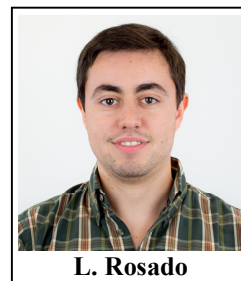


A Review of Automatic Malaria Parasites Detection and Segmentation in Microscopic Images

Luís Rosado^{1*}, José M. Correia da Costa^{2*}, Dirk Elias¹ and Jaime S. Cardoso³

¹Fraunhofer Portugal AICOS, Porto, Portugal; ²Instituto Nacional de Saúde Dr. Ricardo Jorge, Porto, Portugal; ³INESC TEC and University of Porto, Porto, Portugal



L. Rosado

Abstract: *Background:* Malaria is a leading cause of death and disease in many developing countries, where young children and pregnant women are the most affected groups. In 2012, there were an estimated 207 million cases of malaria, which caused approximately 627 000 malaria deaths. Around 80% of malaria cases occur in Africa, where the lack of access to malaria diagnosis is largely due to a shortage of expertise, being the shortage of equipment the secondary factor. This lack of expertise for malaria diagnosis frequently results on the increase of false positives, since prescription of medication is based only on symptoms. Thus, there is an urgent need of new tools that can facilitate the rapid and easy diagnosis of malaria, especially in areas with limited access to quality healthcare services. *Methods:* Various image processing and analysis approaches already proposed on the literature for the detection and segmentation of malaria parasites in blood smear microscopic images were collected and reviewed. This timely review aims to support the increasing interest in the development of low cost tools that can facilitate the rapid and easy diagnosis of malaria, especially in areas with limited access to quality healthcare services. *Results:* Malaria parasites detection and segmentation techniques in microscopic images are, in general, still in need of improvement and further testing. Most of the methodologies reviewed in this work were tested with a limited number of images, and more studies with significantly larger datasets for the evaluation of the proposed approaches are needed. Despite promising results reported during the past years, the great majority of the computer-aided methods found on the literature for malaria diagnosis are based on images acquired under well controlled conditions and with proper microscopic equipment. However, one should take into account that 80% of malaria cases occur in Africa, where this type of equipment is scarce or even nonexistent in common healthcare facilities. *Conclusion:* This work collects and reviews various image processing and analysis approaches already proposed on the literature for the detection and segmentation of malaria parasites in blood smear microscopic images. This timely review aims to support the increasing interest in the development of image processing-based systems to be used in rural areas of developing countries, which might be the next future trend in malaria computer-aided diagnosis.

Keywords: Malaria, computer-aided diagnosis, image analysis, segmentation, feature extraction, classification.

1. INTRODUCTION

Malaria is one of the most severe public health problems worldwide. It is a leading cause of death and disease in many developing countries, where young children and pregnant women are the groups most affected. In 2012, there were an estimated 207 million cases of malaria, which caused approximately 627 000 deaths. An estimated 3.4 billion people continue to be at risk of malaria, mostly in Africa and south-east Asia.

Around 80% of malaria cases occur in Africa [1]. It is worth taking into account that the number of malaria cases and their geographical distribution are not stable because of several factors, like the increasing prevalence in some areas due to expanding drug resistance; the widespread availability of fake and substandard medicines; global warming and

expansion of malaria into favorable areas at higher elevations; and population mobility of different kinds [2].

The increasing interest in the development of computer-aided diagnosis (CAD) systems for malaria diagnosis is closely related with common practical difficulties experienced in under-resourced health facilities of developing countries, such as the excessive workload due to shortage of staff. Image processing approaches are often used in CAD systems to reduce the dependence of manual microscopic examination of blood smears, which is an exhaustive and time consuming activity, simultaneously requiring a considerable expertise of the laboratory technician.

During the last years, several image processing techniques have been proposed for malaria diagnosis using microscopic images, addressing the detection of a wide variety of different malaria parasites, in different growth stages and using images acquired from different types of blood smears. Under the scope of this paper, various image processing and analysis approaches already proposed on the literature for the

*Address correspondence to this author at the Research and Development Department of Fraunhofer Portugal AICOS, Porto, Portugal, Rua Alfredo Allen, 455/461, Porto, Portugal; Tel: +351 22 0430 336; E-mail: luis.rosado@fraunhofer.pt

detection and segmentation of malaria parasites in blood smear microscopic images were collected and reviewed. This timely review aims to support the increasing interest in the development of low cost tools that can facilitate the rapid and easy diagnosis of malaria, especially in areas with limited access to quality healthcare services.

This document is structured into five sections. Section 1 corresponds to Introduction and presents the motivation and objectives of this bibliographic survey. Section 2 gives an overview of the malaria disease in terms of parasite stages, species and life cycle stages; Section 3 provides an overview about the current malaria diagnosis methodologies, with a focus on the characterization of stained components in thin and thick blood smears; Section 4 gives a literature review regarding the analysis of malaria infected blood smears using image processing and analysis; Section 5 summarizes and gives a critical appreciation of the review works; Section 6 provides the final remarks about the discussed subjects.

2. MALARIA DISEASE CHARACTERIZATION

Malaria is caused by a parasite in the blood and can be seen only under a microscope with high magnification. For the visualization of the parasites, a blood film must be made, dried, stained and examined under the microscope. When the microscopist sees stained parasites, the diagnosis of malaria is confirmed by identifying the stage and species of the malaria parasite, as well as the infection density [2].

2.1. Malaria Parasites Stages

In the human host, malaria parasites pass through 3 different growth stages that can be detected in the peripheral blood: the trophozoite stage, the schizont stage and the gametocyte stage. Trophozoites are often called the ring stage, being the most commonly seen stage, appear incomplete in thick films, and can vary from small to quite large within the host cell. Usually, trophozoites have one chromatin dot, but two are common for the *P.falciparum* species. The cytoplasm takes different shapes, from a well-defined, fine ring to forms that are irregular or bizarre, sometimes called 'amoeboid' [1]. The schizont stage begins when the trophozoite has reached its full capacity and the parasite starts to divide into daughter cells called merozoites. Several more divisions of the chromatin follow, which mark the growth of the schizont, until there are many chromatin bodies, each with its accompanying cytoplasm. The number of chromatin and merozoite divisions helps to identify the species. These clearly delineated new parasites are now ready to leave the host cell to invade new red blood cells [2]. Gametocytes are round or banana-shaped, depending on the species. The way in which the parasite takes up the stain helps to identify the sex of the parasite in thin films, being difficult to differentiate between male and female in thick films [3].

2.2. Malaria Parasites Species

Four species of Plasmodium can infect and be transmitted by humans: the *P.falciparum*, *P.vivax*, *P.ovale* and *P.malariae*. *P.falciparum* is the commonest species in the tropical parts of the world and can evolve rapidly to severe illness and death if not recognized and treated with effective

medicines [2]. It is the species responsible for most cases of severe malaria and death. *P.vivax* is the commonest species in the cooler parts of the tropics, being the largest of the human malaria parasites and the cause of much illness and absenteeism from work and school [2]. *P.ovale* is considered a rare species, but relatively common in West Africa and other parts of the African continent. Because of morphological similarities, *P.ovale* is sometimes mistaken for *P.vivax* by less experienced microscopists [2]. Moreover, the existence of a new genotype for *P.ovale* has been recently hypothesized [4]. *P.malariae* is found worldwide and causes a chronic infection that in some cases can last a lifetime. In some chronically infected patients, *P.malariae* can cause serious complications such as the nephrotic syndrome [5]. As a final note, *P.knowlesi* is a malaria parasite that is found in nature in macaques, and naturally acquired human infections were thought to be extremely rare, however a large focus of human infections was reported in 2004 [6].

