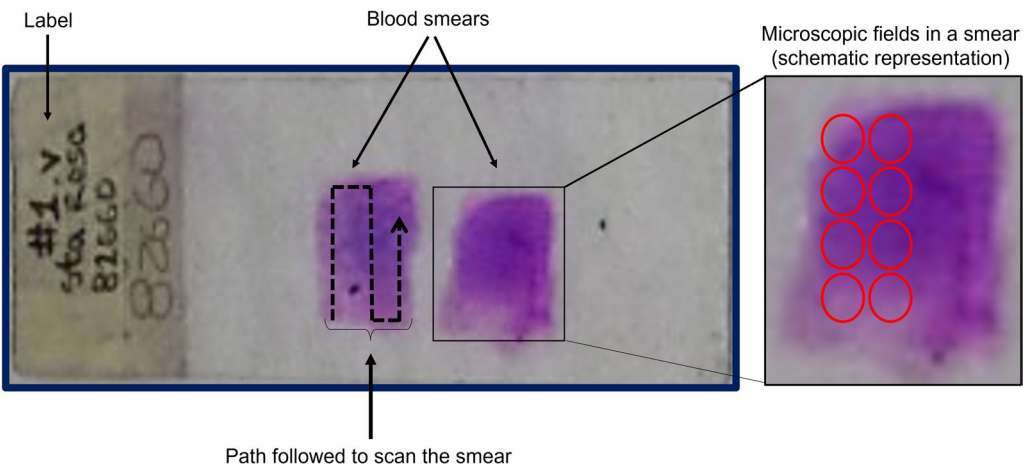
All sets contained balanced classes. The database instructions and labels can be found using the <https://doi.org/10.5281/zenodo.7191423> [38].

### Coloration quality criteria

Depending on the staining procedure, the staining of a thick blood smear can be of good or bad quality, and the presence of parasites can be easily detected or not.

### Biological criteria

Factors such as the background colour of a thick smear, the colour of the nucleus and the cytoplasm from leukocytes and parasites (Fig. 3), the colour and morphology of the platelets, and minor differences in pH are indicator



**Fig. 2** Indication about how a TBS is read when malaria is diagnosed. The two staining squares represent two blood drops spread over a slide. The left written label corresponds to smear identification. It can be numerical or descriptive. The dashed arrow shows the route of the microscopic visualization recommended by the WHO to guarantee a correct reading. The red circles indicate the schematic representation of 8 microscopic fields

**Table 1** Database of TBS stained with the modified Romanowsky stain

Labels	Classification	Quantity	Database split		
			Training	Validation	Testing
Coloration quality	Good	300	210	45	45
	Bad	300	210	45	45

factors of the quality of the stain in a thick smear. They give information about how well-stained a smear is. There are more details about this section in a previous article where it was mentioned that the appearance of the smear background is one of the most critical features that indicates information about the coloration quality in a thick blood smear [29].

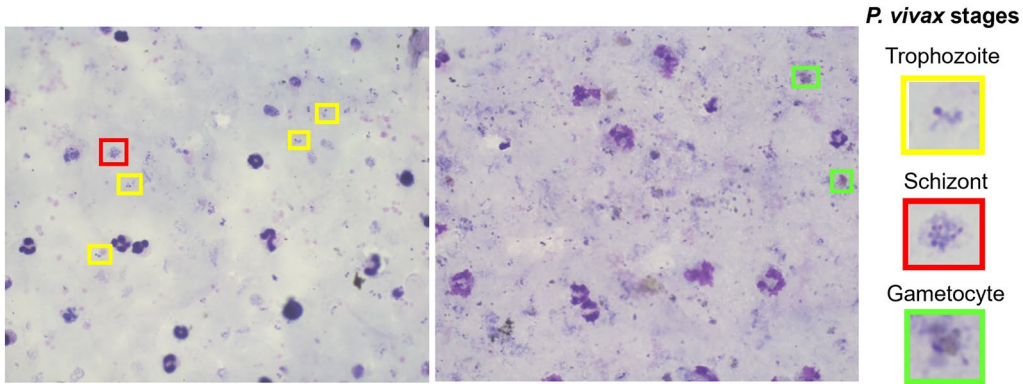
Image-based criteria of quality

Background segmentation

The biological criteria of quality showed that the appearance of the background of the smear is a crucial feature in evaluating the quality of the coloration. Therefore, from the microscopic images of the TBS, the background information had to be separated from the foreground information (leukocytes, parasites, and platelets) to ensure that the image-based analysis was conducted only on the background information.

The microscopic image is captured in the RGB colour space. However, this image can be transformed into other colour models, called colour spaces, to extract specific features. Colour as a feature can give us valuable information to characterize the coloration in TBS.

In the literature, previous works focused on parasite detection and classification have mainly used the RGB



**Fig. 3** Two microscopic images obtained from TBS with good coloration quality. This figure shows examples of the life cycle stages of the parasite *P. vivax*, where the yellow, red, and green squares correspond to the trophozoite, schizont, and gametocyte stages, respectively



and HSV colour spaces [10, 11, 13, 14, 18–21, 32, 39, 40]. These colour spaces have been used to segment the parasite nucleus, contrast enhancement, erythrocyte segmentation, and even background removal in images from thin blood smears [41]. HSV colour space has been used in studies related to parasite detection [13] and classification [32] to segment foreground elements such as parasites, leukocytes, and platelets. In the case of the study done by Chakrabortya et al. [13], they used the H and S components to extract the foreground elements. However, their TBS were stained with the Jaswant Singh Battacharya stain, which gives different colour saturation to the biological elements from the smear [13].

