Heuristic Scale to Estimate Premature Malaria Parasites: Scope in Microscopic Blood Smear Images

Chaya D. Jagtap* and N. Usha Rani

Department of ECE, Vignan University, Guntur - 522213, Andhra Pradesh, India; chaya123jadhav@gmail.com, usharani.nsaii@gmail.com

Abstract

Objective: Malaria is one of the epidemic diseases and early detection of malaria symptoms in the patients using the current manual procedures are skeptical, as the diagnosis patterns depends more on the experience of professionals. To overcome the challenges, in this paper we are proposing computer aided model to support in malaria detection at early stages using Microscopic Blood Smear Images analysis using machine learning. **Methods/Statistical Analysis**: There are many computer aided models that were proposed and adapted in the process of addressing the diagnosis models. Some models like machine learning, image processing, neural network based solutions etc are adapted, which reflects more insights into the process. However, the issue of gaps in accuracy still persists, and the proposed model of Heuristic Scale to Estimate Premature Malaria Parasites Scope (SEMPS) with multi stage processing of the microscopic images of blood smear is processed. **Findings**: The proposed model is compared with the benchmark models like SVM and Bayesian, the outcome in terms of efficiency of the model is imperative from the results. The proposed model has resulted in more effective and accurate detection of malaria symptoms in the test cases, and the result accuracy is higher than the other two benchmarking models of SVM and Bayesian techniques chosen for comparative analysis. **Improvements:** The computational complexity of the SEMPS is evinced as linear, where the majority of benchmarking models are found to be up-hard.

Keywords: Blood Samples, Case Based Reasoning, Disease Diagnosis, Erythrocyte, Heuristic Scale, Machine Learning, Malaria Parasite, Soft Computing

1. Introduction

Malaria is an endemic disease more prevalent in African and Asian continents, and the number of cases reported for malaria related deaths are on surge. One of the key challenges in the diagnosis of Malaria is about the detection at the premature levels, and there is significant need for early diagnosis of the conditions to ensure better treatments and reducing the implications of resulting impacts like life threat etc. From the review of literature on Malaria[1], malaria parasites (Species of Parasitic Protozoa) are categorized in to four types as:

- *Plasmodium ovale.*
- *Plasmodium falciparum[2].*
- *Plasmodium vivax[2].*
- *Plasmodium malaria[3].*

In the hot and humid weather conditions, the parasite *Plasmodium vivax* is found in significant manner[4]. To ensure that suitable drugs are administered to the human body ailing with disease, early detection of such parasites in the human blood at premature level is profoundly important. WHO guidelines to the medical practitioners[5, 6], recommends that the suspected cases of malaria parasite existence are recommended to microscopic diagnostic tests of blood smears, if can help identification and treating the disease at its early stages.

Certainly the microscopic diagnostic tests shall support in identifying parasite type differentiation and also

---

Author for correspondence

in quantification of the presence to ensure that the disease severity is analyzed. Rapid-diagnostic tests are the other alternative method in terms of diagnostics, but one of the key limitations in Rapid-diagnostic model is the non-feasibility of detecting the disease at pre-mature stages.

Minimal costs and scope of scalability is leading to microscopic diagnostic tests being preferred and adapted. Key strategy in microscopic diagnostics is pertinent to thick and thin blood smears in the blood samples collected from the suspect cases. In identifying the parasite influence, thick smears are analyzed, while the type of parasite can be determined using thin smears[6]. Thick smears can be very resourceful than the thin smears, in identification of the malaria symptoms at premature stage.

Microscopic diagnostic tests provide inputs on the parasite type and quantification observed in the blood smears. The results from the tests have to be affirmed by the authorized professionals having substantial domain knowledge, thus resulting in many of inaccurate detections or non-diagnosis. Adapting computational technology aided solutions for addressing the limitations and using the computing solutions to overcome the problem of early diagnosis issues could be an effective solution.

For instance, the machine learning model could be very significant for developing a solution with computer aided models. A solution like features obtained from microscopic images of erythrocytes and other normal blood cells could be very vital in differentiating the premature state of disease scope.

In the collection of microscopic images, digital image processing solutions can be very vital[7–18] and more specifically the edge based segmentation could be an effective model[19] for analysis. However, some of the limitations in the edge based segmentation of microscopic images for erythrocytes could be with limited identification of parasites due to contrast issues, lack of clarity on the edges resulting from color similarities , irregular edges and also in terms of noise intensity in the parasite effected;

Considering such limitations in the models, the proposed model is about a new machine learning strategy termed CUCKOO Search[20] developed on the basis of evolutionary computational model, which adapts morphological features and benchmark textures as the basis. The proposed solution shall address the limitations envisaged in the edge based segmentation.

Numerous researches and developments has emerged since last decade of time, over the process of computer aided malaria diagnosis, which has grabbed the attention for microscopic image analysis. Many benchmarking contributions have been proposed in the earlier studies on the model. Some of them are based on supervised learning methods[21–23] and the decision support system[24] based solution, digital image analysis[25–28], and some based on the pattern recognition solution[29,30]. Also, some of the solutions were proposed adapting artificial neural network solutions[31,32].

With the rising adaptation of cross modeling, some significant models like the segmentation based on histogram equalization which is used to classify the overlapping infected cells were also considered in the process[33]. There are many unsupervised models also found in the review of literature[34–37]. Also some of the machine learning models[38] and content based image retrieval models[39] has also been proposed in the earlier studies, like the parasite estimation using segmented digitized blood smears are some of the significant contributions that are reviewed in the literature.

Some of the critical constraints that are envisaged in the benchmarking models are about the dullness in contrast, intensities observed in similar fashion for both effected and normal areas of blood smear images for microscopic image segmentation modeling, edge formation issues and other such factors are turning out to be a major impact. Also, apart from one model devised in[38], all the other models depend on signatures of effected images to identify the parasite scope in a new image. Such dependency, in the instances of trivial variations could lead to more of false alarming conditions. However, such constraints could be effectively addressed in the machine learning models, but one of the constraints envisaged in the machine learning models is about need for considerable volume of microscopic images as inputs essential for training the system, and it should adapt optimal features.

Around 94 features are used in the machine learning method proposed in[38], and the SVM machine learning models[40] and also the Bayesian Classifiers[41] are also adapted to train and test the model for effective machine learning. Among the varied models, the optimal range of feature selection has resulted in One-way-ANOVA[42]. Though the detection accuracy levels of the model is reaching to 84%, still there is significant instability in divergent count of features. Also, the issues of process complexity towards resource utilization is also not linear for divergent count of features in the model, and the other key constraint is about feature extraction which is carried out by segmentation of image using an algorithm called

Marker Controlled Watershed model[14]. The algorithm in[14] uses the gradients that happen under segmentation for estimation, thus the chances of feature value optimality may not be so effective.

