Automatic Identification of Malaria Using Image Processing and Artificial Neural Network

Mahendra Kanojia[1(✉)], Niketa Gandhi[2], Leisa J. Armstrong[3], and Pranali Pednekar[4]

[1] JJT University, Jhunjhunu, Rajasthan, India
kgkmahendra@gmail.com
[2] Machine Intelligence Research Labs (MIR Labs), Auburn, Washington, USA
niketa@gmail.com
[3] Edith Cowan University, Perth, Western Australia, Australia
l.armstrong@ecu.edu.au
[4] Sathaye College, University of Mumbai, Mumbai, Maharashtra, India
panupednekar@gmail.com

Abstract. Malaria is a mosquito-borne infectious disease, which is diagnosed by visual microscopic assessment of Giemsa stained blood smears. Manual detection of malaria is very time consuming and inefficient. The automation of the detection of malarial cells would be very beneficial in the treatment of patients. This paper investigates the possibility of developing automatic malarial diagnosis process through the development of a Graphical User Interface (GUI) based detection system. The detection system carries out segmentation of red blood cells (RBC) and creates a database of these RBC sample images. The GUI based system extracts features from smear image which were used to execute a segmentation method for a particular blood smear image. The segmentation technique proposed in this paper is based on the processing of a threshold binary image. Watershed threshold transformation was used as a principal method to separate cell compounds. The approach described in this study was found to give satisfactory results for smear images with various qualitative characteristics. Some problems were noted with the segmentation process with some smear images showing over or under segmentation of cells. The paper also describes the feature extraction technique that was used to determine the important features from the RBC smear images. These features were used to differentiate between malaria infected and normal red blood cells. A set of features were proposed based on shape, intensity, contrast and texture. These features were used for input to a neural network for identification. The results from the study concluded that some features could be successfully used for the malaria detection.

Keywords: Feature extraction · Image processing · Malaria
Microscope image analysis · Matlab program · Plasmodium · Red blood cell
Segmentation

© Springer International Publishing AG, part of Springer Nature 2018
A. Abraham et al. (Eds.): ISDA 2017, AISC 736, pp. 846–857, 2018.
https://doi.org/10.1007/978-3-319-76348-4_81

1 Introduction

Malaria is caused by protozoan parasites of the genus Plasmodium [1]. It is a serious global disease and a leading cause of morbidity and mortality in tropical and sub-tropical countries. As per Centers for Disease Control and Prevention, it affects between 350–500 million people and causes more than 1 million deaths every year [2]. Yet, malaria is both preventable and curable. Rapid and correct diagnosis is an essential requirement to control the disease. There are four species of Plasmodium that infects humans and result in four types of malarial fever: P. falciparum, P. vivax, P. ovale, and P. malariae. P. vivax shows the widest distribution and is characterized by reappearances of symptoms after a latent period of up to five years. With the similar characteristics, P. ovale appears mainly in tropical Africa. P. falciparum is most common in tropical and subtropical areas [1]. It causes the most dangerous and malignant kind of malaria and contributes to the majority of deaths associated with the Malaria disease. P. malariae is also widely distributed but much less than P. vivax or P. falciparum [1, 3, 8] (Fig. 1).

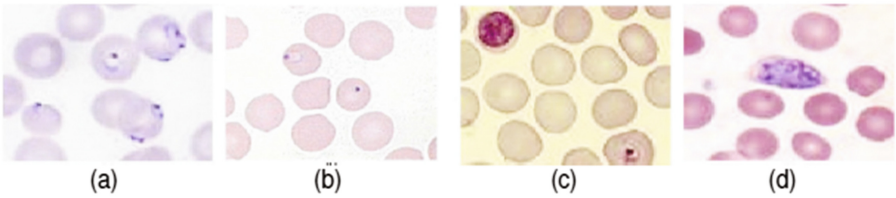

Fig. 1. (a) P. Falciparum (b) P. Vivax (c) P. Malariae (d) P. Ovale [3]

Two kinds of blood film are used in malaria microscopy. The thick film is always used to search for malaria parasites. The thin film is used to confirm the malaria parasite species. The most widely used technique for determining the development stage of the malaria disease is visual microscopical evaluation of Giemsa stained blood films [4, 8, 15]. This process consists of manually counting the infected red blood cells against the number of red blood cells in a slide prepared. The manual analysis of slides is time consuming and requires a trained operator. The accuracy of the final diagnosis also depends on the skill and experience of the technician and the time spent studying each slide. It has been observed that the agreement rates among the clinical experts for the diagnosis are surprisingly low.