3. MALARIA DIAGNOSIS CHARACTERIZATION

Malaria infection can be suspected based on the patient's symptoms, travel history or physical findings at examination. However, for a definitive diagnosis, laboratory tests must be made to prove the presence of the malaria parasites. The microscopy examination remains the gold standard for laboratory confirmation of malaria, which consists in preparing a blood smear, staining it (most often with the Giemsa stain) and examining it through a microscope [5]. The importance of reliable malaria diagnoses cannot be overstated, since false negatives can be potentially fatal, and false positives increase the drug resistance of the patients, leading consequently to unnecessary economic burden [7]. Laboratory diagnosis of malaria can be made through microscopic examination of two kinds of blood smears, thin and thick, taken most often from a finger prick. Thick blood smears are 20-40 times more sensitive in detecting malaria parasites because the blood is more concentrated, which allows for a greater volume of blood to be examined. The thick smear is approximately 6-20 times as thick as a single layer of red blood cells, which results in a larger volume of blood being examined. However, thick smears are more difficult to read, so thin smears aid in parasite species identification and quantification [5], [7].

3.1. Malaria Parasites Stages

The images used on the reviewed works can be divided in two different groups according to their characteristics: full view images (FV) and manually cropped sub-images (CS). FV consists on images corresponding to the entire microscopic field of view (see Figs. 1 and 2). CS consists on cropped patches of the FV images, corresponding to regions of interest manually cropped (see Figs. 3 and 4).

Moreover, the vast majority of the proposed approaches found on literature use high quality equipment in the acquisition process, particularly commercial cameras that are specifically customized for the acquisition of microscopic images, for instance easily attached to microscopes. The exception is [7], which uses a smartphone built-in camera to acquire images (see Fig. 5).

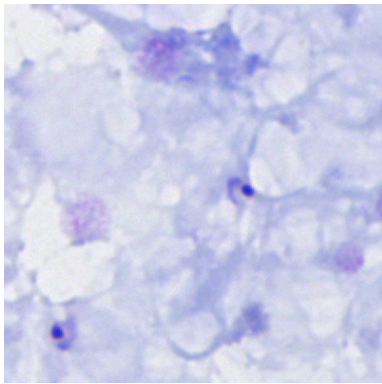


Fig. (1). Example of full view image on thick smear (from [8]).

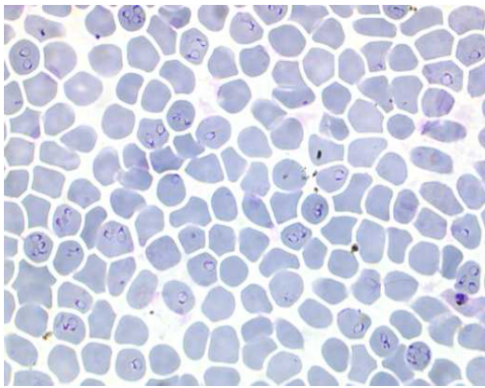


Fig. (2). Full view image on thin smear (from [9]).

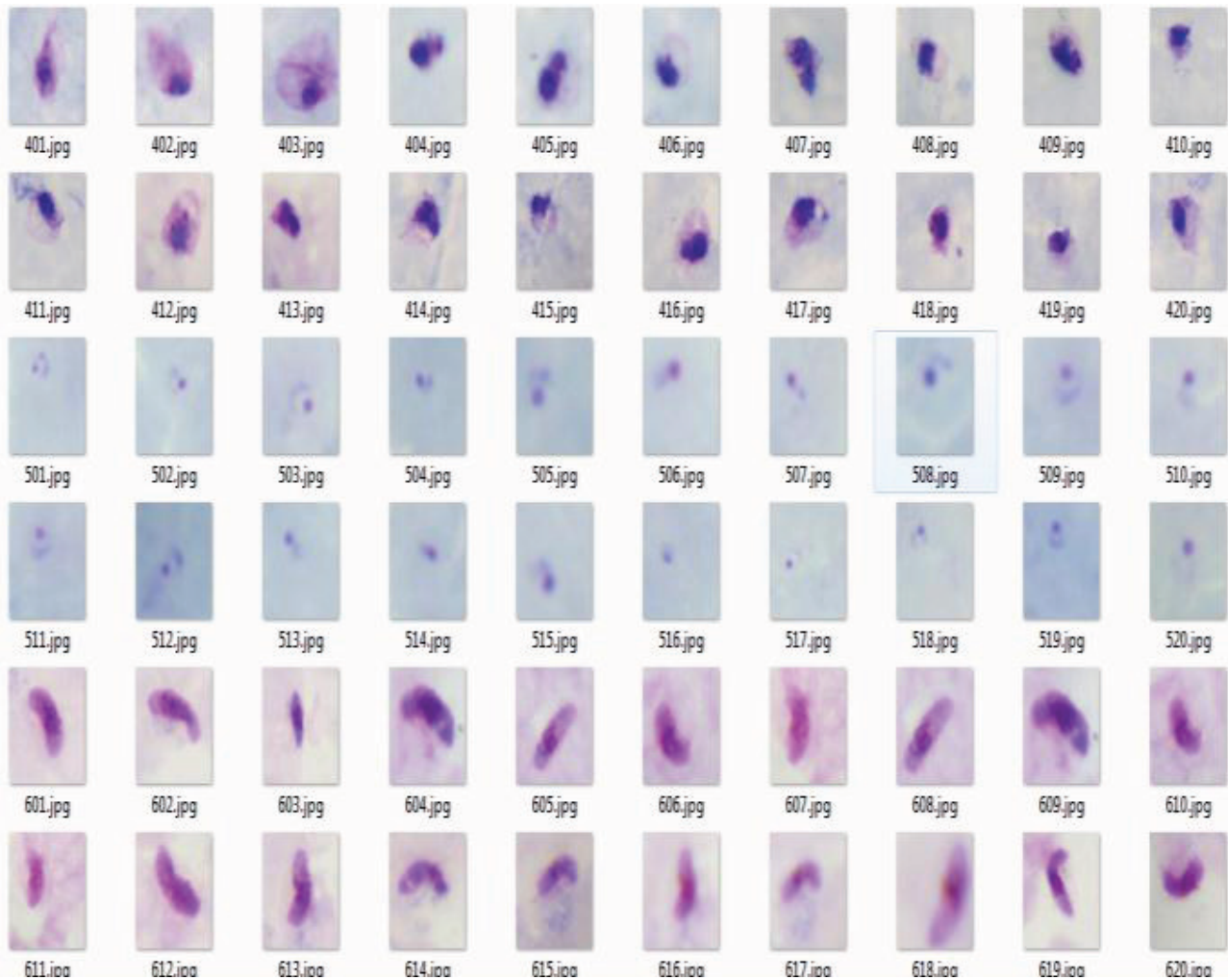


Fig. (3). Cropped images on thick smear (from [10]).

3.2. Performance Metrics

The classification results of the reviewed works are usually presented in terms of two metrics ordinarily used for this purpose: 1) Sensitivity (SE), *i.e.* the percentage of structures correctly classified as positive cases of malaria parasites; and 2) Specificity (SP), *i.e.* the percentage of structures correctly classified as negative cases of malaria parasites.

4. IMAGE PROCESSING TECHNIQUES

This section critically reviews the main studies found in the literature regarding the analysis of malaria infected blood smears using image processing and analysis. Since typical approaches usually comprise four different image processing and analysis tasks, the reviewed works were divided into the following sub-sections: 1) Segmentation; 2) Feature Extrac-

tion; 3) Feature Selection; and 4) Classification. For each sub-section, the methods proposed to date for malaria parasites (MP) stained components analysis, both on thin and thick blood smears, were separately reviewed.