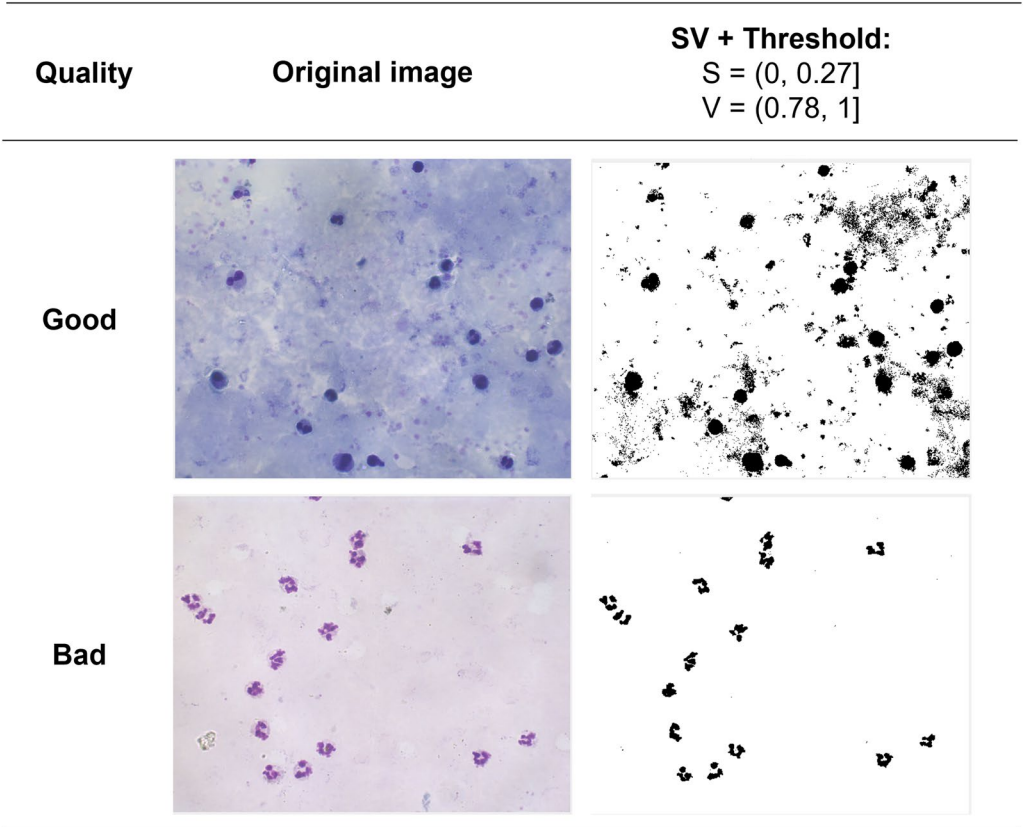
Because previous work had not studied the background information as a feature for estimating the coloration quality in TBS, an exploratory analysis with different components of the HSV and RGB colour spaces (analyzed in our previous work) [28, 29] was conducted to find the best threshold and the best colour space for background segmentation. The best colour space was carefully selected, ensuring stained leukocytes were not confused as background information. The S and V

components were selected as they showed similar background segmentation capability for TBS images of good and bad coloration quality. The threshold values used were  $S = (0, 0.27]$  and  $V = (0.78, 1]$  (Fig. 4).

**Feature extraction**

Computer vision techniques applied to parasite detection and classification in blood smears have been used in previous studies to respond to the need for new alternatives to reduce malaria by 2030 [42]. Combined with machine learning, image features such as colour, texture, shape, and size are commonly used for parasite detection [13, 14, 20, 21, 23, 24, 32]. However, the challenge is to include the preprocessing stages of the smears in the analysis, including varying parameters for image acquisition and slide preparation [16, 26]. Any automatic malaria diagnosis system should include those critical steps before the classification system.

Despite the few studies focused on TBS [13, 14, 20, 21, 23, 24, 32], some have used features based on histograms for malaria parasite detection, as in Hanif et al. [23] and Salamah et al. [24]. These studies used threshold-based



**Fig. 4** Threshold selection from the S and V combined components from the HSV colour space. SV + Threshold: the RGB image was transformed into the HSV colour space. From this image, it was obtained a threshold by each component (S and V), and the pixels that fell under the threshold were considered background; otherwise, they were foreground

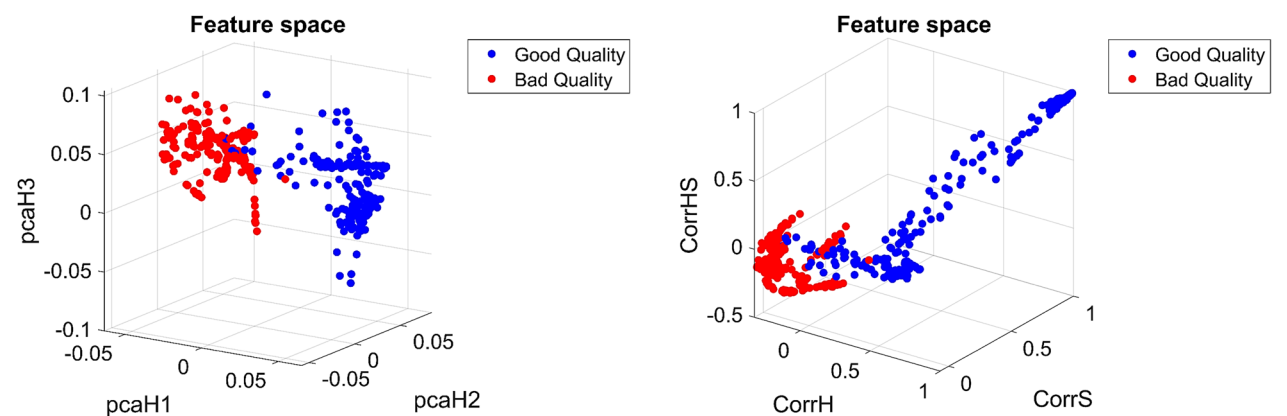
segmentation and colour-based histograms, respectively. On the other hand, the machine-learning techniques for malaria diagnosis have focused on parasite detection and classification using k-means clustering and support vector machines [16]. According to previous research in the field, a histogram-based feature extraction procedure was considered one of the strategies for our work.