Considering such factors, the model of evolutionary computation based machine learning system is adapted to overcome the limits that are identified in[38]. The proposed model CUCKOO of search shall be used over the edge based segmentation model to extract the features from the chosen blood sample images, as the edge based segmentation is one of the effective methods for image based analysis of parasite type detection[37].

In the further sections of this report, the model proposed in section-II. The experimental study results are depicted in Section 3 along with performance analysis and the conclusion for the model is depicted in Section 4.

2. Heuristic Scale to Estimate Premature Malaria Parasites Scope

The definition of scale is defined using hierarchy of multiple stages like acquiring the images of blood smears in the image inputs, ensuring that the images are preprocessed. Extracting some of the bench marking features identification of the optimal features and then applying the Cuckoo search on the optimal features in a relative manner, for defining the heuristic scales to depict the normal blood smears and the parasite prone blood smears at premature level.

2.1 Conversion to Grayscale

Using the Colorimetric gray conversion [43]Luma coding grayscale conversion[44], and Green channel gray conversion[44], the RGB (three channels) microscopic image is converted to a single channel. Also PCA- related grayscale conversion[45] is also adapted in the approach for conversion, and among all the conversion methods adapted, PCA based grey scale has provided more optimal quality images of microscopic blood smear[46–48]. Hence PCA is adapted as an effective approach for RGB microscopic image conversion to grayscale.

Using the linear least-square model, the maximum contrast in the grayscale is developed in the PAC based model. RGB color coordinate is used for assessing the primary axis of the RBC color and the best fit regression line

developed by PCA regression which certainly reduces the distance between point and the axis line that are influenced by parasite impacted cell image in the regression space. Also, the illustrative visualization of the regression line, with inputs to RGB image and the output grayscale is depicted in Figure 1.

The variation of angles amidst the angles could be represented by R, G and B that are obtained first and sequentially, the cosine values of them are transformed to grayscale values. Equation that is used to gather the regression weights are:

$$\wedge(x,y,z) = \frac{\sum_{i=0}^{|P_{ix}|}(xr_k + yg_k + zb_k)}{\sqrt{x^2+y^2+z^2}} \quad (1)$$

$$agr(x,y,z) = \sum_{i=0}^{|P_{ix}|}(xr_k + yg_k + zb_k) \quad (a)$$

$$sqrt(x,y,z) = \sqrt{x^2+y^2+z^2} \quad (b)$$

$$\wedge(x,y,z) = \frac{agr(x,y,z)}{sqrt(x,y,z)}$$

In the above depicted equation (Eq.1), minimum of the weights x, y, z that are applied to R, G and B. The image of the pixel count is depicted by $|P_{ix}|$ and r_i, g_i, b_i denotes the red, green and blue values of the ith pixel.

2.2 Contrast Correction in Illumination

The poor illumination of the microscopic images could be attributed to varied range of conditions. However, in terms of identifying infected erythrocytes any kind of dullness persisting in contrast shall be a major constraint. Hence, the emphasis is on improving the contrast levels is a key
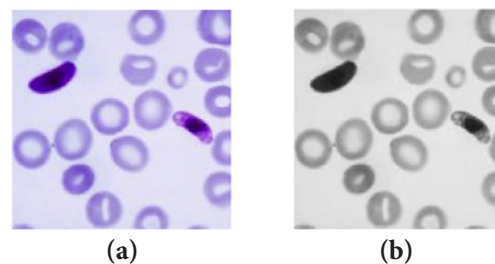

(a)　　　　(b)

Figure 1. Imagery example of PCA image conversion from RGB image to grayscale. **(a)** RGB input image. **(b)** Grayscale output image.

step in the segmentation process. To increase the contrast level of gray scale image $b(\overline{m},\overline{n})$ the technique termed gamma equalization (GE) [49,14] is adapted as depicted in the equation below(see Eq2):

$$b(\overline{m},\overline{n}) = \gamma_{max}(\overline{m},\overline{n}) * \left| \frac{(\gamma(\overline{m},\overline{n}) - \gamma_{min}(\overline{m},\overline{n})^{\tau})}{(\gamma_{max}(\overline{m},\overline{n}) - \gamma_{min}(\overline{m},\overline{n})^{\tau})} \right| \quad (2)$$

Under the gamma threshold levels of 0.5, the contrast gray scale image for the input gray scale can be seen in the Figure 2.

The given grayscale input image and the resultant image from the gamma equalization process under gamma threshold 0.5 can be found in Figure 2.

2.3 Noise Tumbling

The noise of some common type called Salt & Pepper and other super imposed patterns always impact the microscopic images of blood smears[50]. Technique of Median filter is adapted for removing such noise patterns[51], and also the Median filter is used in combination of Gaussian Filter [50] to reduce the super imposed patterns in the blood smear images shown in Figure 3.

- Identifying spectral peaks of pattern noise

Special filtering process is applied to any kind of Moiré pattern noise from Fourier amplitude spectrum in a given image using (Equation (3))

$$L(a,b) = \sum_{m=0}^{S_1-1} \sum_{n=0}^{R_1-1} \left(l(\overline{m},\overline{n}) \gamma^{2\pi((a\overline{m}/R_1)+(b\overline{n}/S_1))} \right) \quad (3)$$

$l(c, d)$ Is the image of size $R_1 \times S_1$. The amplitude of the spectral $(a, b)th$ coefficients is $L (a, b)$. The peaks shall
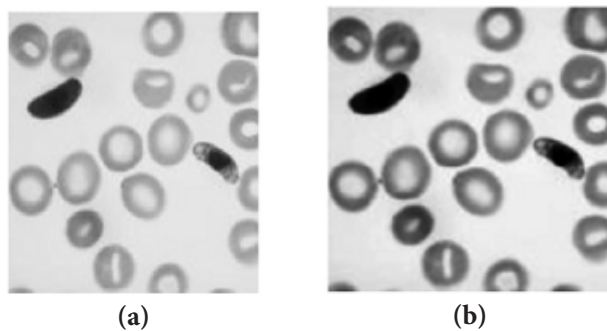

(a)   **(b)**

Figure 2. Imagery of Gamma equalization process. **(a)** Input Image for Gamma Equalization. **(b)** Output image of the Gamma Equalization for $(\tau = 0.5)$.
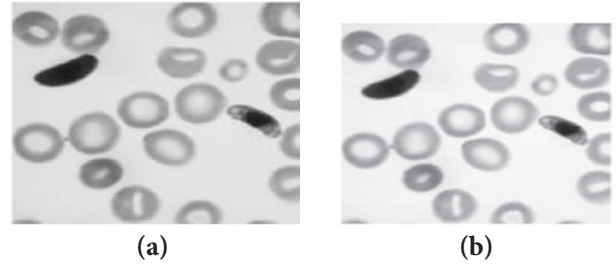

(a)   **(b)**

Figure 3. Input and resultant blood smear images of applying combined Gaussian and median filter. **(a)** Grayscale image input. **(b)** Output image.

be visible as bright spots in amplitude spectra, similar to impulses in visual conditions. Any kind of impulsive noises can be detected by profound solution of median filter[52–54] and hence such solution is adapted in detection and filtering of noise. Also the following two steps are followed in detection o spectral peak in the Fourier amplitude spectrum.