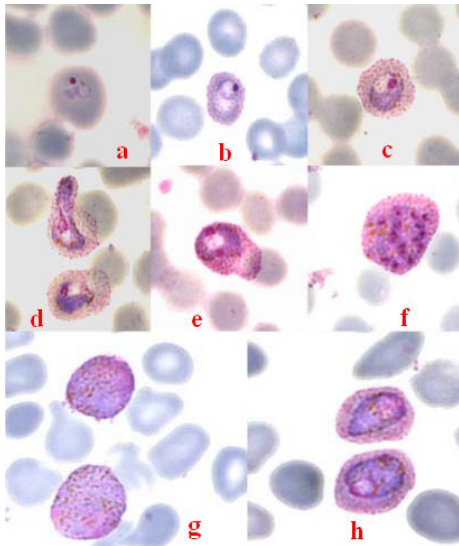


Fig. (4). Cropped sub-images on thin smear (from [11]).

4.1. Segmentation

Image segmentation is the process that partitions a digital image into disjoint (non-overlapping) regions, each of which typically corresponds to one object. Once isolated, these objects can be measured and classified, as discussed in the following sub-sections. This sub-section groups and reviews the methods proposed on the literature for the segmentation of malaria-stained components on thin and thick blood smears.

4.1.1. Thresholding

Thresholding is an essential region-based image segmentation technique that is particularly useful for scenes containing solid objects resting on a contrasting background. All pixels at or above/below the threshold are assigned to the foreground and all pixels below/above the threshold are assigned to the background [12].

4.1.1.1. Thin Blood Films

In [13], the authors suggest a scheme based on HSV color space that segments red blood cells (RBC) and identifies the RBC infected with MP. The RBC segmentation is achieved by dividing the entire hue range of 360° into six segments and finding the dominant color type to be the representative of the background. Each segment is centered on a color type, which is defined according to the follow degree

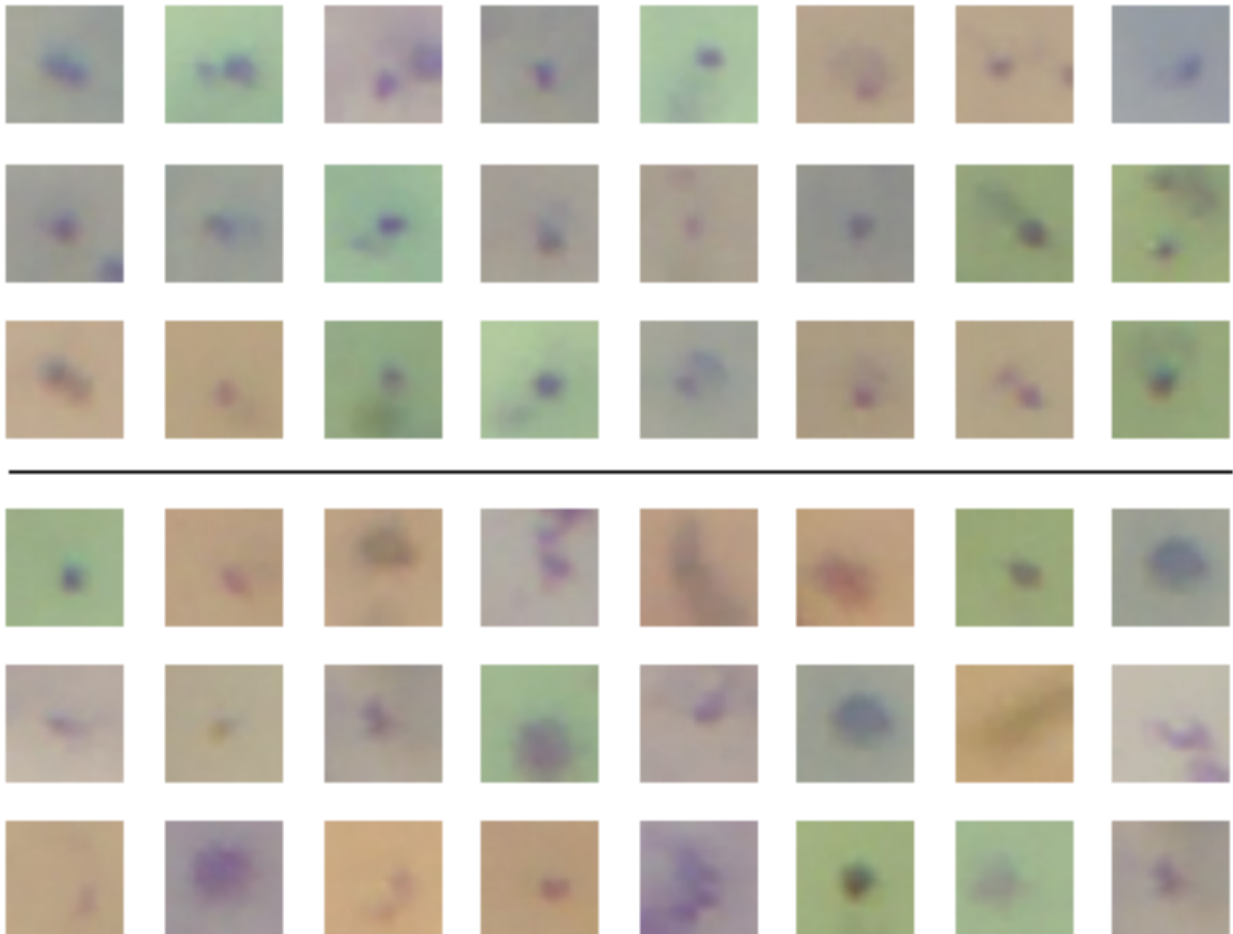


Fig. (5). Mobile acquired image patches on thin smear, with positive cases at the top and negative cases at the bottom (source: [7]).

values: 0° for red; 60° for yellow; 120° for green; 180° for cyan; 240° for blue; and 300° for magenta. The authors state a SE and SP of 83% and 98%, respectively, however this method uses images taken from Leishman-stained blood smears, while the gold standard recommended by WHO for malaria diagnosis is the usage of Giemsa staining. Furthermore, they assume that the dominant color in these images is representative of the background, which might not be true in images highly populated of RBC (see Fig. 3).

Another segmentation approach proposed in [3] for RBC on thin blood smears is based on the Annular Ring Ratio (ARR) transform. This transform consists in obtaining a ratio transformed image by calculating the ratio between the average intensities of a dilated image using annular concentric ring structuring element and an eroded image using a disk shaped structuring element. The ARR transform method mainly aims at locating the center of each cell present in the image. It is worth taking into account that this segmentation methodology uses fixed parameter values defined manually, such as the radius of the annular ring and disk, which can substantially vary depending on the image resolution.

The authors in [14] used a modified ARR transform method for the detection of stained cells through the direct application of the ratio transform on grayscale images, eliminating the morphological dilation and erosion. A threshold-based peak detection algorithm is then used to determine the coordinates of each stained component. This approach results in locating all the stained components in the image, with the drawback of picking up artifacts and other noises present in the image.

For blood cells components segmentation in thin smears, [15] uses the Otsu's Method, a well-known histogram shape-based image thresholding routine. This method assumes that the input image has two classes of pixels, and calculates the threshold that minimizes the intra-class variance. A hole filling process is then applied, due to the biconcave nature of the RBC. Using the knowledge that MP cytoplasm appear lighter while MP nucleus appear darker than the cytoplasm of the RBC, the authors state that it is possible to confirm that it is actually an infected cell by dividing it into 3 different regions (MP nucleous, MP cytoplasm and RBC cytoplasm) using multiple thresholding. However, how this multiple thresholding segmentation is achieved is not explained in the paper, and one should also take into account that possible artifacts could also be separated into 3 different regions.