The objective of this paper was to improve malaria microscopy diagnosis by removing the dependency on the performance of a human operator for the accuracy of diagnosis. This paper proposes a system to segment red blood cells, extract image features and classify the images into infected or uninfected red blood cells using thin blood smear film images and identify the type of malaria parasite using thick blood smear film images. Image processing techniques, Levenberg – Marquardt Backpropagation Neural Network and Euclidean distance measure were used for feature extraction, classification and parasite identification respectively.

2 Related Work

The majority of the research carried out so far for automated detection of malaria either includes image processing or artificial neural network techniques. There are very few instances where both the techniques are used. Further some researchers have used thick or thin blood simmers film images for the detection purpose whereas such methods cannot be termed as complete.

A study by [4] proposed a thin smear blood image classification technique using local image histogram equalization and adaptive thresholding and gradient edge detection technique. The proposed work was semiautomatic and claimed to reduce human error in detection. Another study by [3] used median filtering thresholding and morphological operations for image processing. The study applied the clump splitting algorithm for object identification and reported a sensitivity of 85.5%. Another study by [5] reported on the color normalization with K-nearest neighbor and Bayesian decision algorithm. The study used a stained pixel classifier which resulted in 88.5%, 5.6% true and false detection, parasite/non-parasite classifier achieves 74% sensitivity, 98% specificity, 88% positive prediction, 95% negative prediction. Another study by [6] reviewed the use of minimum watershed transformation and color normalization image processing methods for segmentation, whereas Bayesian pixel classifier with double threshold is reported as preferred classifier with accuracy of 96.72%.

Other studies have investigated the possibility of rapid and accurate automated diagnosis of red blood cell disorders and described a method to detect and classify malarial parasites in blood sample images [7]. The study applied image processing techniques including morphological operations and thresholding with feed forward back propagation neural network as soft computing methodology. A study by [8] adopted frequency domain discrete fourier transform to convert the images to log-polar coordinates and neural network support vector machine and reported accuracy of 96.72%. Similarly, [9] uses images taken from leishman-stained blood smears. The computing technique used are Zack's thresholding, euclidian distance and clustering. Another similar study conducted by [10] used highest quality oil immersion views (10×1000), of Giemsa stained blood films images. Two stage tree classifiers were used to achieve a sensitivity of 99%. Other studies have implemented support vector machine neural network with image processing techniques which resulted in sensitivity of 93.12% [11]. Another similar study which used only image processing techniques achieve a accuracy of 73.75% [12]. A study by [13] used machine learning model based on convolution neural network to automatically classify single cells in thin blood smears and reported accuracy of 97.37%. Work done by [14] include an automatic microscopic image acquisition system equipped with GUI based image processing modeled software to quantify the number of red blood cells. Several features grouped under shape features, intensity features, and texture features are classified and the result is presented to physician for further inspection. Review on computer vision malaria diagnostic technologies by [15] reports the challenges in commercial malaria diagnosis system such as availability of malaria parasite database to train the algorithm and disability of computer vision system to catch basic hematologic abnormalities.

3 Research Methodology

This section outlines the methods and phases used to develop a system for the accurate identification of malaria cells using either the microscopic examination using either thin or thick blood smears. A description of the techniques used for image features extraction from both the types of images for the diagnosis of malaria and malaria parasite are provided. The features extracted using thin blood smear images were simulated using Levenberg – Marquardt backpropagation neural network for diagnosis of malaria. The euler thin smear images were used to diagnose malaria whereas the thick smear images were used to identify the type of parasite.

3.1 Malaria Detection Using Thin Blood Smears Images

The input thin blood smear images are Giemsa stained. The images are processed through a number of stages as shown in Fig. 2. Details of the stages are provided below.

Image Enhancement: Image processing techniques were used to enhance the required features in the image [1, 2, 9, 10].

Color to Gray Scale Conversion: Thin smear blood images are in Red Green Blue (RGB) color space. Applying gray scale conversion gives darker shade to red blood cells and the background gets lighter shades.