- Defining low-frequency area

The transformation of the initial image in to 2D wavelet is carried out, and if the relevant functions are identified to be discrete, the scaling and aggregate of wavelets shall be bipartite in to two phases. Hence, the wavelet transform is applied to each axis initially. In extension to such bipartite signal (in terms of image in the form of 2D signal) has been portioned in to sub bands as LL, HL, LH and HH[55].

The sub bands HL and LH depict the signal deviation in X-axis and Y-axis in the decomposed image. To improvise the ability of coding, the maximal bits and minimal or zero bits are spent on low and high frequency bands in respective manner.

- Localize a spectral peak

The $(a, b)th$ spectral coefficient $L_{ab}^{\overline{mn}}$ is considered as peak, if the following justification exists:

$$\left\{ \frac{L_{ab}}{MED_{A\times B}(L_{ab})} \geq T \exists (a,b) \notin A_{BF} \right\} \quad (4)$$

The window of size with local median of $A \times B$ surrounding to $(a, b)^{th}$ spectral coefficient L_{ab} is $MED_{A\times B}(L_{ab})$ and pre-defined threshold value denoted by T.

Number of peaks identified the threshold value than T are found inversely proportional.

- Gaussian Filtering:

Detection of spectral peak is adapted for Gaussian filtering which corrects spectral coefficient of interest and

amplitude towards spectrum coefficients and the filtering process is carried out as

$$\overline{L}_{ab}^{\overline{mn}} = \begin{cases} L_{ab}^{\overline{mn}} \otimes G_{\overline{mn}}, if \left(\dfrac{L_{ab}}{MED_{A \times B}(L_{ab})} \geq T \exists (a,b) \notin A_{BF} \right) \\ L_{ab}^{\overline{mn}} \end{cases}$$

(5)

$L_{ab}^{\overline{mn}}$ Indicates the set of amplitude spectrum coefficients chosen from the window of size $\overline{m} \times \overline{n}$ in surround of $(a,b)^{th}$ spectral coefficient L_{ab}. The Gaussian filtering process (Equation (5)) results set of amplitude spectrum coefficients $\overline{L}_{ab}^{\overline{mn}}$ and the Gaussian filter used is represented by $G_{\overline{mn}}$.

Also a typical Gaussian filter shall be applied even to explore Gaussian surface which is covered by two connected peaks. The process overhead shall be found to be adapted in Gaussian filtering process and it contains some pairs of noise peaks and also the process overhead could be substantially high.

It is imperative from the above process that the usage of Median and Gaussian filters are very much divergent but effective in terms of noise filtering. Also, the median filters restrict the size of region around noise peaks and also further when Gaussian filter is adapted it shall perform over the region defined by median filter.

2.4 Edge Detection

The initial and key objective is to improving the visibility of the borders of erythrocytes, which is done by Canny-based filters[56] that preserves continuous edges in effective manner. In order to this the median filter will be used initially to smoothen the contours formed on target image of size $q \times q$ pixels of noise free image. Since the infected area of the erythrocyte appears as dark area, the edges and the borders related to such darkest regions are highlighted in Figure 4.
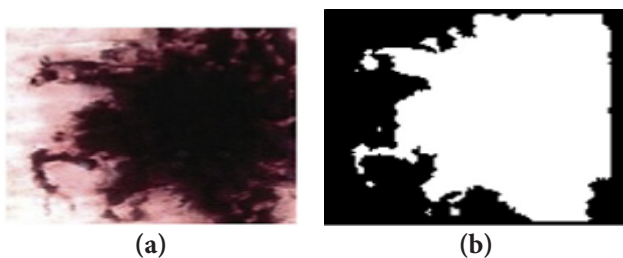

(a)   **(b)**

Figure 4.   Edge detection on image using canny edge detection approach. **(a)** Microscopic image input. **(b)** Image with edge detection as output.

2.5 Segmentation by K-Means Clustering to Identify Erythrocyte

In the preprocessing stage, the grayscale Microscopic blood smear image is delivered and it is used as input for process of segmentation by using K-Means algorithm that has K value as 2, which is since the pixels either fall into infected area or normal region of the erythrocyte Figure 5.

The simple clustering technique of K-Means[57] shall be adapted for clustering the microscopic image data sets that are considered. If the dataset U is clustered in d_i dimension space as k clusters, as in the selected context the value of k is 2. Initially the normal and infected erythrocytes shall be used to create prototypes in such a manner that it denotes respective cluster.

Then each entry of dataset U shall be moved to respective clusters that are based in nearest prototype. Also, for each of the prototype cluster that is identified, if any other cluster prototype is different to the earlier prototype, the clustering is done according to the cluster prototypes. However, if no change in the prototypes is envisaged, all the clusters observed then shall be used with respective entries that are finalized.

The objective function that estimates the squared error is adapted in K-means clustering for identifying the nearness of each entry to the dataset U and hence respective prototypes of the clusters are adapted as follows:

The objective function, which estimates the squared error, is used in K-means clustering to identify the nearness of the each entry of the dataset U and respective prototypes of the clusters that is as follows.

$$f(x) = \sum_{j=1}^{c} \sum_{i=1}^{n} \left\| U_i^{(j)} - v_j \right\|^2$$
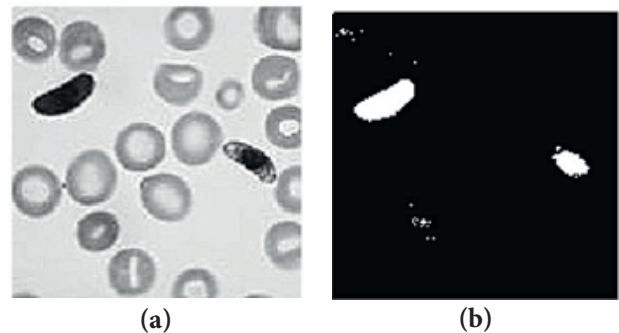
(6)

(a)   **(b)**

Figure 5.   The input and output imageryof K-Means Clustering. (a) Microscopic image for K-Means clustering – Input. (b) Image of two clusters effected (white), uneffected (dark).