In [16], *P.vivax* parasites from Leishman-stained thin blood film are segmented using divergence based threshold selection. The proposed modified fuzzy divergence method is based on Cauchy membership function, and applied to the C channel in the CMYK color space, which was the color channel that delivered the best results. However, it is not clear if this approach can be also applied to Giemsa stained thin blood films.

A methodology based on phase spectrum is used in [17] for malaria parasite detection in thin blood films. The method uses the Quaternion Fourier Transform (QFT) to obtain the amplitude spectrum and phase spectrum for blood smear images. Afterward, the reconstructed image is obtained using the Inverse Quaternion Fourier transform

(IQFT) on a constant amplitude spectrum and the original phase spectrum. The authors concluded that the most sensitive channels for MP detection were the B and G channels from the RGB space, as well as the I channel from the HSI space.

The proposed automated method in [18] for parasite detection and identification on thin blood film comprises two different segmentation steps: The foreground-background segmentation and the stained pixels segmentation. The proposed segmentation is performed using morphological area top-hats (using the average cell area value) and morphological double thresholding. For the stained pixels segmentation, the authors modeled the stained and unstained pixel distributions with RGB space histograms and used the probability density function to determine whether a pixel on the input image is stained or not. This work also comprises significant pre-processing effort, like the granulometry-based cell size estimation, which considers the peak index of the granulometry distribution as the average cell area. It also addresses problems like the non-uniform illumination, which is removed using a pre-recorded illumination image, or using a morphological closing operation with a structuring element with size of 5 times the average cell area. Despite this work being only focused on thin blood films, the authors highlighted the importance of future work to investigate the automatic analysis of thick films, since they are more sensitive for malaria parasites density estimation.

In [19], the authors combined the binary images from Otsu's method with the edge detection images from Canny's method for RBC segmentation, followed by a hole filling process. This work also proposes a pre-processing step based on the estimation of the image illumination, which is then subtracted from the original image.

4.1.1.2. Thick Blood Films

For non-background objects detection on thick blood films, *i.e.* MP, white blood cells (WBC) and possible Giemsa stain-derived artifacts, the authors in [20] used adaptive threshold found according to information of the V-Value histogram. However this methodology needs further validation since it was only tested on 20 microscopic images.

The G and B channels of the RGB color space were identified in [8] as very good features to identify objects containing chromatin in Giemsa stained blood films, being not only considered highly discriminative but also almost independent of differences in illumination and staining intensity. They transformed the color input image into a monochrome image $I(x,y)$, that highlights objects containing chromatin: $I(x,y) = \arctan(I_{\text{GREEN}}(x,y)/I_{\text{BLUE}}(x,y))$. The authors also used a black-top-hat morphological operator to separate MP from both leukocytes and platelets, with a non-flat paraboloid structuring element of radius of 9 and a slope of 1 pixel. It should be taken into account that these fixed parameters might not be suitable for images with different pixel resolutions. The black-top-hat operator is followed by a thresholding operation with a fixed threshold, which according to the authors is reliable given the independence of the G and B channels with regard to illumination and staining intensity. However, the authors do not define the value of this fixed threshold on the publication.

A dark stretching technique is applied in [21] for contrast enhancement of MP-infected thick blood smear images. The dark stretching is a process that uses an auto scaling method, being the dark areas stretched and the bright areas compressed. The authors state that the dark areas correspond to the MP in the infected images. To segment the MP, they use a single threshold value on the stretched image. However, the thresholds have been chosen empirically, and different thresholds had to be applied for different images in order to achieve good segmentation results.

4.1.2. Boundary-based Segmentation

Boundary-based techniques seek to extract object boundaries directly, based on identifying the edge pixels located at the boundaries in the image [12].

4.1.2.1. Thin Blood Films

An improved Circle Hough Transform (CHT) was used in [22] to detect RBC, based on the consideration that RBC have a circular pattern. The authors highlighted that CHT has the advantage of handling the segmentation for highly overlapping and oval cells, and noted that the RBC are much more noticeable in the green channel of the RGB color space. However, RBC might present shapes significantly different from the expected circular/oval pattern, which might difficult the detection using this approach.

4.1.2.2. Thick Blood Films

A combination of Absence of Gradients and Nernstian Equilibrium Stripping (AGNES), and Morphological Gradient techniques is used in [23] for the detection of *P.vivax* parasites in thick blood films. A morphological gradient method is first applied in order to enhance the borders of the objects present in the image. This is followed by a threshold detection stage using the K-Median method. The AGNES and K-Median techniques were then used to assign the remaining number of pixels to each region, using as starting points the image regions previously identified as objects and background. These techniques were locally applied on rectangular sectors corresponding to 20% of the input image, a percentage defined experimentally. According to the authors, this approach might induce the appearance of objects that are not part of the original image in the border of the defined rectangular sectors.

4.1.3. Clustering Methods

Clustering is a machine learning technique that can be used for segmentation purposes when applied as an indicator of the similarity of different image regions, based on a set of measurements that describes those regions.

4.1.3.1. Thin Blood Films

The application of 3 different clustering algorithms to segment blood cell images is studied in [24]: Mean-shift, K-means and Fuzzy C-means. The authors support that K-means clustering algorithm achieves best results for blood cell images segmentation, and they used a Median-cut algorithm after applying the K-means in order to reduce the number of regions to an optimum level. However, this ap-

proach needs further validation since it only used 5 cropped sub-images.

A segmentation approach using k-means clustering is also proposed in [25] for the detection of *P.vivax* parasites. Different color components of RGB, HSI and C-Y color models have been analyzed, and the S component of C-Y color model has proven to be the best, with segmentation accuracy and F-score of 99.46% and 0.9370, respectively.

In [26], the authors used the b*-color channel from the CIE L*a*b* color space for k-means clustering as an unsupervised segmentation approach to identify malaria parasite tissues. They used 118 Leishman-stained microscopic images with reported results of 76% and 60% for SE and SP, respectively. It is worth noting that these results refer to classifying the images as infected/not infected, thus not giving information about the precise number of correctly identified MP in each image.

4.1.4. Graph Partitioning Methods

Graph partitioning methods model the images as a weighted and undirected graph, which defines the (dis)similarity between the neighborhood pixels. It can be used for segmentation purposes when the graph is partitioned according to a specific criterion that creates reliable pixel clusters.

4.1.4.1. Thick Blood Films

The authors in [11] present an optimized normalized cut (NCut) algorithm for the segmentation of RBC infected with MP in thick blood smears. The NCut algorithm is based on a global criterion, and it maximizes both the total dissimilarity between the different groups and the total similarity within the groups. For detection of trophozoites and schizonts, the NCut algorithm performs best in the HSV color space. However, several artifacts appear on the segmented images, and the usage of a global criterion makes this methodology prone to error with significant variation of the illumination conditions and/or staining intensity.

4.1.5. Classifier-based Methods

Classifier-based methods for segmentation rely on the usage of two-class models for segmentation purposes.

4.1.5.1. Thin Blood Films

In [9], a Bayesian pixel classifier has been employed in order to differentiate between the stained and non-stained pixels. The class conditional probability density functions of the stained and the non-stained classes were estimated using the non-parametric histogram method.

4.2. Feature Extraction

The primary objectives of feature extraction are reducing the computational complexity of the subsequent process and facilitating a reliable and accurate recognition for unknown novel data, being the last objective particularly important for computer vision and pattern recognition systems. Moreover, the in-depth understanding of the domain-specific knowledge gained by human experts on the problem being addressed

can be of extreme importance for the design of a reliable and effective feature extraction engine [27].