The analysis of 19 feature vectors involved variances, correlation coefficients, and histograms (specific variables from histograms, full histograms, and principal components from the histograms). In previous work [29], the ability of 19 feature vectors to classify images based on good and bad coloration quality was analyzed. As a result, four feature vectors ( $[H_{PCA1}, H_{PCA2}, H_{PCA3}]$ ,  $[H_{corr}, Scorr, HScorr]$ ,  $Hist_{15-H}$ , and  $[H_{PCA1}, H_{PCA2}, H_{PCA3}, S_{PCA1}, S_{PCA2}, S_{PCA3}, HS_{PCA1}, HS_{PCA2}, HS_{PCA3}]$ ) were selected as candidates to describe the background information according to the coloration quality. Figure 5

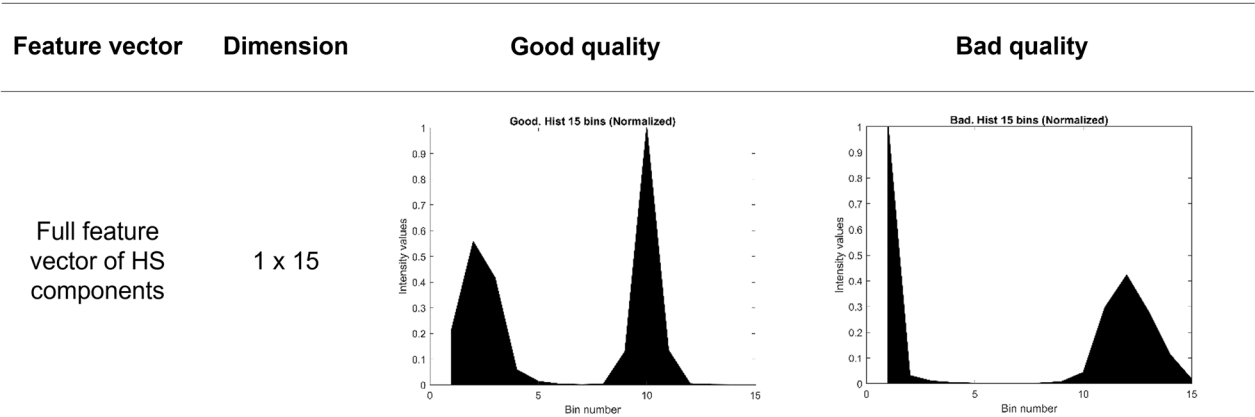
shows the feature space of two feature vectors such as  $[H_{PCA1}, H_{PCA2}, \text{ and } H_{PCA3}]$  and  $[H_{corr}, Scorr, HScorr]$ . The histogram obtained for  $Hist_{15-H}$  is shown in Fig. 6.

**Machine learning classifiers**

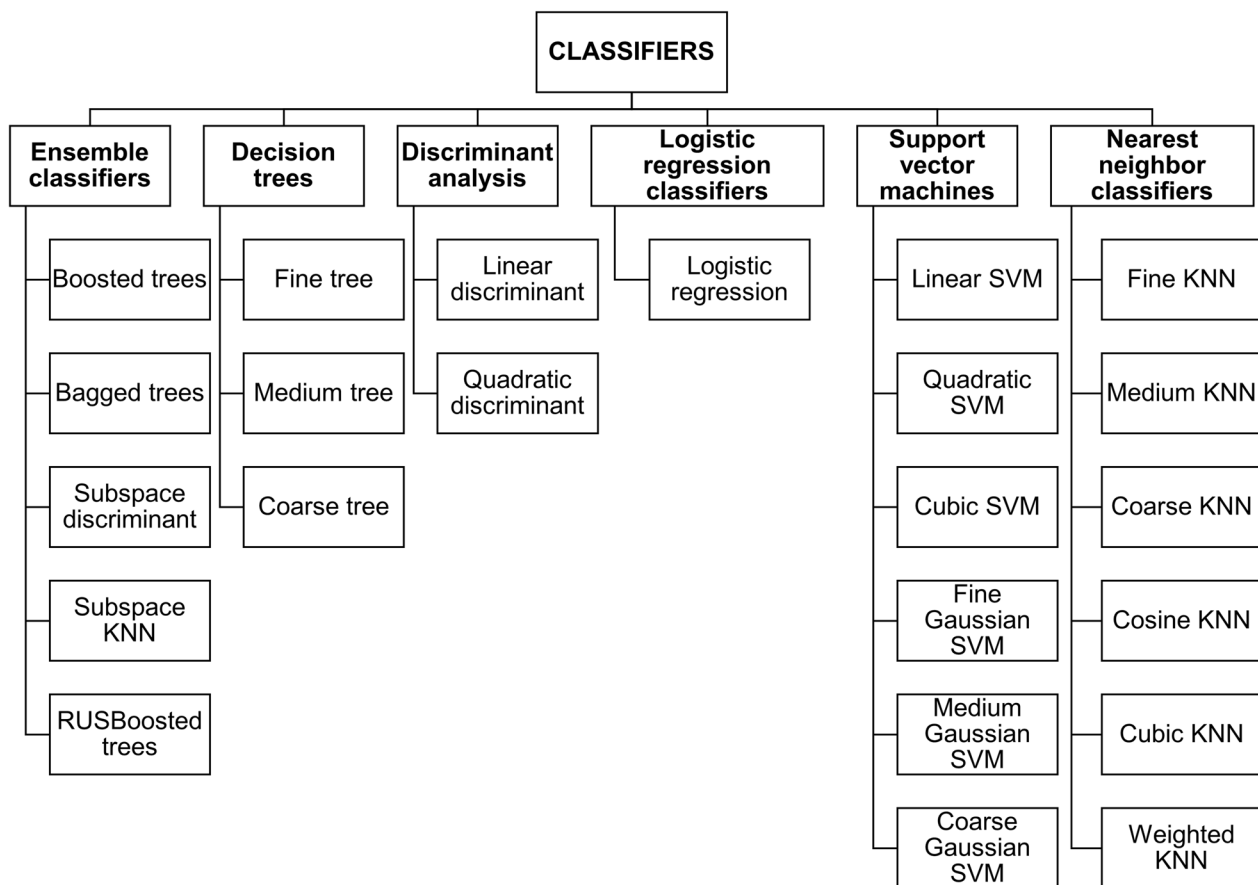
The Machine Learning Matlab Toolbox [43] was used to select the best feature vector and machine learning classifier. The 23 classifiers available in the toolbox (Fig. 7) were trained with their default parameters, using as feature vectors the ones presented in section “Feature extraction”. The training and validation set from the database (“Image database of TBS” section) were used. The True Negative Rate (TNR), the True Positive Rate (TPR), and the F1 score were analysed to evaluate the results. Nevertheless, the TNR was considered the most critical metric. To predict ‘good’ when in reality ‘bad’ is the attribute for assessing coloration quality could harm patient health. A smear that is not