Median Filter: The noise in the gray scale image was removed using nonlinear median filter [11]. The 5×5 kernel is identified to give best noise reduction results. Noise reduction was needed in order to remove false object detection and improve the results in the later stages.

Histogram Equalization for Contrast Enhancement: Filtering was used to smoothen the edges with the noise. Enhancement of the edges of the red blood cells and the contrast of the image is achieved using adaptive histogram equalization (AHE).

Image Segmentation: Ostu global thresholding and watershed algorithm was used for image segmentation. The goal of image segmentation was to segment individual red blood cells from the background [5, 9, 10, 14].

Ostu's Global Thresholding: Ostu's method [11] is the clustering based global thresholding technique. It minimizes within class variances and auto calculates optimum threshold to separate the foreground (red blood cells) and background pixels.

Flood-Fill Algorithm: Red blood cells in some blood smear images have ring-like shapes. The corresponding binary mask obtained by thresholding of such an image can result in holes in the centre of the cells. These holes need to be filled to obtain correct masks and to allow the subsequent methods to work properly. This operation was used starting with the border pixels of the background to fill the background area connected to the edges and identify the holes in the image as background pixels.

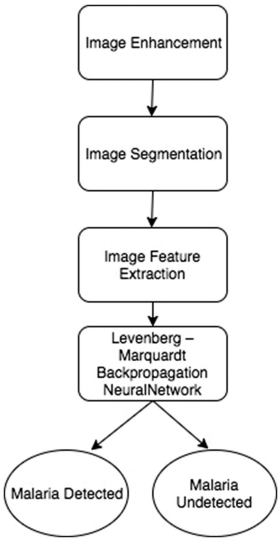

Fig. 2. Malaria detection using thin smear blood film images.

Marker Controlled Watershed Segmentation: The final stage of segmenting all the individual red blood cells from the background was achieved by using marker controlled watershed transformation [11]. This allowed the overlapping red blood cells to be uniquely identified.

Image Feature Extraction: To train the proposed neural network to diagnose malaria, all the principal image component features corresponding to red blood cells were extracted. The extracted feature vector space has large between-class distance and small within-class variance. The set of features that extracted were used to discriminate between infected and uninfected red blood. A feature based dataset was created to train the neural network [2, 5, 6, 8, 11, 13].

Shape Descriptor Features: This feature was used to identify the roundness of the identified objects.

First Order Statistics Features: The set of features such as mean, standard deviation, variance, skewness, kurtosis, fifth and sixth moment (Higher moments) and entropy under this category were extracted.

Second Order Statistics Features (GLCM): Gray-level co-occurrence matrices, contrast, correlation, energy and homogeneity are all higher order statistical features which covers the texture based information.

Levenberg – Marquardt Backpropagation Neural Network [18]. The Levenberg-Marquardt algorithm is the damped least-squares method. The loss function is the sum of squared error. It computes the gradient vector and the Jacobian matrix without computing exact Hessian matrix. This process was based on the following algorithm

1. Compute the Jacobian matrix J
2. Compute the error gradient

$$g = J^T E \text{ where E is vector of all error}$$

3. Approximate the Hessian using the cross-product Jacobian

$$H = J^T J$$

4. Solve $(\mathbf{H} + \boldsymbol{\lambda}\mathbf{I})\boldsymbol{\delta} = \mathbf{g}$ to find $\boldsymbol{\delta}$ where I is identity matrix
5. Update the network weights **w** using $\boldsymbol{\delta}$
6. Recalculate the sum of squared errors using the updated weights
7. If the sum of squared errors has not decreased.
8. Discard the new weights, increase $\boldsymbol{\lambda}$ using **v** and go to step iv.
9. Else decrease $\boldsymbol{\lambda}$ using **v** and stop.

3.2 Malaria Detection Using Thin Blood Smears Images

The thick blood smear image analysis process was used to count the number of malaria parasites which existed in each digitized red blood specimen image. Thick blood smear examination was considered a necessary part for rapid screening of malaria parasite. The proposed system carried out malaria detection based on analysis of colors. The detection of malaria parasites from thick smear images thresholding used segmentation techniques, which separates foreground (parasite) from background. The images were pre-processed to enhance the image details followed by application of thresholding technique. The stages for automatic diagnosis of malaria using thin smear blood images is shown in Fig. 3.