4.2.1. Binary Objects Measures

A binary object can be described in terms of its size (e.g. area, perimeter), pose (e.g. centroid, orientation), shape measures (e.g. thinness ratio, rectangularity, circularity, Euler number, moments, elongation), shape descriptors (e.g. differential chain code, fourier descriptors, media axis transform, graph representation) or distance to other objects (e.g. Euclidean distance, city-block distance, chessboard distance) [12].

4.2.1.1. Thin Blood Films

In [13], the infected RBC detection in thin blood smears is based on relative ratios between chromatin dots and RBC terms of area and centroids distances. For the same purpose, the work in [18] uses shape measurements like compactness and moments of inertia.

4.2.1.2. Thick Blood Films

The segmented non-background objects are distinguished in [20] according to size, and the MP species *P. falciparum* and *P. vivax* are distinguish according to the chromatin size. However, this methodology was tested only on 20 images and the thresholds were defined in number of pixels, thus not suitable for images with different pixel resolutions.

4.2.2. Gray-Level Objects Measures

Gray-level objects measures consist on measurements derived from the intensity distribution of the object. There are 3 main categories of gray-level object measurements: Intensity measures (e.g. integrated and average optical intensity, contrast) histogram measures (e.g. mean, standard deviation, skew, entropy, energy) and texture measures (e.g. statistical texture measures, power spectrum features) [12].

4.2.2.1. Thin Blood Films

For detection and differentiation between WBC and *P. falciparum* gametocytes, the authors in [14] start by finding the centroids of each detected stained component. The proposed algorithm considers as candidates the neighborhood of the listed centroid coordinates within the diameter of the WBC. The candidates are marked as WBC if the neighborhood region presents low mean intensity and less variance, and as gametocyte of *P. falciparum* if the region has high eccentricities (variance) and average intensity.

The authors in [22] observed that nucleated components result in distinctively high intensity values in the B channel of the RGB color space, while the same nucleated components in the G channel exhibit very low intensity values. Therefore, they proposed an intensity measure based on the differences between the B and the G color intensity channels, in order to stretch the contrast for visual perception and emphasize nucleated objects.

The methodology proposed by [15] aims to detect 3 regions inside the infected RBC, particularly the MP nucleous, MP cytoplasm and RBC cytoplasm. Features of those re-

gions are then extracted for classification purposes, based on area ratios and range of intensities of each region.

In another work [9], the following gray-level measures were extracted for further MP/non-MP classification in thin blood films: Hu moments, relative shape measurements, histogram and color auto-correlogram. According to the authors, the last two features are the most effective for MP classification.

The color histogram is used in [18] after quantization into 32 colors, as well as the local area granulometry for each RGB channel. It worth noting that the features were normalized (for zero mean, 1 variance) first on the training set, being the mean and standard deviation calculated and then used for the normalization of the testing set.

The authors in [19] extracted four different features: 1) Gradient; 2) Flat texture (determined by computing the difference between the original image and the filtered image using median filter); 3) Color histogram; and 4) Area granulometry.

4.2.2.2. Thick Blood Films

In order to differentiate MP from artifacts in thick blood films, the authors in [8] extracted 174 different features. The features were grouped in 3 main groups: 1) Statistical moment features - 4 central moments (mean, variance, skewness, and kurtosis), Hu's set of 7 invariant moments [28], 49 Zernike moments of orders up to 12 [29]; 2) Texture features: Haralick's 13 co-occurrence matrix features [30]; Unser's 18 sum and difference histogram features [31]; Chen's 16 statistical geometrical features [32]; 5 features proposed by Young *et al.* [33] that describe the distribution of chromatin in the ROI; 3) Color features: 60 features representing a 60-bin histogram of the hue channel of the ROI; 2 features described by Kovalev *et al.* [34] that represent cyan shifts in the ROI.

The following features were extracted in the work presented in [10]: mean; standard deviation; kurtosis; skewness; entropy of the histograms of R,G and B channels from RGB space, H channel from HSV space, and H channel of HSI space.

In another recent work, 3 groups of features were extracted in [23] for each RGB channel: 1) Color features: standard deviation, 7 Hough moments of color and color range; 2) Texture features based on the co-occurrence matrix (in 4 directions): homogeneity, contrast, GLMSR (General Linear Model Simple Regression), standard deviation, second angular moment and correlation; 3) Texture features based on the Wavelet transform: energy, mean and standard deviation. Four wavelet families were used (Haar; biorthogonal 1,3; Daubechies 2; and Daubechies 8) considering 2 decomposition levels, being applied in each level a low and high pass filter.

In [7], the authors extract two types of features: 1) Connected component features: Perimeter; Moment of Inertia; Elongation; Jaggedness; and Maximum λ (consists in the maximum child gray level minus the current gray level); 2) Moment features: the moment m_{00} , the central moments, u_{11} , u_{20} , u_{02} and Hu moments h_0 , h_1 , h_2 .

4.3. Feature Selection

In order to build a good classification model, the reduction of the number of attributes used on the classification process may not only have positive impact in terms of the processing time, but also in terms of the classification results. Feature selection techniques play an important role in this context, and they can be organized into three categories: filter methods, wrapper methods and embedded methods [35].

Filter Methods rank each feature according to some univariate searching function and select the highest-ranking features, where the scoring should reflect the discriminative power of each feature. Some of the most common univariate filter methods includes Bayesian Network, Information Gain, Signal-to-Ratio, Euclidean Distance or Correlation Squares (R^2) [35]. Filter methods are usually very efficient and fast to compute, but comprise some significant drawbacks like the redundancy of the selected features, which can carry the same information. Another important disadvantage of the filter methods is the fact that this selection does not consider some important relationships between features, since features can receive a low score by the ranker algorithm when used by itself, but be very useful when combined with other features.

Opposite to filter techniques that consider the ranking of each feature independently, Wrapper and Embedded Methods are specific to a given machine-learning algorithm. In the Wrapper Method, a search is conducted using a specific classifier in order to find the subset of features with which the classification algorithm performs the best. For instance, using forward selection, the Wrapper Method estimates the accuracy of adding each unselected feature to the feature subset, and the feature that most improves the accuracy is selected. These methods typically terminate when the estimated accuracy of adding any feature is less than the estimated accuracy of the feature set already selected [36].

4.3.1. Thick Blood Films

An optimal feature subset was chosen in [8], based on the 174 previously extracted features for MP detection. The features were first normalized to have zero mean and a standard deviation of one. The authors proposed a 2-step feature selection methodology: an univariate ranking is first applied in order to keep only the 60 features that have the highest univariate discriminative power, and then a genetic algorithm (GA) is used for automatic selection of an even smaller feature subset.

Another example of feature selection for MP detection is referred in [23], where the Principal Component Analysis (PCA) was used to identify features that are redundant and do not carry new information, thus also involving feature transformation.

4.4. Classification

In machine learning and statistics, classification is the problem of identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known. In the terminology of machine learning, classification is considered an instance of

supervised learning, *i.e.* learning where a training set of correctly identified observations is available [37].

4.4.1. Naïve Bayes Classifier

The Naïve Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature, meaning that it assumes the independency of variables given the class. In spite of its naive design and apparently over-simplified assumptions, Naïve Bayes can often outperform more sophisticated classification methods applied in many complex real word situations. An advantage of the Naïve Bayes classifier is that it requires a small amount of training data to estimate the parameters of the model [38].

4.4.1.1. Thin Blood Films

In [15], the classification model for the identification of infected RBC achieved a SE of 92.59% and a SP of 99.65% using a Bayes Decision Rule classifier. However, only 60 images were used for training and 20 for testing.