**Fig. 5** Features space for the  $[H_{PCA1}, H_{PCA2}, H_{PCA3}]$  (at left) and  $[H_{corr}, Scorr, HScorr]$  (at right) features vector. The red points represent the images with bad coloration quality, and the blue points represent the images with good coloration quality



**Fig. 6** Histograms using 15 bins with HS combined components. The histograms did not include the black pixels from leukocytes. The figure shows the data distribution: the left shape in each graph corresponds to the pixel distribution for the S component, and the right shape shows the pixel distribution for the H component



**Fig. 7** Matlab classifiers used for selecting the preliminary 4-best feature vectors

stained correctly can lead to errors in detecting the parasite. Therefore, it may result in errors in calculating patients' dosage and parasite density. That is why TNR was the interest metric. This metric would facilitate the evaluation of the system's ability to discriminate bad-quality samples.

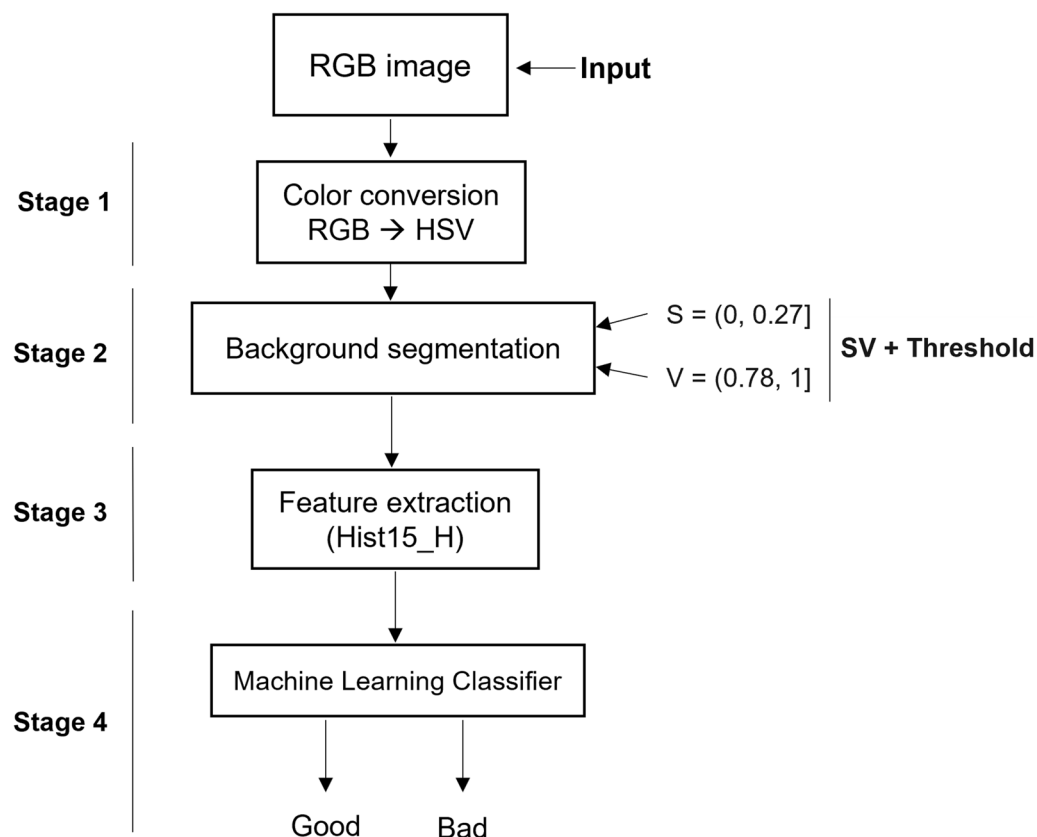
Table S3 in Supplementary material shows the classifiers that presented the best training results in terms of the TNR and F1 score in each of the tested feature vectors (SVM Quadratic, SVM Cubic, SVM Medium Gaussian, SVM Fine Gaussian, Logistic Regression).

Since most classifiers achieved accuracy rates exceeding 95% across all candidate feature vectors (including  $[H_{PCA1}, H_{PCA2}, H_{PCA3}]$ ,  $[H_{corr}, S_{corr}, H_{Scorr}]$ ,  $H_{ist15\_H}$ , and  $[H_{PCA1}, H_{PCA2}, H_{PCA3}, S_{PCA1}, S_{PCA2}, S_{PCA3}, H_{S_{PCA1}}, H_{S_{PCA2}}, H_{S_{PCA3}}]$ ), Quadratic SVM and Cubic SVM were selected for further evaluation due to their superior performance. The classifiers mentioned before were used to develop the system adjustments (Table S2 in Supplementary material).

#### CAM: coloration analysis in malaria

CAM is an algorithm that automatically analyses the coloration quality of TBS. It was designed using MATLAB® Student, Version R2018a, Mathworks, USA [43]. The proposed algorithm consists of four stages, as presented in Fig. 8 and described below.

- Stage 1: As input, CAM receives an image in RGB format, which is then transformed into the HSV colour space.
- Stage 2: The histograms of each component of HSV were extracted. The S and V components were integrated into a histogram called SV, which was thresholded to segment the foreground elements (leucocytes and parasites) from the background, obtaining the thresholded image. To isolate the background information of interest in the HSV color space, the thresholded image was multiplied by the original HSV image, resulting in only the background information remaining and the foreground elements appearing black.