The distance amid each point $U_i^{(j)}$ and the cluster v_l is done by $\left\| U_i^{(j)} - v_j \right\|^2$ (see Equation (6)). Using this objective functions, the results that are obtained denotes the distance among the data points and respective cluster. Key steps that are adapted in the K-Means process are:

For the K clusters, K data points shall be considered as K data points (k value is considered as Equation (2))

1. Distance between k centroids and data points are identified.
2. Moving the data point to cluster once the cluster of a centroid is found within minimal distance.
3. Application of step 3 to all the data points.
4. For each cluster optimal centroid is searched for.
5. The process is repeated of 2, 3, and 4 are repeated until no changes are observed in any centroid.

I. Apply the step III to all data points.
II. Find optimal centroid of each cluster.
III. Repeat steps II, III and IV until the centroid of any cluster not changed.

K-means reflects significant times and also the cluster optimality which shall have to be proportionate for centroids that are initially adapted. To the selected binary image by k-means.

Clustering of the given binary image by k-means shall be explored as:

The number of clusters that are set to be 2, when the infected erythrocytes shall contain pixels that has high intensity is used for normal erythrocytes.

As the input one is a grayscale image, the scope of differentiation that is estimated by their intensity.

Also, the set of 2 clusters that are formed by assessing at varied level of pixel shall be darkest area or even it might not be one. To evaluate such conditions, for the initial centroids of cluster 1 and 2, pixels are randomly selected for the darkest area for an image and the pixel shall randomly be selected from other part of the image in respective manner.

2.6 Connected Component Analysis

In terms of noise removal in the resultant images from the FCM process, shall be the emphasis in the process, however, for the connected component analysis, there is morphological method with a kind of erosion process which shall be adapted. Also, the holes that are observed in the resulting images shall be achieved using the optimal segmentation.

i. Morphological Operation
Also, in terms of resulting binary images to the infected erythrocytes in a clustering process comprising of spanned excess regions hovering on infected erythrocytes, such excess regions has to be eliminated. Such process can be done by focusing on morphological binary destruction operation[14]. Also, the structuring element $s(\overline{m}, \overline{n})$ shall be used for destruct binary image $b(\overline{m}, \overline{n})$ that delivers resultant binary image $r(\overline{m}, \overline{n})$. The destruction is carried out as explored in Equation (7).

$$r(\overline{m}, \overline{n}) = \left\{ b(\overline{m}, \overline{n}) \wedge b(\overline{m}^{-1}, \overline{n}^{-1}) \exists b(\overline{m}^{-1}, \overline{n}^{-1}) \in s(\overline{m}, \overline{n}) \right\} \quad (7)$$

To identify optimal size of the STREL, tests were conducted on $3 \times 3, 5 \times 5 \, and \, 7 \times 7$ square particles of STREL and it is imperative that 3×3 squared particles are optimal.

ii. Filling Holes observed in clustered erythrosine areas
The resulting images from k-means depict erythrocytes that have holes and also the common obstacles shall be around segmentation accuracy for infected erythrocytes[58]. Hence the holes have to be filled as follows.

If $b(\overline{m}, \overline{n})$ be the output of the K-means process for a parasite binary image. And $\overline{m}_i(\overline{m}, \overline{n})$ be the marker image used for filling the image and surroundings to highlight the border of the image. Then resultant image $r(\overline{m}, \overline{n})$ is $b(\overline{m}, \overline{n})$ with hole filled (see Equation (8)). The result generated from connected component analysis which is applied on resultant image of K-means is visualized in Figure 6.

2.7 Features Extraction

The features for the contextual differentiating in the texture and also the morphological patterns for the given greyscale images are profoundly related in the literature on the domain. In[7–18], which are not considered towards distinguishing the normal and diseased erythrocytes.

• Entropy
In terms of entropy which indicates the level of uncertainty, it is important to differentiate the infected and normal erythrocytes, and the model usually relies on available entropies that are explored in the literature [15, 16]. For measuring the entropies, some of the factors like histogram of the region of interest have to be taken in to account. Also, in terms of entropies that have to be evaluated, there
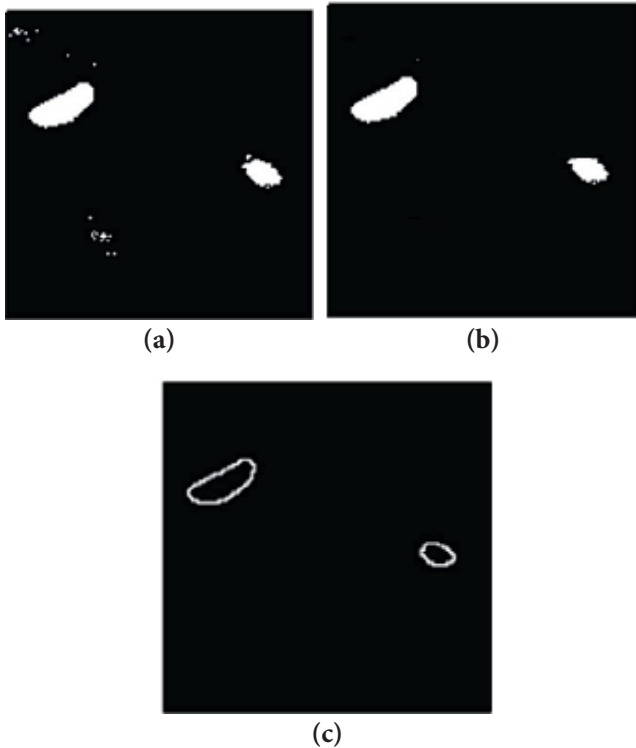
(a)   **(b)**

(c)

Figure 6. The input and resultant images for connected component analysis. **(a)** Input image for connected component analysis. **(b)** Resultant image with removal of falsely clustered erythrosite. **(c)** Resultant image with filled holes.

are five entropies that are considered (Kapur's[15], Renyi's[15], Charvat's[15], Havarda's[15], and Yeager's[16] entropies).

• GLCM features

The total GLCM features are 19 and all of them are related to information measure and variance, entropy and energy related aspects [29,30], that are substantial for exploring the texture information[8]. Also, the GLCM matrix which indicates the divergent grey shades that is found in the image. The matrix shall be used for describing the metrics defined.

• Gray Level Run Length Matrix[8]-based textural features

The features analyze the granular structure for a grayscale image. If $RLM_{(i,j)}$ is a matrix for run lengths, of size j, towards a specified direction for grey value i of a grayscale image $I(x, y)$, that is used for defining all 11 texture features[8,10,11].

• Fractal dimension

Using the fractal dimension model, the surface coarseness in an image can be figured out [59]. To ensure that the

infected erythrocyte in terms of surface coarseness, the grayscale image shall be processed as third dimension for the chosen 2 dimension image and the resulting variation from the process shall conclude coarseness or texture deviation in the infected erythrocyte. Also, the approach that is adapted for identification of the variations towards third dimension shall be as "modified differential box counting with sequential algorithm"[9,12].