4.4.2. K-Nearest Neighbor

One of the most fundamental and simple classification methods is K-nearest neighbor (kNN), being often used when there is little or no prior knowledge about the distribution of the data. Nearest neighbor algorithm is based on the principle that the instances within a dataset will generally exist in close proximity to other instances with similar properties [39].

4.4.2.1. Thin Blood Films

For detection of MP in thin films, in [9] the authors proposed a distance weighted kNN classifier trained with the extracted features, and a detailed performance comparison is presented. The authors state that the proposed method achieved 74% of SE and 98% of SP.

The authors in [18] proposed an automated method for MP detection and identification on thin blood film, where three classifiers were tested: the kNN classifier achieved 72.4% of SE and 97.6% of SP; the Fisher Linear Discriminant achieved 71.5% of SE and 93.9% of SP; and the Back Propagation Neural Network achieved 70.5±1.7% of SE and 96.6±0.4% of SP. Furthermore, the authors proposed and compared three different classification models for species and life-cycle-stage identification: The first model (20-class) performs detection, species, and life-cycle-stage recognition in a single classification, simultaneously considering non-parasite classes. The second and third models perform a binary detection beforehand followed by a single 16-class classification or two 4-class classifications for identification, respectively. The 20-class model was considered the most favorable model.

In another recent work, the authors in [19] tested five different classifiers, and stated that the kNN outperformed others with 80% of SE and 95.5% of SP.

4.4.3. Support Vector Machine

A support vector machine (SVM) constructs a hyperplane or set of hyperplanes in a high-or infinite-dimensional space,

which can be used for classification. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier [40].

4.4.3.1. Thick Blood Films

For MP classification, the authors in [8] used a SVM with a radial basis function kernel with gamma $G = 0.125$, and a cost factor of $C = 1.0$. The stated results showed a SE of 97% combined with 0.8 false-positive detections per image.

4.4.4. Neural Networks

Neural Networks (NN) are based on the way biological nervous systems process information and are applied to a large number of real world problems. They are especially well suited to problems that people are good at solving, but the algorithm solution is too complex to be defined. The most common NN model is the multilayer perceptron, which is a feed-forward artificial NN that maps sets of input data onto a set of appropriate output, using three or more layers of neurons with nonlinear activation functions [41].

4.4.4.1. Thick Blood Films

A NN with 5 neurons in the output layer was used in [23], of which three neurons represent the parasites of interest (gametocytes of *P.falciparum*, squizonts of *P.vivax* and

gametocytes of *P.vivax*), and the two remaining neurons correspond to sediments and WBC.

4.4.5. Genetic Programming

Genetic programming is an evolutionary algorithm-based methodology inspired by biological evolution to find computer programs that perform a user-defined task. Essentially genetic programming is a set of instructions and a fitness function to measure how well a computer has performed a task. It is a specialization of genetic algorithms where each individual is a computer program, being a machine learning technique used to optimize a population of computer programs according to a fitness landscape determined by a program's ability to perform a given computational task [42].

4.4.5.1. Thick Blood Films

In [10], the classification was made using Genetic Programming to identify parasites and also to detect type and phase of the parasite. Two different classification models were considered using a dataset of 180 thick blood film sub-images cropped manually: a first model (two classes) with an accuracy of 95.49% for non-parasites and 95.58% for parasites, and a second model (six classes) with an average accuracy of 90.25% for non-parasites, 82.25% for *P.vivax* trophozoites, 75.83% for *P.vivax* schizonts, 81.75% for *P.vivax* gametocytes, 90.75% for *P.falciparum* trophozoites and 86.75% for *P.falciparum* gametocytes.

Table 1. Proposed approaches for the detection and/or segmentation of malaria parasites in thick blood films.

Author	Year	Segmentation	Features	Classifier	#Images	SE/SP (%)
(Kaewkamnerd <i>et al.</i> , 2011)	2011	Adaptive Thresholding	- Hue histogram. - Chromatin size.	-	20 (FV)	- / -
(Elter <i>et al.</i> , 2011)	2011	Thresholding	- 60 statistical moment features. - 52 texture features. - 62 color features.	SVM	256 (FV)	97 / -
(Yunda <i>et al.</i> , 2012)	2012	AGNES + Morphological gradient techniques	- 27 color features. - 72 co-occurrence matrix texture features. - 62 Wavelet transform texture features.	Neural Network	248 (FV)	- / -
(Purnama <i>et al.</i> , 2013)	2013	-	- Mean and standard deviation. - Kurtosis. - Skewness. - Entropy of the histograms of: (a) R, G and B channels from RGB space; (b) H channel from HSV space; (c) H channel of HIS space.	Genetic Programming	180 (CS)	96 / 96
(Quinn <i>et al.</i> , 2014)	2014	-	- Connected component features (Perimeter; Moment of Inertia; Elongation; Jaggedness; and Maximum λ). - Moment features (moment m_{00} , the central moments, u_{11} , u_{20} , u_{02} and Hu moments h_0 , h_1 , h_2).	Extremely Randomized Trees	2703 (M)	*ROC AUC = 0.97

#Images indicates the number of images used to validate the proposed approach; SE and SP indicates the reported sensitivity and specificity of the proposed approach, respectively; AGNES stands for "Absence of gradients and Nernstian equilibrium stripping"; FV stands for full view images; CS stands for manually cropped sub-images; M stands for mobile acquired images.

4.4.6. Extremely Randomized Trees

Extremely Randomized Trees is a tree-based ensemble method for classification that consists of randomizing strongly both attribute and cut-point choice while splitting a tree node. In the extreme case, it builds totally randomized trees whose structures are independent of the output values of the learning sample. The strength of the randomization can be tuned to problem specifics by the appropriate choice of a parameter [43].

4.4.6.1. Thick Blood Films

The authors in [7] used a Extremely Randomized Trees classifier with an ensemble of 250 trees and a maximum tree depth of 5. The authors stated that this classifier has the ad-

vantages of being fast and memory-efficient, which might be useful in situations where classification must be carried out for instance on a mobile device with limited computational resources. The authors stated an area under the curve (AUC) of 0.97 for the receiver operating characteristic (ROC) curve.

5. SUMMATION AND CRITICAL APPRECIATION

The proposed methodologies were divided in two main groups: MP in thick blood films (see Table 1) and MP in thin blood films (see Table 2).

Malaria parasites detection and segmentation techniques in microscopic images are, in general, still in need of improvement and further testing. Most of the methodologies reviewed in this work were tested with a limited number of

Table 2. Proposed approaches for the detection and/or segmentation of malaria parasites in thin blood films.

Author	Year	Segmentation	Features	Classifier	#Images	SE/SP (%)
(Tek <i>et al.</i> , 2006)	2006	Bayesian pixel classifier	- Histogram. - Hu moments. - Relative shape measurements. - Color auto-correlogram.	kNN	260 (FV)	74 / 98
(Makkapati and Rao, 2009)	2009	Thresholding	- Area ratios. - Centroids distances between RBC measurements and chromatine dots.	-	- (FV)	83 / 98
(Zou <i>et al.</i> , 2010)	2010	Circle Hough transform	-	-	(FV)	-
(Mandal <i>et al.</i> , 2010)	2010	Normalized cut	-	-	37 (CS)	-
(Tek <i>et al.</i> , 2010)	2010	Thresholding	- Color histogram. - Local granulometry. - Shape measurements (e.g. compactness, moments of inertia).	kNN	630 (FV)	72 / 98
(Kareem <i>et al.</i> , 2011)	2011	Thresholding	-	-	- (FV)	-
(Muda and Salam, 2011)	2011	K-means clustering + Median-cut	-	-	5 (CS)	-
(Anggraini <i>et al.</i> , 2011)	2011	Otsu's method	- Range of intensity of infected and normal RBC. - Area ratios.	Bayes Classifier	80 (FV)	93 / 99
(Ghosh <i>et al.</i> , 2011)	2011	Fuzzy divergence	-	-	150 (CS)	-
(Fang <i>et al.</i> , 2011)	2011	Thresholding	-	-	100 (FV)	87 / 97
(Kareem <i>et al.</i> , 2012)	2012	Thresholding	-	-	200 (FV)	-
(Malihi <i>et al.</i> , 2013)	2013	Otsu's method + Canny's method	- Gradient. - Flat texture. - Color Histogram. - Area Granulometry.	kNN	363 (FV)	80 / 96
(Abdul <i>et al.</i> , 2013)	2013	K-means clustering	-	-	100 (CS)	94 / 96
(Khan <i>et al.</i> , 2014)	2014	K-means clustering	- b*-color channel from the CIE L*a*b* color space.	-	118 (FV)	76 / 60