**Fig. 8** Algorithm for the automatic quality analysis of TBS in malaria called Coloration Analysis in Malaria (CAM)

- Stage 3: This stage relates to feature extraction, as found in [28, 29], that extracting features in the HSV colour space makes it possible to find image criteria of coloration quality. The feature vector is used as input for Stage 4. With that information, two possible features can be extracted, which have been demonstrated to provide good results: the 15-bin histogram of the H channel (Hist<sub>15\_H</sub>) and the three main PCA components from the histogram of the H channel [ $H_{PCA1}$ ,  $H_{PCA2}$ ,  $H_{PCA3}$ ].
- Stage 4: A machine learning classifier is in charge of estimating the coloration quality of the image background. The algorithm's output will be “good” if the quality of the coloration is good or bad if the background coloration quality of the image is not appropriate for diagnosis, according to the image's criteria of quality that was presented in [28, 29].

A system like CAM can be used in isolation to evaluate the abilities of microscopists, e.g., when assessing microscopists' smear staining capabilities, to support microscopists by making them aware that a careful analysis of each microscopic field is required, or as a subsystem of an automated malaria diagnosis system.

## Results

The first test below presents the process for selecting the best feature vector and classifier for CAM. Once selected, CAM's performance is evaluated in terms of TNR, TPR, accuracy, misclassification rate, and F1 Score.

### Classifier optimization

The Quadratic SVM and the Cubic SVM classifiers were selected for tuning to optimize the coloration quality estimation algorithm. Code for adjustments was obtained from the Machine Learning Toolbox in MATLAB [43].

The default values for the Quadratic and Cubic SVM include a polynomial kernel function, the polynomial order (2 for Quadratic SVM and 3 for Cubic SVM), and the automatic kernel scale (default value 1). The supplementary material (Table S3) contains details about the results obtained during the optimization process of the classifiers. Table S3 summarizes the final candidate classifiers and the features and parameters that showed the best results for assessing the background quality using the validation set. According to the results, the best features that can be used to analyse the coloration quality in thick blood smear images were feature vectors 1 and 3, respectively: [ $H_{PCA1}$ ,  $H_{PCA2}$ ,  $H_{PCA3}$ ] and (Hist<sub>15\_H</sub>).



**CAM performance**

The algorithm’s performance was evaluated using the test set from the database, considering both the selected classifier and the two candidate feature vectors. The obtained rates are presented in Table 2. Analysis of the results revealed that one of the feature vectors, Hist<sub>15</sub>\_H, exhibited superior predictive behaviour (TNR: 95%, TPR: 97%, F1-score: 98%) than the [H<sub>PCA1</sub>, H<sub>PCA2</sub>, H<sub>PCA3</sub>] feature vector (TNR: 95%, TPR: 95%, F1-score: 95%). The latter is because the values (in terms of TPR, Accuracy, and F1-score) obtained from the feature vector Hist<sub>15</sub>\_H were higher than 95%. Although the misclassification rate was lower in the feature vector [H<sub>PCA1</sub>, H<sub>PCA2</sub>, H<sub>PCA3</sub>], the TNR, TPR, accuracy, and F1-score rates were also considered to make the selection.

The confusion matrixes in Table 3 indicate the classifier performance obtained using the feature vectors [H<sub>PCA1</sub>, H<sub>PCA2</sub>, H<sub>PCA3</sub>] and [Hist<sub>15</sub>\_H].

Based on Table 3, the images where the algorithm did not estimate the quality correctly were analysed. Table 3 shows images corresponding to two good and bad coloration quality. For the analysis, it is essential to consider that a good background coloration quality is related to a blue colour, a homogeneous sample spreading, and other factors mentioned in [4, 25]. The images in Fig. 9 present differences in the coloration background.

Analysing those images, the wrong labels occurred with images that contained parasite presence. The images without parasites did not present that misclassification. In the images shown in Fig. 9, three presented a non-homogeneous background. This means that the blood sample was not spread adequately above the slide. Thus, the colour information is not homogeneous,

which could affect the label. On the contrary, the image c44\_9 (Fig. 9) showed a homogeneous background.

Analysis revealed differences in the classifier’s predictive ability for good quality images when using each feature vector. Notably, the Cubic SVM classifier exhibited one additional misclassification using the [H<sub>PCA1</sub>, H<sub>PCA2</sub>, H<sub>PCA3</sub>] feature vector compared to Hist<sub>15</sub>\_H.

Images c7\_10 and c44\_9 from Fig. 9 show examples of false-negative predictions according to the ground truth column. Using both feature vectors ([H<sub>PCA1</sub>, H<sub>PCA2</sub>, H<sub>PCA3</sub>] and Hist<sub>15</sub>\_H), the background of the image c7\_10 (left image from Fig. 9) was wrongly predicted. This image presented a no homogeneous stain distribution and a low sample thickness. The second image (c44\_9, right image from Fig. 9) was only misclassified by the feature vector [H<sub>PCA1</sub>, H<sub>PCA2</sub>, H<sub>PCA3</sub>]. This image presents a pale blue colour in the background and a sample spreading of type homogeneous. The higher bin number of the feature vector Hist<sub>15</sub>\_H than the feature vector [H<sub>PCA1</sub>, H<sub>PCA2</sub>, H<sub>PCA3</sub>] allows for distinguishing different colour tonalities, which helps correct the image classification.

On the other hand, in the images with bad coloration quality, the classifier with the two different feature vectors misclassified the same images (c67\_8 and c23\_1). Both images present a pale blue colour in the background (blue background is characteristic of good coloration quality). However, the images show a non-homogeneous colour distribution with pink colour areas related to the misclassification in both feature vectors (there were areas without blood-stained samples).

The results described above about the predictive capability of the feature vectors suggest that the Hist<sub>15</sub>\_H feature vector is the best vector for describing and classifying the background coloration quality in TBS stained with modified Romanowsky stain.

Thus, the feature vector Hist<sub>15</sub>\_H and the Cubic SVM classifiers were selected to be part of CAM.