• Local Binary Pattern (LBP)

The grayscale image termed Local Binary Pattern (LBP) reflects the similarity of the local neighbor regions and the other total number of LBP features are 6[13,17]. The circular neighborhood and the bilinear interpolation are some of the key factors that are used for computation of LBP.

Let P_{ix} be the set of pixels found in the range of radius R of erythrocyte image I, let p_c be the pixel found at center of the radius R, let Gp_c be the grey value of the p_c and $\left\{ Gp_i \exists p \in P_{ix} \wedge i = 1,2,3,...\left| P_{ix} \right| \wedge p_i \neq p_c \right\}$ be the grey value of the neighborhood pixel $\left\{ p \exists p \in P_{ix} \right\}$. Then each pixel $\left\{ p \exists p \in P_{ix} \wedge p \neq p_c \right\}$ shall be converted in to binary pattern (0 or 1) based on the Gp_c.

• Morphological feature

Some the features that are proposed in [14], that has invariant moments[7,60], shall be considered as Morphometric information features that are extensive for depicting anomalous erythrocytes recognition. Also this could be attributed to variance of shape and size depicted amidst of infected and normal erythrocytes.

2.8 Features Selection

When E and F shall be set of records, for each record $\left\{ re \exists re \in E \right\}, \left\{ rf \exists rf \in F \right\}$ denotes all features for an normal erythrocyte and also infected and respectively

The task of finding hamming distance amid unique values towards each attribute of E for the counter part of F

Selection of attributes with hamming distance for more than the threshold *hdt* that set of optimal attributes E_a of size n, F_a of size m from E and F respectively

i. Assessing Hamming Distance is as follows:
Difference between the unique values of same attribute for the records identified as true and false are generated from Hamming Distance. One of the key strategies for assessing the difference for elements is in the coding theory adapted. Such strategy shall be applied for managing distance between various unique values that are observed and also for an attribute for record set that is labeled as true or false.

For a given two vectors $V = \{v_1, v_2, ..., v_m\}$ and $W = \{w_1, w_2, ..., w_n\}$ of size c and d respectively. Hamming Distance can be measured as follows

Let $Z \leftarrow \phi$ // is a vector of size 0

$foreach\ \{i \exists i = 1, 2, 3, ... \max(c, d)\}$ Begin

$if\ (\{v_i \exists v_i \in V\} - \{w_i \exists w_i \in W\}) \equiv 0\ then$

$Z \leftarrow \{v_i \exists v_i \in V\} - \{w_i \exists w_i \in W\}$

Else

$Z \leftarrow 1$

End

$$hd_{V \leftrightarrow W} = \frac{\sum_{j=1}^{|Z|} Z\{i\}}{\max(c, d)}$$

// $hd_{V \leftrightarrow W}$ is the hamming distance between V and W, $Z\{i\}$ is the ith element of the vector Z with size $|Z|$.

2.9 Applying CUCKOO Search

i. Nest Formation

Prepare $2^n - 1$ unique subsets $\{E_a\}$ such that at least one subset of sizes $\{n, n-1, n-2, ---, 1\}$ and $2^m - 1$ unique subsets $\{F_a\}$ such that at least one subset of sizes $\{m, m-1, m-2, ---, 1\}$ from E_a and F_a respectively.

Prepare a set $\{E_f\}$ such that $\{E_f\}$ contains the set of respective values appeared in 1 or more records of E for attributes of each subset of $\{E_a\}$

Prepare a set $\{F_f\}$ such that $\{F_f\}$ contains the set of respective values appeared in 1 or more records of E for attributes of each subset of $\{F_f\}$

Build a hierarchical order TI such that subset $[s \exists s \in \{I_a\}]$ with size of $|E_a|$ as root and subsets with size $|E_a| - 1$ as level 1 nodes and continue further building of that order with all subsets of the $\{E_a\}$ as the subsets of size $|E_a| - (i+1)$ as the next level nodes to the level formed by the subsets of size $|E_a| - i$

Similarly build hierarchical order TN such that subset $[s \exists s \in \{F_a\}]$ with size of $|F_a|$ as root and subsets with size $|F_a| - 1$ as level 1 nodes and continue further building of that hierarchy with all subsets of the $\{F_a\}$ as the subsets of size $|F_a| - (i+1)$ as the next level nodes to the level formed by the subsets of size $|F_a| - i$

Further the nodes at different levels of these hierarchies are considered as nests identified by the combination of level id and node id

For each node of hierarchy TI that representing a nest, the respective value sets those belongs $\{E_f\}$ are considered as eggs

$\{E_f\}_{ij}$ is the set of eggs belongs to the nest j in level i of hierarchy TI.

For each node of hierarchy TN, the respective value sets those belongs $\{F_f\}$ are considered to be as eggs in the nest represented by that node.

$\{F_f\}_{ij}$ is the set of eggs belongs to the nest j in level i of hierarchy TN.

ii. Cuckoo Search

For a given record R that represents the features of the erythrocyte, form the 2^n subsets $\{RE_f\}$ from the values of the optimal features reflected by E_a and also form the 2^m subsets $\{RF_f\}$ from the values of the optimal features reflected by F_a.

The $\{RE_f\}$ and $\{RF_f\}$ are considered as set of cuckoo eggs to be placed in nests represented by TI and TN respectively.

a) Cuckoo Search on TI.

Sort the $\{RE_f\}$ in descending order of sizes

$S(E) \leftarrow \phi$ // vector that represents the number of compatible nests in each level of hierarchy TI

For each level $\{l \exists l = 1, 2, 3, ... n\}$ Begin

// l represents the level of the hierarchy from root to leaves respectively

$s_l = 0$ // possible number of nests in level l accommodates cuckoo eggs

For each cuckoo egg $[ce \exists ce \in \{RE_f\} \wedge |ce| \equiv (n - l + 1)]$ Begin

// $(n - l + 1)$ represents the number of features in cuckoo egg ce

For each $[j \exists j = 1, 2, 3, ... |TI_l|]$ Begin

// $|TI_l|$ represents the number of nests in level l of hierarchy TI

$(\{E_a\}_{ij} \subset \{ce_a\} \wedge \{ce_a\} \subset \{E_a\}_{ij})$ Begin

// Attributes representing the nest $\{E_f\}_{ij}$ must be identical to the attributes of the features in ce

$$\{t\} \leftarrow \{ce\} \cup \{E_f\}_{ij}$$

// forming a temporary set from the union of $\{ce\}$, which is a set with one element and $\{E_f\}_{ij}$ that is a set of eggs in the nest j of level l

If $|\{t\}| \equiv |\{E_f\}_{ij}|$ then $s_l += 1$

End

End

$s(E) \leftarrow s_l$

End

End

Further, the similar search process also be performed on TN to obtain the $s(F)$, which is a vector that repre-

sents the number of compatible nests in each level of hierarchy*TN*.

b) Cuckoo Search on Hierarchy TN.