#Images indicates the number of images used to validate the proposed approach; SE and SP indicates the reported sensitivity and specificity of the proposed approach, respectively; kNN stands for the k-Nearest Neighbor; FV stands for full view images; CS stands for manually cropped sub-images.

images, and more studies with significantly larger datasets for the evaluation of the proposed approaches are needed. Despite promising results reported during the past years, the great majority of the computer-aided methods found on the literature for malaria diagnosis are based on images acquired under well controlled conditions and with proper microscopic equipment. However, one should take into account that 80% of malaria cases occur in Africa, where this type of equipment is scarce or even nonexistent in common healthcare facilities.

Furthermore, the analysis of thin blood smears is much more addressed in the literature when compared with thick blood smears. This is probably because the image processing tasks required on thick blood smears analysis are considerably more challenging, given the fact that thin blood smears consist on a single layer of blood elements. Thus, several additional image artifacts are avoided in thin blood smear analysis, like the overlap of blood elements or the recurrent appearance of unfocused structures caused by the location of those structures in different focal planes. Thick blood smears are considered 20-40 times more sensitive in detecting malaria parasites, and therefore should be equally addressed, despite the additional image processing challenges.

CONCLUSION

This work collects and reviews various image processing and analysis approaches already proposed on the literature for the detection and segmentation of malaria parasites in blood smear microscopic images. This timely review aims to support the increasing interest in the development of image processing-based systems to be used in rural areas of developing countries, which might be the next future trend in malaria computer-aided diagnosis.

The development of new mobility-aware microscopic devices (and ideally low cost) is an area that can greatly improve the chances of the successful deployment of computer vision CAD solutions for malaria diagnosis in the field. Taking into account the high customs taxes and import duties currently in practice in most of the African countries, the easy replicability of these microscopy devices in third world countries should also be an issue to address. Several others additional requirements for this type of microscopic devices can be equally considered, like automating the device as much as possible, discarding the need of considerable expertise and train of the technician in terms of maneuvering the microscope, or supplying the energy needed for the illumination and/or any type of automation through the mobile device battery, thus discarding the need of an additional power source.

The mobile phone is currently Africa's most important digital technology, and is boosting African health as it emerges as a platform for diagnosis and treatment. In 2000, few Africans had a phone, but today about three-quarters do [44]. Just as African telecommunications largely skipped over landline infrastructure and went straight to mobile phones, some experts say African medicine can skip over centralized labs [45]. Considering the recent significant improvements of the new generation of mobile devices in terms of image acquisition and processing power, if a reliable automatic diagnostic performance is ensured through the

usage of those devices, one would dramatically reduce the effort in the exhaustive and time consuming activity of microscopic examination.

Moreover, the lack of highly trained microscopists on malaria diagnosis in rural areas could then be complemented by a significantly less specialized technician that knows how to operate the system and prepare blood smears. The usage of mobile devices in the system architecture can also bring significant improvements in terms of portability and data transmission. Finally, malaria diagnosis might be just one element of a suite of diagnostic software tests running on this type of system. Several other tests could simultaneously be carried out using the same images, for instance cell counting or detection of other hemoparasites like microfilaria or trypanosoma.

CONFLICT OF INTEREST

The author(s) confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

We would like to acknowledge the financial support obtained from North Portugal Regional Operational Programme (ON.2 - O Novo Norte), Portuguese National Strategic Reference Framework (NSRF) and the European Regional Development Fund (ERDF) from European Union through the project ICT4DCC (ICT4D Competence Center) NORTE-07-0124-FEDER-000042.

REFERENCES

- [1] World Health Organization. WHO | World Malaria Report 2013 http://www.who.int/malaria/publications/world_malaria_report_2013/report/en/ (accessed Mar 12, 2014).
- [2] World Health Organization. WHO | Basic malaria microscopy: Part I. Learner's guide. Second edition <http://www.who.int/malaria/publications/atoz/9241547820/en/> (accessed Apr 8, 2014).
- [3] Kareem, S.; Morling, R.C.S.; Kale, I. A Novel Method to Count the Red Blood Cells in Thin Blood Films. In *2011 IEEE International Symposium on Circuits and Systems (ISCAS)*; 2011; pp 1021-1024.
- [4] Sutherland, C.J.; Tanomsing, N.; Nolder, D.; Oguike, M.; Jennison, C.; Pukrittayakamee, S.; Dolecek, C.; Hien, T.T.; Rosário, V.E. do; Arez, A.P. Pinto, J.; Michon, P.; Escalante, A.A.; Nosten, F.; Burke, M.; Lee, R.; Blaze, M.; Otto, T.D.; Barnwell, J.W.; Pain, A.; Williams, J.; White, N.J.; Day, N.P.; Snounou, G.; Lockhart, P.J.; Chiodini, P.L.; Imwong, M.; Polley, S.D. Two Nonrecombining Sympatric Forms of the Human Malaria Parasite Plasmodium Ovale Occur Globally. *J. Infect. Dis.*, **2010**, *201* (10), 1544-1550.
- [5] Centers for Disease Control and Prevention <http://www.cdc.gov/> (accessed Apr 13, 2014).
- [6] Singh, B.; Daneshvar, C. Human Infections and Detection of Plasmodium Knowlesi. *Clin. Microbiol. Rev.*, **2013**, *26* (2), 165-184.
- [7] Quinn, J.; Andama, A.; Munabi, I.; Kiwanuka, F. Automated Blood Smear Analysis for Mobile Malaria Diagnosis. In *Mobile Point-of-Care Monitors and Diagnostic Device Design*; Devices, Circuits, and Systems, **2014**.
- [8] Elter, M.; Hasslmeyer, E.; Zeffass, T. Detection of Malaria Parasites in Thick Blood Films. *Annu. Int. Conf., IEEE Eng. Med. Biol. Soc.* **2011**, *2011*, 5140-5144.
- [9] Tek, F.B.; Dempster, A.G.; Kale, I. Malaria Parasite Detection in Peripheral Blood Images, **2006**.
- [10] Purnama, I.K.E.; Rahmanti, F.Z.; Purnomo, M.H. Malaria Parasite Identification on Thick Blood Film Using Genetic Programming. In *2013 3rd International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME)*, **2013**, pp 194-198.