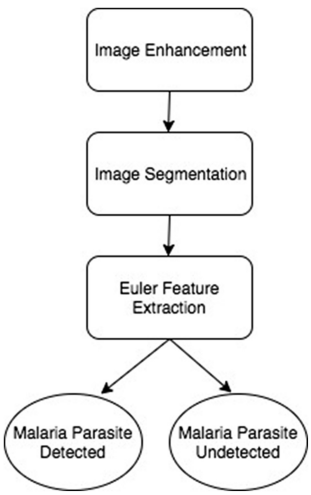

Fig. 3. Malaria detection using thick smear blood film images.

Image Enhancement: The red, green and blue planes of input color image was separated using the adaptive histogram equalization applied to each plane. This method found to be suitable for improving the local contrast and enhancing the definitions of edges in each red blood cells of an image [3] (Fig. 4).

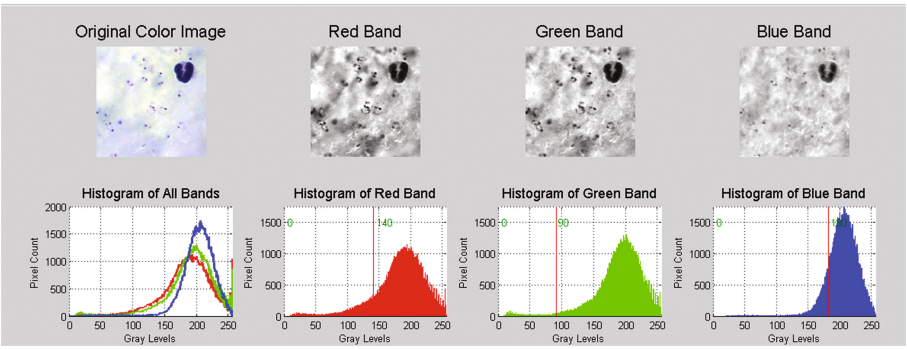

Fig. 4. Separated RGB planes of input image

Image Segmentation: Low and high threshold values for each red, green and blue planes were selected manually which provided a better segmentation result by extracting out parasites from background of blood images. The binary image produced after thresholding of all three planes were concatenated for generating parasite mask image. The borders were smoothened using morphological closing operation. Connected components where then counted for determining the number of malaria parasites present in thick blood smear images [3–5, 8–10].

Feature Extraction: The Euler number [19] for malaria thick blood smear images is the total number of parasites in the image minus the total number of holes in those images. Objects were connected sets of pixels, that is, pixels having a value of 1.

$$E = C - H$$

Where E is Euler Number, C is Connected components and H is the holes in Object. If Euler value is less than zero than the counted objects are infer as malaria parasite.

4 Proposed System

The proposed system was tested using microscopic blood images of both thick and thin smear which were obtained from an open source library CDC DPDx - Malaria image library [16]. DPDx is an education resource designed for health professionals and laboratory scientists. CDC's DPDx Parasite Image Library where parasites and

parasitic diseases are listed alphabetically and are cross-referenced [17]. The image processing techniques were applied to the data set of 134 downloaded images which included 76 thick and thin smear images respectively. System testing was carried out on 20 images collected from a pathology lab in India. The sample included 15 images which were infected and 5 images were uninfected. All the images were based on 300 X 300 in .jpg format with resolution of 72 dpi.

The MatLab software tools were used to implement image processing techniques and Levenberg – Marquardt backpropagation neural network. The GUI was designed to execute all the image processing functionality, feature extraction and saving the malaria cell identification results (Fig. 5).