**Table 2** Performance of the cubic SVM classifier predicting the coloration quality of the images of the test set

Features	Rates				
	Misclassif <sup>a</sup>	TNR	TPR	Accuracy	F1-score
[H <sub>PCA1</sub> , H <sub>PCA2</sub> , H <sub>PCA3</sub> ]	0.0190	0.9556	0.9556	0.9556	0.9556
Hist <sub>15</sub> _H	0.0214	0.9556	0.9778	0.9667	0.9888

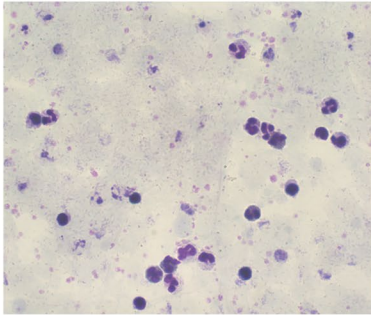
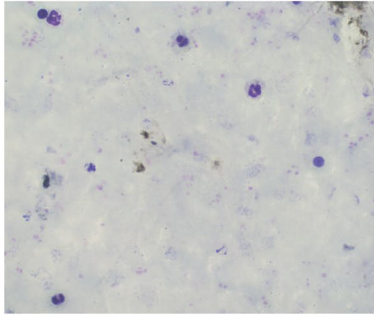
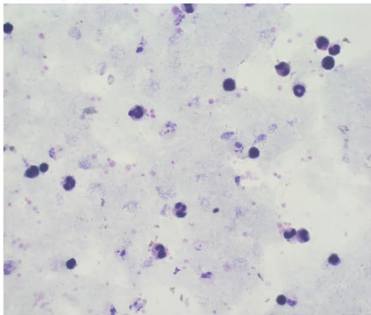
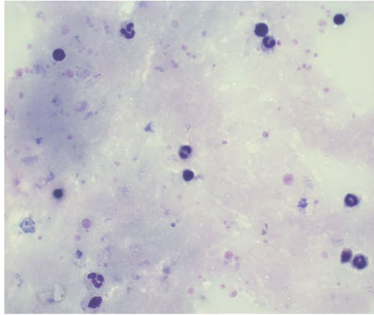
<sup>a</sup> Misclassif.: out-of-sample misclassification rate

**Discussion**

Currently, available literature has focused on methods to support malaria diagnosis. Previous studies have noted that the coloration quality of images obtained

**Table 3** Confusion matrixes obtained for each feature vector ([H<sub>PCA1</sub>, H<sub>PCA2</sub>, H<sub>PCA3</sub>] and [Hist<sub>15</sub>\_H]) using the cubic SVM

[H <sub>PCA1</sub> , H <sub>PCA2</sub> , H <sub>PCA3</sub> ]			[Hist <sub>15</sub> _H]		
Total: 90	Predicted: bad quality	Predicted: good quality	Total: 90	Predicted: bad quality	Predicted: good quality
Actual: bad quality	43	2	Actual: bad quality	43	2
Actual: good quality	2	43	Actual: good quality	1	44

Ground truth	Predicted labels by the Cubic SVM							
	[HPCA1, HPCA2, HPCA3]			Hist15_H	[HPCA1, HPCA2, HPCA3]			Hist15_H
	Bad	c7_10	Bad		Bad	c44_9	Good	
Good								
Bad								

**Fig. 9** Errors by the Cubic SVM using the [H<sub>PCA1</sub>, H<sub>PCA2</sub>, H<sub>PCA3</sub>] and Hist<sub>15</sub>\_H feature vectors

from TBS complicates the comparison of studies involved in automatic parasite detection. Despite achieving reasonable sensitivity and specificity rates, the different systems developed so far cannot be compared due to differences in coloration quality [16, 44].

This paper presents CAM, the first algorithm proposed to automate the analysis of coloration quality in TBS. This is the first report on using an SVM classifier in studies that analyse TBS coloration quality for malaria diagnosis.

According to Table 2, the algorithm achieves a True Positive Rate (TPR) of 97% in predicting the background quality in images with good coloration. The best results for predicting coloration quality were obtained with the Hist<sub>15</sub>\_H feature vector, corresponding to the 15-bin histogram of the Hue channel from the HSV colour space. The following evidence supports the previously reported findings [28, 29]. The study represents a pioneering step towards analysing image quality, providing a basis for comparing studies by considering the number of images with superior and inferior coloration quality and their respective performance outcomes.

This study’s findings offer initial steps towards automating the assessment of coloration quality procedures

in Departmental Laboratories of Public Health to ensure effective diagnostics. Thus, CAM is introduced as a novel aid system for analysing the coloration quality of thick blood smears using image processing and machine learning techniques.

While CAM represents a significant advancement in automating the analysis of coloration quality in thick blood smears (TBS) for malaria diagnosis, it is important to acknowledge its limitations. For instance, deploying such technology in the field, where expert microbiologists are scarce, faces significant practical challenges related to budgeting. Despite this challenge, CAM’s contribution to standardizing the evaluation of TBS coloration quality offers a novel approach that complements traditional methods, especially in resource-limited settings where expert microscopists may not be accessible.

**Conclusions**

This work successfully engineered a novel aid system named CAM. It evaluates the staining procedures used during sample preparation, offering a new perspective on automated malaria diagnosis that can lead to more accurate and efficient disease detection.

After testing, the 15-bin histogram of the 'H' component from the HSV colour space was the most effective feature vector for training CAM. Built on a Cubic SVM classifier, CAM demonstrated a True Negative Rate (TNR) of 95% and a True Positive Rate (TPR) of 97%.

CAM could help ensure accurate malaria diagnosis in remote diagnostic centres, regardless of variations in smear staining. This solution is a supportive tool for analysing the coloration quality of thick blood smears, empowering microscopists to examine each microscopic field carefully when necessary. Moreover, this solution holds potential for external performance evaluations conducted by official laboratories. Microscopist abilities, including smear staining capabilities, are closely monitored and evaluated in these settings. Implementing CAM can facilitate objective assessments that are not subject to evaluation personnel's visual perception biases, improving accuracy and consistency in malaria diagnosis.

#### Abbreviations

TBS	Thick blood smears
INS-Colombia	Instituto Nacional de Salud de Colombia (National Institute of Health-Colombia)
WHO	World Health Organization
SVM	Support vector machine
TNR	True negative rate
TPR	True positive rate

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12936-024-05025-7>.