Sort the $\{RN_f\}$ in descending order of sizes

$s(F) \leftarrow \phi$ // a vector represents the compatible nests at each level of the hierarchy *TN*

For each level $\{l \exists l = 1, 2, 3, \ldots m\}$ Begin

// *l* represents the level of the hierarchy from root to last level respectively

$s_l = 0$ // possible number of nests in level *l* accommodates cuckoo eggs

For each cuckoo egg $[ce \exists ce \in \{RN_f\} \wedge |ce| \equiv (m - l + 1)]$ Begin

// $(m - l + 1)$ represents the number of features in cuckoo egg *ce*

For each $[j \exists j = 1, 2, 3, \ldots |TN_l|]$ Begin

// $|TN_l|$ represents the number of nests in level *l* of hierarchy *TN*

$if(\{N_a\}_{lj} \subset \{ce_a\} \wedge \{ce_a\} \subset \{N_a\}_{lj})$ Begin

// Attributes representing the nest $\{N_f\}_{lj}$ must be identical to the attributes of the features in *ce*

$$\{t\} \leftarrow \{ce\} \cup \{N_f\}_{lj}$$

// forming a temporary set from the union of $\{ce\}$, which is a set with one element and $\{N_f\}_{lj}$ that is a set of eggs in the nest *j* of level *l*

If $||t| \equiv |\{N_f\}_{lj}|$ then $s_{l+} = 1$

End

End

$s(F) \leftarrow s_l$

End

End

2.10 Assessing the State of a Record of given Erythrocyte Features

For each level $\{l_{TI} \exists l_{TI} = 1, 2, 3, \ldots n\}$ of hierarchy *TI* Begin

$$cnr(l_{TI}) = \frac{cnc(l_{TI})}{nc(l_{TI})}$$

// $cnc(l_{TI})$ is compatible number of nests in level l_{TI}, $nc(l_{TI})$ is total number of nests in level l_{TI}

End

For each level $\{l_{TN} \exists l_{TN} = 1, 2, 3, \ldots m\}$ of hierarchy *TN* Begin

$$cnr(l_{TN}) = \frac{cnc(l_{TN})}{nc(l_{TN})}$$

// $cnc(l_{TN})$ is compatible number of nests in level l_{TN}, is total number of nests in level

End

For each level Begin

End

Begin

//is the infected ratio for each level of, is the infected ratio threshold given

Erythrocyte record is confirmed to be infected

Else

Erythrocyte record is confirmed to be Normal

End

3. Experimental Study and Results Analysis

3.1 The Dataset

A dataset is developed based on the samples gathered from varied cellular counterparts that are collected from varied diagnostics, using statistical guidelines essential for contributions in a medical journal[61]. Total sample size collected is 180 (48 P. falciparum, 84 P. vivax, and 48 normal) from the 800 samples evaluated (661 P. vivax, 67 normal, and 72 P. falciparum) collected. Selection of samples was carried out on the basis of visible clarity that could be gained from under microscopic tests. Labels that are essential for the samples were affirmed by five pathologists.

Results that are derived from study experiments reflect upon the performance graphs and the tables. Levels of prediction accuracy shall be stable for hamming distances within or equal to 0.50 towards a trained and tested images in Figure 7.

Also in terms of overhead and the resource utilization for SEMPS, there is optimal ratio level essential for time of completion pertaining to the input records that are envisaged as linear in Figure 8. Also the ration of memory usage towards linear levels is divergent to the number of records inspected seen in Figure 9.

The heuristic scale definition from CUCKOO search shall be as optimal for the prediction accuracy that is concrete in terms of results to the count of optimal features and the prediction accuracy estimated (87%), which is substantially an effective count compared to SVM based prediction (83.555)[38] and also Naïve Bayes prediction accuracy resulting (84%)[38].
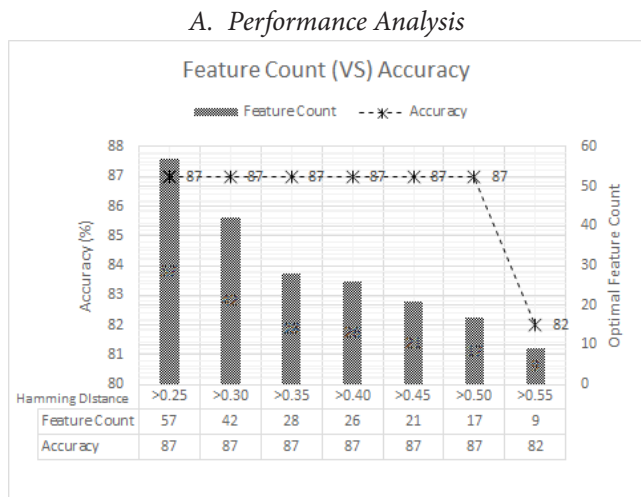
A. Performance Analysis

Figure 7. The accuracy observed for different number of features under varying hamming distance thresholds.
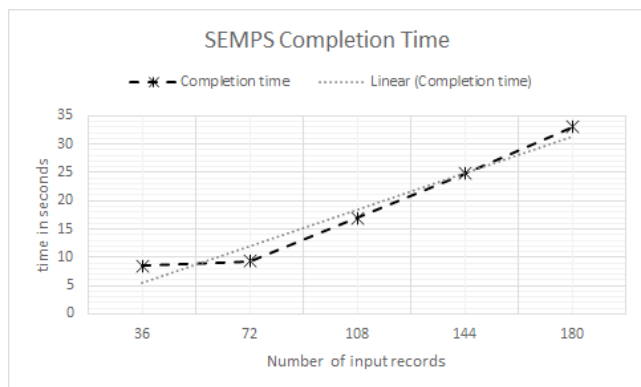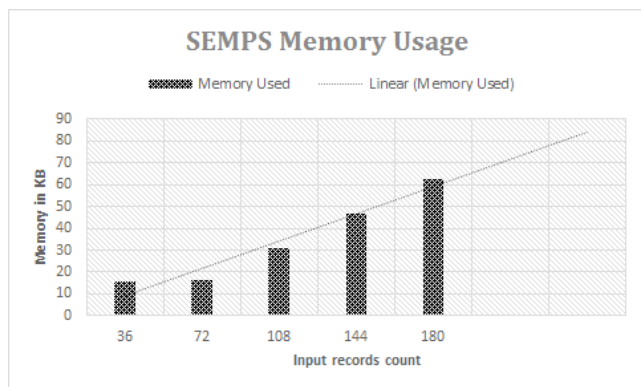

Figure 8. SEMPS process completion time vs linearity.