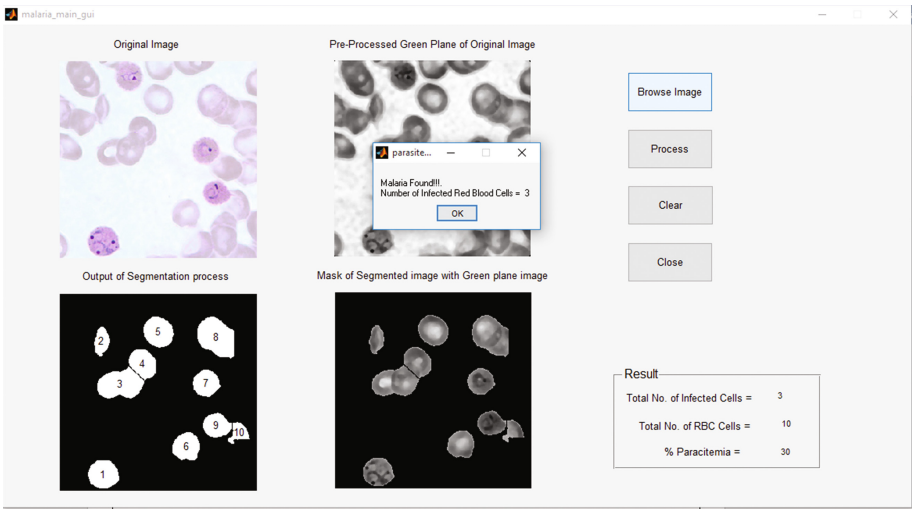

Fig. 5. GUI for thin smear blood cell image processing

5 Result

Table 1 provides results of the accuracy of identification of the malaria testing dataset. The system tested 20 images of which 5 images where from blood infected with malaria. Extracted features were simulated using SVM algorithm where the results found were less than 50% accuracy, hence Levenberg - Marquardt Backpropagation Neural Network was selected owing to better accuracy.

The proposed system correctly identified malaria cells with an accuracy of 78%. Overall performance of the system is found to be 80%.

Table 1. Result comparison (Infected by malaria)

Image No.	Actual parasite count	System count
1	1	3
2	1	1
3	3	6
4	1	1
5	1	1
6	2	2
7	2	2
8	4	3
9	2	2
10	2	1

Table 2 gives comprehensive comparison of the image processing techniques and soft computing techniques used by the related work done with the proposed system.

Table 2. Proposed system comparison with other methods

Reference	Image processing	Soft computing technique	Result	Comparative remark
[1]	Local histogram equalization adaptive thresholding Image segmentation and intensity based malaria identification	Not used	Localization of infected RBC	Feature extraction not done, soft computing not implemented
[2]	Median filtering thresholding and morphological operations	Decision tree and back propagation feed forward algorithm	Sensitivity = 85.5%. Positive predictive value = 81%	Analysis of overlapping cells is absent
[3]	Color normalization	Bayesian classifier, KNN algorithm	Sensitivity = 98%. Positive prediction = 88%,	Data set used is very less to develop reliability
[6]	Discrete fourier transform	Support vector machine	Accuracy: Neural network = 7853% SVM = 96.72%	GUI and complete automation missing
[7]	HSV color space, sequential edge linking algorithm and clustering	Not used	Localization of infected RBC	Less image database, soft computing not implemented

(*continued*)

Table 2. (*continued*)

Reference	Image processing	Soft computing technique	Result	Comparative remark
	based on Euclidian distance			
[13]	Not used	Convolutional Neural Network (CNN)	Accuracy of 97.37%	Image processing not implemented
[14]	RGB to HSV, adaptive estimated threshold, Ostu's thresholding	Not used	Test results are compared with human experts	No statistical results reported
Proposed system	Color to gray scale conversion, median filter, histogram equalization for contrast enhancement, Ostu's global thresholding, flood-fill algorithm, marker controlled watershed segmentation	Levenberg – Marquardt Back Propagation Neural Network	Accuracy = 72%	GUI implemented, 134 image dataset, real time implementation

6 Conclusion

This paper presents a proposed method for segmentation of red blood cells in microscopic blood smear images infected with malaria cells. The paper also described the design of a graphical user interface for automation of this process. The main purpose of the GUI was to execute the segmentation method, label the samples and save the information about individual segmented cells. The GUI based system was used to create a database of labeled red blood cells containing non-infected cells and cells infected by malaria parasites is created. A set of features was extracted from the area of a red blood cell in order to detect malaria.

The segmentation method presented a relatively simple solution of separating red blood cells from the background and isolating overlapping or occluded cell. The method was based on processing of a binary image obtained by thresholding and utilizes the watershed transformation for separating cell compounds. The method produced good results on all input images with only occasional over-segmented and under-segmented cells.

The designed GUI provided the tools necessary for segmenting cells, labeling of samples, manipulating with files, and saving the results. As some future enhancement certain improvements could be made in the GUI including the implementation of some semi-automatic methods for the correction of cell contours, a set of shortcut keys, the use of mouse wheel for zooming, etc. Some of these improvements would be dependent on the programing environment. Future studies will be devoted to enhance the accuracy of the proposed system using larger data set.