- [11] Mandal, S.; Kumar, A.; Chatterjee, J.; Manjunatha, M.; Ray, A.K. Segmentation of Blood Smear Images Using Normalized Cuts for Detection of Malarial Parasites. In *2010 Annual IEEE India Conference (INDICON)*, **2010**, pp 1-4.
- [12] Wu, Q.; Merchant, F.A.; Castleman, K.R. *Microscope Image Processing*; Elsevier/Academic Press: Amsterdam; Boston, 2008.
- [13] Makkapati, V.V.; Rao, R.M. Segmentation of Malaria Parasites in Peripheral Blood Smear Images. In *IEEE International Conference on Acoustics, Speech and Signal Processing, 2009. ICASSP 2009*, **2009**, pp 1361-1364.
- [14] Kareem, S.; Kale, I.; Morling, R. C. S. Automated P.falciparum Detection System for Post-Treatment Malaria Diagnosis Using Modified Annular Ring Ratio Method. In *2012 UKSim 14th International Conference on Computer Modelling and Simulation (UK-Sim)*; **2012**; pp 432-436.
- [15] Anggraini, D.; Nugroho, A.S.; Pratama, C.; Rozi, I.E.; Pragesjvara, V.; Gunawan, M. Automated Status Identification of Microscopic Images Obtained from Malaria Thin Blood Smears Using Bayes Decision: A Study Case in Plasmodium Falciparum. In *2011 International Conference on Advanced Computer Science and Information System (ICACSIS)*, **2011**, pp 347-352.
- [16] Ghosh, M.; Das, D.; Chakraborty, C.; Ray, A.K. Plasmodium Vivax Segmentation Using Modified Fuzzy Divergence. In *2011 International Conference on Image Information Processing (ICIIP)*; **2011**; pp 1-5.
- [17] Fang, Y.; Xiong, W.; Lin, W.; Chen, Z. Unsupervised Malaria Parasite Detection Based on Phase Spectrum. *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Conf.*, **2011**, 2011, 7997-8000.
- [18] Tek, F.B.; Dempster, A.G.; Kale, I. Parasite Detection and Identification for Automated Thin Blood Film Malaria Diagnosis. *Comput. Vis. Image Underst.*, **2010**, 114 (1), 21-32.
- [19] Malihi, L.; Ansari-Asl, K.; Behbahani, A. Malaria Parasite Detection in Giemsa-Stained Blood Cell Images. In *2013 8th Iranian Conference on Machine Vision and Image Processing (MVIP)*, **2013**, pp 360-365.
- [20] Kaewkamnerd, S.; Intarapanich, A.; Pannarat, M.; Chaotheing, S.; Uthairibull, C.; Tongsim, S. Detection and Classification Device for Malaria Parasites in Thick-Blood Films. In *2011 IEEE 6th International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS)*, **2011**, Vol. 1, pp 435-438.
- [21] Hanif, N.S.M.M.; Mashor, M.Y.; Mohamed, Z. Image Enhancement and Segmentation Using Dark Stretching Technique for Plasmodium Falciparum for Thick Blood Smear. In *2011 IEEE 7th International Colloquium on Signal Processing and its Applications (CSPA)*; **2011**; pp 257-260.
- [22] Zou, L.; Chen, J.; Zhang, J.; Garcia, N. Malaria Cell Counting Diagnosis within Large Field of View. In *2010 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, **2010**, pp 172-177.
- [23] Yunda, L.; Alarcón, A.; Millán, J. Metodo Automatizado de Analisis de Imagenes Para Deteccion Del Parasito de La Malaria P-Vivax En Imagenes de Gota Gruesa. *Sist. Telemat.*, **2012**, 10 (20), 9-25.
- [24] Muda, T.Z.T.; Salam, R.A. Blood Cell Image Segmentation Using Hybrid K-Means and Median-Cut Algorithms. In *2011 IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, **2011**, pp 237-243.
- [25] Abdul, N.; Aimi, S.; Mashor, M. Y.; Mohamedq, Z. Colour Image Segmentation Approach for Detection of Malaria Parasites Using Various Colour Models and k-Means Clustering <http://connection.ebscohost.com/c/articles/93385472/colour-image-segmentation-approach-detection-malaria-parasites-using-various-colour-models-k-means-clustering> (accessed Mar 12, 2014).
- [26] Khan, N.A.; Pervaz, H.; Latif, A.K.; Musharraf, A.; Saniya. Unsupervised Identification of Malaria Parasites Using Computer Vision. In *2014 11th International Joint Conference on Computer Science and Software Engineering (JCSSE)*; **2014**; pp 263-267.
- [27] Jiang, X. Feature Extraction for Image Recognition and Computer Vision. In *2nd IEEE International Conference on Computer Science and Information Technology, 2009. ICCSIT 2009*, **2009**, pp 1-15.
- [28] Hu, M.-K. Visual Pattern Recognition by Moment Invariants. *IRE Trans. Inf. Theory* **1962**, 8 (2), 179-187.
- [29] Khotanzad, A.; Hong, Y.H. Invariant Image Recognition by Zernike Moments. *IEEE Trans. Pattern Anal. Mach. Intell.*, **1990**, 12 (5), 489-497.
- [30] Haralick, R.M.; Shanmugam, K.; Dinstein, I. Textural Features for Image Classification. *IEEE Trans. Syst. Man Cybern.*, **1973**, SMC-3 (6), 610-621.
- [31] Unser, M. Sum and Difference Histograms for Texture Classification. *IEEE Trans. Pattern Anal. Mach. Intell.*, **1986**, PAMI-8 (1), 118-125.
- [32] Chen, Y.Q.; Nixon, M.S.; Thomas, D.W. Statistical Geometric Features for Texture Classification. *Pattern Recognit.*, **1995**, 28 (4), 537.
- [33] Young, I.T.; Verbeek, P.W.; Mayall, B.H. Characterization of Chromatin Distribution in Cell Nuclei. *Cytometry*, **1986**, 7 (5), 467-474.
- [34] Kovalev, V.A.; Grigoriev, A.Y.; Ahn, H.-S. Robust Recognition of White Blood Cell Images. In *Proceedings of the 13th International Conference on Pattern Recognition, 1996*, **1996**, Vol. 4, pp 371-375 vol.4.
- [35] George, G.V.S.; Raj, V.C. Review on Feature Selection Techniques and the Impact of SVM for Cancer Classification Using Gene Expression Profile. *CoRR*, **2011**, abs/1109.1062.
- [36] Das, S. Filters, Wrappers and a Boosting-Based Hybrid for Feature Selection. In *Proceedings of the Eighteenth International Conference on Machine Learning*; ICML '01; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, **2001**; pp 74-81.
- [37] K.Jayavani. Statistical classification in machine intelligent. **2014**, 1 (1).
- [38] John, G. H.; Langley, P. Estimating Continuous Distributions in Bayesian Classifiers. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*; UAI'95; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, **1995**, pp 338-345.
- [39] Aha, D. W.; Kibler, D.; Albert, M. K. Instance-Based Learning Algorithms. *Mach Learn.*, **1991**, 6 (1), 37-66.
- [40] Support vector machine - Wikipedia, the free encyclopedia http://en.wikipedia.org/wiki/Support_vector_machine (accessed Jun 3, 2014).
- [41] Umbaugh, S.E.; Wei, Y.; Zuke, M. Feature Extraction in Image Analysis. A Program for Facilitating Data Reduction in Medical Image Classification. *IEEE Eng. Med. Biol. Mag. Q. Mag. Eng. Med. Biol. Soc.*, **1997**, 16 (4), 62-73.
- [42] Genetic programming - Wikipedia, the free encyclopedia http://en.wikipedia.org/wiki/Genetic_programming (accessed Jun 3, 2014).
- [43] Geurts, P.; Ernst, D.; Wehenkel, L. Extremely Randomized Trees. *Mach. Learn.*, **2006**, 63 (1), 3-42.
- [44] Zachary, G. P. Technology Alone Won't Improve Health in Africa [Spectral Lines]. *IEEE Spectr.*, **2015**, 52 (1), 7-7.
- [45] Dolgin, E. Portable Pathology for Africa. *IEEE Spectr.*, **2015**, 52 (1), 37-39.