Figure 9. SEMPS memory usage vs linearity.

4. Conclusion

In the proposed model of Heuristic Scale to Estimate Premature Malaria Parasites Scope (SEMPS) with multi

stage processing of the microscopic images of blood smear ismuch effective than the other models that were depicted earlier. The process of image processing and the heuristic scale evaluation is carried out using multiple phases in order to ensure that the levels of accuracy and the process outcome shall be more effective. Some of the key factors considered in the process are segmentation, optimal feature selection, feature extraction and evolutionary compuation that can be adapted on the baiss of heurisitc scale definition.

Some of the critical constraints that are envisaged in the benchmarking models are about the dullness in contrast, intensities observed in similar fashion for both effected and normal areas of blood smear images for microscopic image segmentation modeling, edge formation issues and other such factors are turning out to be a major impact, from the earlier processes.

With the proposed model of CUCKOO search the outcome in terms of evaluating the process using the image comparisons and the optimal feature utilization has been resulting effective outcome. Despite the fact that there significant stages involved in the process and the adaptation of various benchmarking methods like SVM and Bayesian model are considered in the experimental study, the results that are envisaged from the process signify that the proposed model of SEMPS shall be very resourceful in premature stage detection of malaria from the microscopic images. Also, the study has depicted scope towards using varied directions like the identification of correlation of features and the impact of such features on the scale definition, and the scope for using the genetic algorithm for identifying the optimal features.

5. References

1. Rougemont MVS. Detection of four Plasmodium species in blood from humans by 18S rRNA gene subunit-based and species-specific real-time PCR assays. Journal of Clinical Microbiology. 2004; 42(12):5636–43. PMid:15583293 PMCid:PMC535226. https://doi.org/10.1128/JCM.42.12.5636-5643.2004

2. Florens LW. A proteomic view of the Plasmodium falciparum life cycle. Nature. 2002; 419(6106):520–6. PMid:12368866. Available from: https://doi.org/10.1038/nature01107

3. Pain AB. The genome of the simian and human malaria parasite Plasmodium knowlesi. Nature. 2008;455(7214):799–803. PMid:18843368 PMCid:PMC2656934. Available from: https://doi.org/10.1038/nature07306

4. Snow RW. The global distribution of clinical episodes of Plasmodium falciparum malaria. Nature. 2005; 434(7030):214–17. PMid:15759000 PMCid:PMC3128492. Available from: https://doi.org/10.1038/nature03342

5. WHO WH. Guidelines for the treatment of malaria. World Health Organization; 2006.

6. Reyburn H. New WHO guidelines for the treatment of malaria. BMJ. 2010. PMCid: PMC2847687. Available from: https://doi.org/10.1136/bmj.c2637

7. Hu MK. Visual pattern recognition by moment invariants. IRE Transactions on Information Theory. 1962; 8(2):179–87. Available from: https://doi.org/10.1109/TIT.1962.1057692

8. Galloway MM. Texture analysis using gray level run lengths. Computer graphics and image processing. 1975; 4(2):172–9. Available from: https://doi.org/10.1016/S0146-664X(75)80008-6

9. Mandelbrot B. The fractal geometry of nature/Revised and enlarged edition. New York: WH Freeman and Co; 1983.

10. Chu AS. Use of gray value distribution of run lengths for texture analysis. Pattern Recognition Letters. 1990; 11(6):415–9. Available from: https://doi.org/10.1016/0167-8655(90)90112-F

11. Dasarathy BV. Image characterizations based on joint gray level-run length distributions. Pattern Recognition Letters. 1991; 12(8):497–502. Available from: https://doi.org/10.1016/0167-8655(91)80014-2

12. Sarkar N. An efficient differential box-counting approach to compute fractal dimension of image. IEEE Transactions on Systems, Man and Cybernetics. 1994; 24(1):115–20. Available from: https://doi.org/10.1109/21.259692

13. Ojala TP. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002; 24(7):971–87. Available from: https://doi.org/10.1109/TPAMI.2002.1017623

14. Gonzalez RC. Digital image processing. Prentice Hall; 2002.

15. Pharwaha AP. Hannon and non-shannon measures of entropy for statistical texture feature extraction in digitized mammograms. World Congress on Engineering and Computer Science. 2002; 2:20–2.

16. Ghosh MD. Entropy based divergence for leukocyte. IEEE International Conference on Systems in Medicine and Biology; 2010. p. 409–13.

17. Krishnan MM. Textural characterization of histopathological images for oral sub-mucous fibrosis detection. Tissue and Cell. 2011; 43(5):318–30. PMid: 21824635. Available from: https://doi.org/10.1016/j.tice.2011.06.005

18. Krishnan MM. Statistical analysis of textural features for improved classification of oral histopathological images. Journal of Medical Systems. 2012; 36(2):865–81. PMid: 20703647. Available from: https://doi.org/10.1007/s10916-010-9550-8

19. Celebi ME. Lesion border detection in dermoscopy images. Computerized Medical Imaging and Graphics. 2009; 33(2):148–53. PMid: 19121917 PMCid: PMC2671195. Available from: https://doi.org/10.1016/j.compmedimag.2008.11.002

20. Yang XS. Cuckoo search via Lévy flights. World Congress on Nature and Biologically Inspired Computing; 2009. p. 210–4.

21. Ross NE. Automated image processing method for the diagnosis and classification of malaria on thin blood smears. Medical and Biological Engineering and Computing. 2006; 44(5):427–36. PMid: 16937184. Available from: https://doi.org/10.1007/s11517-006-0044-2

22. Kaewkamnerd SU. An automatic device for detection and classification of malaria parasite species in thick blood film. BMC Bioinformatics. 2012; 13(17):1. Available from: https://doi.org/10.1186/1471-2105-13-s17-s18

23. Díaz GG. A semi automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images. Journal of Biomedical Informatics. 2009; 42(2):296–307. PMid:19166974. Available from: https://doi.org/10.1016/j.jbi.2008.11.005

24. Prasad KW. Image analysis approach for development of a decision support system for detection of malaria parasites in thin blood smear images. Journal of Digital Imaging. 2012; 25(4):542–9. PMid: 22146834 PMCid: PMC3389088. Available from: https://doi.org/10.1007/s10278-011-9442-6

25. Frean JA. Reliable enumeration of malaria parasites in thick blood films using digital image analysis. Malaria Journal. 2009; 8(218). Available from: https://doi.org/10.1186/1475-2875-8-218

26. Ching-Hao YSYHL. A protozoan parasite extraction scheme for digital microscopic images. Comput Med Image Graph. 2010; 122–30.