There are many possible classification scenarios that can exist with the malaria cell diagnosis. This paper has described a study which has focused on distinguishing between infected and non-infected red blood cells. The findings from the study indicated the potential to use this system to identify malaria cells from images taken from both thin and thick blood smear.

References

1. World Health Organization. Malaria, Fact sheet no. 094 (2017)
2. Centers for Disease Control and Prevention. CDC Health Information for International Travel 2010. Atlanta, Mosby, pp. 128–159 (2009)
3. Sudhakar, P.: Detection of malarial parasite in blood using image processing. Int. J. Eng. Innov. Technol. **2**(10), 124–126 (2013)
4. Khan, M., Acharya, B., Singh, B., Soni, J.: Content based image retrieval approaches for detection of malarial parasite in blood images. Int. J. Biom. Bioinform. **2**, 97–110 (2011)
5. Tek, B., Dempster, A., Kale, I.: Malaria parasite detection in peripheral blood images. In: BMVC, pp. 347–356 (2006)
6. Tek, B., Dempster, A., Kale, I.: Computer vision for microscopy diagnosis of malaria. Bio-Med Central Malar. J. **8**(1), 153 (2009)
7. Ahirwar, N., Pattnaik, S., Acharya, B.: Advanced image analysis based system for automation detection and classification of malarial parasite in blood images. Int. J. Inf. Technol. Knowl. Manag. **5**(1), 59–64 (2012)
8. Annaldas, S., Shirgan, S.: Automatic diagnosis of malaria parasites using neural network and support vector machine. Int. J. Adv. Found. Res. Comput. (IJAFRC) **2**, 60–66 (2015)
9. Damahe, L., Krrishna, K., Janwe, J., Nileshsingh, T.: Segmentation based approach to detect parasites and RBCs in blood cell images. Int. J. Comput. Sci. Appl. **4**(2), 71–81 (2011)
10. Soni, J., Mishra, N.: Advanced image analysis based system for automatic detection of malarial parasite in blood images. In: International Conference on Advanced Computing Communication and Networks, pp. 59–64 (2011)
11. Savkare, S., Narote, P.: Automatic detection of malaria parasites for estimating parasitemia. Int. J. Comput. Sci. Secur. (IJCSS) **5**(3), 310–315 (2011)
12. Ghate, D., Jadhav, C., Rani, U.: Automatic detection of malaria parasite from blood images. Int. J. Adv. Comput. Technol. **4**(1), 129–132 (2012)
13. Liang, Z., Powell, A., Ersoy, I., Poostchi, M., Silamut, K., Palaniappan, K., Guo, P., Hossain, M., Sameer, A., Maude, R., Huang, J., Jaeger, S., Thoma, G.: CNN-based image analysis for malaria diagnosis. In: IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Shenzhen, pp. 493–496 (2016)
14. Mehrjou, A., Abbasian, T., Izadi, M.: Automatic malaria diagnosis system. In: First RSI/ISM International Conference on Robotics and Mechatronics (ICRoM), pp. 205–211 (2013)
15. Pollak, J., Houri-Yafin, A., Salpeter, S.: Computer vision malaria diagnostic systems—progress and prospects. Front. Public Health **5**, 1–5 (2017)
16. Centers for Disease Control and Prevention (CDC) Laboratory Identification of Parasitic Diseases of Public Health Concern (DPDx) - Malaria image library. http://www.cdc.gov/dpdx/malaria/gallery.html. Accessed 04 May 2017

17. Centers for Disease Control and Prevention (CDC) Laboratory Identification of Parasitic Diseases of Public Health Concern (DPDx) - Malaria image library listed alphabetically and cross referenced. http://www.cdc.gov/dpdx/az.html. Accessed 04 May 2017
18. Ahiwar, N., Patnaik, S.: Texture and intensity based classification of malaria parasite in blood images using Levenberg-Marquardt (LM) algorithm. Int. J. Comput. Intell. Inf. Secur. **2**(12), 4–9 (2011)
19. Chakrabortya, K., Chattopadhyayb, A., Chakrabarti, A., Acharyad, T.: A combined algorithm for malaria detection from thick smear blood slides. Health Med. Inf. **6**(1), 2–6 (2015)