Supplementary Material 1: Table S1. Summary of the TNR, TPR, and F1-score obtained by each classifier during training. Table S2. Rates obtained during the system adjustment. These values were obtained from the validation set using the selected four features. Table S3. Features and parameters that showed the best results using the validation set.

#### Acknowledgements

The authors thank the Colombia National Institute of Health for supplying the thick blood smear samples used in this research, Martha Ayala for her invaluable assistance as a Malaria expert advisor, and the field laboratory staff responsible for malaria diagnosis who participated in the interviews. Lastly, we thank Professor Martha Manrique from Pontificia Universidad Javeriana for her indispensable support throughout the image acquisition process.

#### Author contributions

Conceptualization: WMFA, CM, DRS. Methodology: WMFA, CM. Investigation: WMFA. Visualization: WMFA, DRS, CM, LJCC. Funding acquisition: WMFA, DRS, CM. Project administration: DRS, CM. Supervision: DRS, CM, LJCC. Writing—original draft: WMFA. Writing—review and editing: WMFA, DRS, LJCC, CM.

#### Funding

WM Fong Amaris and this project were partly funded by the M.Sc. program in Bioengineering at Pontificia Universidad Javeriana and Facebook Inc. as a winner of the CV4GC 2019 RFP Research Award (Award Number 3418118842). The project was further supported by Pontificia Universidad Javeriana (IDS PPTA 9053 and PRY 9411).

#### Availability of data and materials

The dataset generated and analyzed during the current study is available at ResearchGate in the link we provided in the methodology section (<https://doi.org/10.5281/zenodo.7191424>). The database is published on: <https://drive.google.com/drive/folders/1Qrv0e4b5EtkeqtPABz-kIQp-6D6OjU-X?usp=sharing>. The database instructions of the database can be found on: [https://www.researchgate.net/publication/359826910\\_Database600Labels\\_Instructionstxt](https://www.researchgate.net/publication/359826910_Database600Labels_Instructionstxt). The labels of the database are published on: [https://www.researchgate.net/publication/359827216\\_Database\\_600Labels.csv](https://www.researchgate.net/publication/359827216_Database_600Labels.csv).

#### Declarations

##### Competing interests

The authors declare no competing interests.

Received: 12 January 2024 Accepted: 25 June 2024

Published online: 07 October 2024

#### References

1. WHO. World malaria report 2023. Geneva: World Health Organization; 2023. p. 356.
2. Ospina OL, Cortés LJ, Cucunubá ZM, Mendoza NM, Chaparro P. Characterization of the national malaria diagnostic network, Colombia, 2006–2010. *Biomedica*. 2016;32:46.
3. Cardona-Arias JA, Salas-Zapata W, Carmona-Fonseca J. A systematic review of mixed studies on malaria in Colombia 1980–2022: what the “bifocal vision” discovers. *BMC Public Health*. 2023;23:1169.
4. WHO. Basic malaria microscopy, learner's guide. 2nd ed. Geneva: World Health Organization; 2010. p. 80.
5. Ministerio de Salud y Protección Social. Guía de Práctica Clínica Diagnóstico y tratamiento de la malaria. Ministerio de Salud y Protección Social Colombia. 2022. p. 1–167. <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/PP/ET/Guia-atencion-clinica-malaria.pdf>.
6. Jazmín L, Cortés C, Patricia A, Vega G. Recomendación técnica sobre el diagnóstico y seguimiento al tratamiento de malaria por gota gruesa y pruebas rápidas y uso de pruebas diagnósticas moleculares en la detección de *Plasmodium* spp. en Colombia. Bogotá: Instituto Nacional de Salud de Colombia; 2020. p. 1–10.
7. WHO. Guidelines for malaria. Geneva: World Health Organization; 2023. p. 1–451. <http://apps.who.int/bookorders>.
8. Maturana CR, de Oliveira AD, Nadal S, Bilalli B, Serrat FZ, Soley ME, et al. Advances and challenges in automated malaria diagnosis using digital microscopy imaging with artificial intelligence tools: a review. *Front Microbiol*. 2022;13:1006659.
9. INS & OMS/OPS. Guía de atención clínica integral del Paciente con malaria. 2010.
10. Abdul-Nasir AS, Mashor MY, Mohamed Z. Modified global and modified linear contrast stretching algorithms: new colour contrast enhancement techniques for microscopic analysis of malaria slide images. *Comput Math Methods Med*. 2012;2012:1–16.
11. Bairagi VK, Charpe KC. Comparison of texture features used for classification of life stages of malaria parasite. *Int J Biomed Imaging*. 2016;2016:7214156.
12. Zou LH, Chen J, Zhang J, García N. Malaria cell counting diagnosis within large field of view. In: *Proceedings—2010 digital image computing: techniques and applications, DICTA*; 2010. p. 172–7.
13. Chakrabortya K, Chattopadhyayb A, Chakrabarti A, Acharyad T, Kr DA. A combined algorithm for malaria detection from thick smear blood slides. *J Health Med Inform*. 2015;6:179.
14. Elter M, Haßmeyer E, Zerfaß T. Detection of malaria parasites in thick blood films. In: *Proceedings of the annual international conference of the IEEE engineering in medicine and biology society, EMBS*. 2011.
15. Khan NA, Pervaz H, Latif A, Musharaff A. Unsupervised identification of malaria parasites using computer vision. *Pak J Pharm Sci*. 2017;30:223–8.