27. Le MT. A novel semi-automatic image processing approach to determine Plasmodium falciparum parasitemia in giemsa-stained thin blood smears. BMC Cell Biology. 2008; 9(15). Available from: https://doi.org/10.1186/1471-2121-9-15

28. Díaz GG. Infected cell identification in thin blood images based on color pixel classification: Comparison and analysis. Pattern Recognition, Image Analysis and Applications. 2007; 812–21.

29. Tek FB. Computer vision for microscopy diagnosis of malaria. Malaria Journal. 2009; 8(1):1–14. PMid: 19594927 PMCid: PMC2719653. Available from: https://doi.org/10.1186/1475-2875-8-153

30. Tek FB. Parasite detection and identification for automated thin blood film malaria diagnosis. Computer Vision and

Image Understanding. 2010; 114(1):21–32. Available from: https://doi.org/10.1016/j.cviu.2009.08.003

31. Memeu DM. Detection of plasmodium parasites from images of thin blood smears. Open Journal of Clinical Diagnostics. 2013; 3(1):183–94. Available from: https://doi.org/10.4236/ojcd.2013.34034

32. Yunda LA. Automated image analysis method for p-vivax malaria parasite detection in thick film blood images. Sistemas y Telemática. 2012; 10(20):9–25. Available from: https://doi.org/10.18046/syt.v10i20.1151

33. Sio SW. An image analysis-based program for the accurate determination of parasitemia. Journal of Microbiological Methods. 2007; 68(1):11–18. PMid:16837087. Available from: https://doi.org/10.1016/j.mimet.2006.05.017

34. Tek FB. Malaria parasite detection in peripheral blood images. British Machine Vision Conference BMVC; 2006. p. 347–56. Available from: https://doi.org/10.5244/c.20.36

35. Makkapati VV. Segmentation of malaria parasites in peripheral blood smear images. IEEE International Conference on Acoustics, Speech and Signal Processing; 2009. p. 1361–4. Available from: https://doi.org/10.1109/icassp.2009.4959845

36. Purwar YS. Automated and unsupervised detection of malarial parasites in microscopic images. Malaria Journal. 2011; 10(1):1–11. PMid: 22165867 PMCid: PMC3254597. Available from: https://doi.org/10.1186/1475-2875-10-364

37. Somasekar JR. Segmentation of erythrocytes infected with malaria parasites for the diagnosis using microscopy imaging. Computers and Electrical Engineering. 2015; 45(1):336–51. Available from: https://doi.org/10.1016/j.compeleceng.2015.04.009

38. Das DK. Machine learning approach for automated screening of malaria parasite using light microscopic images. Micron. 2013; 45:97–106. PMid: 23218914. Available from: https://doi.org/10.1016/j.micron.2012.11.002

39. Khan MI. Content based image retrieval approaches for detection of malarial parasite in blood images. International Journal of Biometrics and Bioinformatics. 2011; 5(2):97–110.

40. Hearst MA. Support vector machines. Intelligent Systems and their Applications. 1998; 13(4):18–28. Available from: https://doi.org/10.1109/5254.708428

41. Langley P. Induction of selective Bayesian classifiers. 10th International Conference on Uncertainty in Artificial Intelligence Morgan Kaufmann Publishers Inc; 1994. p. 399–406. Available from: https://doi.org/10.1016/b978-1-55860-332-5.50055-9

42. Tabachnick BG. Using multivariate statistics. PEARSON; 2001. PMid:16705786

43. Iwaki Y. U. S Patent No. 8; 2014. p. 861–78.

44. Kanan C. Color-to-grayscale: Does the method matter in image recognition. PloS One; 2012. Available from: https://doi.org/10.1371/journal.pone.0029740

45. John JR. Introduction to image processing and analysis. CRC Press; 2008.

46. Abdul-Nasir MM. Colour Image Segmentation Approach for Detection of Malaria Parasites. WSEAS Transactions on Biology and Biomedicine. 2013; 10:41–55.

47. Yeon J, Kim JD, Kim YS, Park CY, Song HY. Effective grayscale conversion method for malaria parasite detection. Advanced Science and Technology Letters. 2014; 78(34):77–81. Available from: https://doi.org/10.14257/astl.2014.78.15

48. Kim JD. Comparison of grayscale conversion methods for malaria classification. International Journal of Bio-Science and Bio-Technology. 2015; 7(1):141–50. Available from: https://doi.org/10.14257/ijbsbt.2015.7.1.14

49. Lai CH. A protozoan parasite extraction scheme for digital microscopic images. Computerized Medical Imaging and Graphics. 2010; 34(2):122–30. PMid: 19699610. Available from: https://doi.org/10.1016/j.compmedimag.2009.07.008

50. Wei JWZ. A median-Gaussian filtering framework for Moiré pattern noise removal from X-ray microscopy image. Micron. 2012; 43(2-3):170–6. PMid: 21803588 PMCid: PMC3858302. Available from: https://doi.org/10.1016/j.micron.2011.07.009

51. Chokkalingam KA. Performance analysis of various lymphocytes images de-noising filters over a microscopic blood smear image. International Journal of Pharma and Bio Sciences. 2013; 4(4):1250–8.

52. Astola JK. Fundamentals of non-linear digital filtering. New Yark: CRC Press; 1997.

53. MathWorks. medfilt2. Available from: http://www.mathworks.com/help/toolbox/images/ref/medfilt2.html

54. Aizenberg IB. New method for impulsive noise filtering using its preliminary detection. SPIE Proceedings of Image Processing: Algorithms and System; 2002. p. 204–14.

55. Kingsbury N. Wavelet transforms in image processing. Signal Analysis and Prediction. 1998; 27–46. Available from: https://doi.org/10.1007/978-1-4612-1768-8_2

56. Gonzalez RE. Digital image processing using MATLAB. Pearson Prentice Hall Pearson Education; 2004.

57. Hartigan JA. Algorithm AS 136: A k-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics). 1979; 100–8. Available from: https://doi.org/10.2307/2346830

58. Christ MJ. Segmentation of medical image using K-Means clustering and marker controlled watershed algorithm. European Journal of Scientific Research. 2012; 71(2):190–4.

59. Krishnan MM. Statistical analysis of textural features for improved classification of oral histopathological images. Journal of Medical Systems. 2012; 36(2):865–81. PMid: 20703647. Available from: https://doi.org/10.1007/s10916-010-9550-8

60. Das DG. Invariant moment based feature analysis for abnormal erythrocyte recognition. IEEE International Conference on Systems in Medicine and Biology; 2010. p. 242–7. Available from: https://doi.org/10.1109/icsmb.2010.5735380

61. Altman DG. Statistical guidelines for contributors to medical journals. Br Med J (Clin Res Ed). 1983; 1489–93. PMid: 6405856 PMCid: PMC1547706.