16. Poostchi M, Silamut K, Maude RJ, Jaeger S, Thoma G. Image analysis and machine learning for detecting malaria. *Transl Res*. 2018;194:36–55. <https://doi.org/10.1016/j.trsl.2017.12.004>.
17. Prasad K, Winter J, Bhat UM, Acharya RV, Prabhu GK. Image analysis approach for development of a decision support system for detection of malaria parasites in thin blood smear images. *J Digit Imaging*. 2012;25:542–9.
18. Rahman A, Zunair H, Sohail Rahman M, Quader Yuki J, Biswas S, Ashraf Al Alam M, et al. Improving malaria parasite detection from red blood cell using deep convolutional neural networks. *arXiv:190710418*. 2019. p. 1–33.
19. Tek FB, Dempster AG, Kale I. Parasite detection and identification for automated thin blood film malaria diagnosis. *Comput Vis Image Underst*. 2010;114:21–32.
20. Yang F, Poostchi M, Yu H, Zhou Z, Silamut K, Yu J, Maude RJ, Jaeger S, Antani S. Deep learning for smartphone-based malaria parasite detection in thick blood smears. *IEEE J Biomed Health Inform*. 2020;24(5):1427–38. <https://doi.org/10.1109/JBHI.2019.2939121>.
21. Yunda L, Alarcón A, Millán J. Automated image analysis method for *P. vivax* malaria parasite detection in thick film blood images. *Sistemas y Telemática*. 2011;10:9.
22. Omuchení DL, Kaduki K, Bulimo W, Angeyo HK. Application of principal component analysis to multispectral-multimodal optical image analysis for malaria diagnostics. *Malar J*. 2014;13:485.
23. Hanif NSMM, Mashor MY, Mohamed Z. Image enhancement and segmentation using dark stretching technique for *Plasmodium falciparum* for thick blood smear. In: *IEEE 7th international colloquium on signal processing and its applications, CSPA*. 2011. p. 257–60.
24. Salamah U, Sarno R, Arifin AZ, Nugroho AS, Gunawan M, Pragesjvara V, et al. Enhancement of low quality thick blood smear microscopic images of malaria patients using contrast and edge corrections. In: *International conference on knowledge creation and intelligent computing (KCIC)*. 2016. p. 219–25.
25. INS & MinSalud. Manual para el diagnóstico de malaria no complicada en puestos de diagnóstico y tratamiento. 2015. p. 1–74. <https://www.minsa.gob.pe/sites/rid/Lists/BibliotecaDigital/RIDE/IA/INS/manual-diagnostico-malaria-no-complicada.pdf>.
26. Linder N, Turkki R, Walliander M, Mårtensson A, Diwan V, Rahtu E, et al. A malaria diagnostic tool based on computer vision screening and visualization of *Plasmodium falciparum* candidate areas in digitized blood smears. *PLoS ONE*. 2014;9: e104855.
27. Fong Amaris WM. Estudio de la calidad de láminas de gota gruesa, para el diagnóstico de malaria por *Plasmodium vivax*, provenientes de pacientes del Departamento de Bolívar. [Tesis de pregrado]. Universidad del Atlántico; 2017.
28. Fong Amaris WM. Image processing for quality analysis of thick blood smears employed in malaria diagnosis [Tesis de Maestría]. Pontificia Universidad Javeriana; 2020. <https://repository.javeriana.edu.co/handle/10554/386>.
29. Fong Amaris WM, Martinez C, Cortés-Cortés LJ, Suárez DR. Image features for quality analysis of thick blood smears employed in malaria diagnosis. *Malar J*. 2022;21:74.
30. OMS. Bases del diagnóstico de paludismo. Organización Mundial de la Salud. 2014;25(1):345–400. [http://apps.who.int/iris/bitstream/handle/10665/164468/9789243547824\\_spa.pdf?sequence=1](http://apps.who.int/iris/bitstream/handle/10665/164468/9789243547824_spa.pdf?sequence=1).
31. WHO. Malaria microscopy quality assurance manual—2nd version. Geneva: World Health Organization; 2016. p. 140.
32. Kaewkamnerd S, Uthairat C, Intarapanich A, Pannarut M, Chaotheing S, Tongsima S. An automatic device for detection and classification of malaria parasite species in thick blood film. *BMC Bioinform*. 2012;13(Suppl 17):S18.
33. Maturana CR, de Oliveira AD, Nadal S, Serrat FZ, Sulleiro E, Ruiz E, et al. iMAGING: a novel automated system for malaria diagnosis by using artificial intelligence tools and a universal low-cost robotized microscope. *Front Microbiol*. 2023;14:1240936.
34. WHO. World malaria report 2019. Geneva: World Health Organization; 2019. <https://www.who.int/publications-detail/world-malaria-report-2019>.
35. Labelbox Online. 2023. Labelbox: the leading training data platform for data labeling. <https://labelbox.com>. Accessed 6 Dec 2023.
36. Field JW, Sandosham AF. A morphological study of the erythrocytic parasites in thick blood films. In: *The microscopical diagnosis of human malaria*. 2nd ed. Kuala Lumpur: Institute for Medical Research, Federation of Malaya; 1963. p. 30.
37. López-Antuñano F, Schmuniz G. Diagnóstico de malaria. Washington, DC: Publicación; 1988. p. 143.
38. Fong Amaris WM, Martinez C, Cortés-Cortés LJ, Suárez DR. Database that contains all images (plus 180 more) employed in the article: "Image features for quality analysis of thick blood smears employed in malaria diagnosis". Zenodo. 2022. <https://doi.org/10.5281/zenodo.7191424>.
39. Kalkan SC, Sahingoz OK. Deep learning based classification of malaria from slide images. In: *IEEE J Biomed Health Inform*. 2019. p. 1–4.
40. Mushabe MC, Dendere R, Douglas TS. Automated detection of malaria in Giemsa-stained thin blood smears. In: *Proceedings of the annual international conference of the IEEE engineering in medicine and biology society, EMBS*. 2013. p. 3698–701.
41. Hegde RB, Prasad K, Hebbar H, Singh BMK. Development of a robust algorithm for detection of nuclei of white blood cells in peripheral blood smear images. *J Med Syst*. 2018;42:110. <https://doi.org/10.1007/s10916-018-0962-1>.
42. OMS. Estrategia Técnica Mundial Contra la Malaria 2016–2030. 2016.
43. The MathWorks. Matlab student. EEUU: Mathworks; 2018. <https://www.mathworks.com>.
44. Loddó A, Di Ruberto C, Kocher M. Recent advances of malaria parasites detection systems based on mathematical morphology. *Sensors*. 2018;18:513.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.