Quantitative characterisation of *Plasmodium vivax* in infected erythrocytes: a textural approach

Madhumala Ghosh, Devkumar Das and Chandan Chakraborty*

School of Medical Science and Technology, Indian Institute of Technology, Kharagpur – 721302, West Bengal, India
E-mail: madhumala.ghosh@gmail.com
E-mail: dev.biomedical@gmail.com
E-mail: chandanc@smst.iitkgp.ernet.in
*Corresponding author

Ajoy Kumar Ray

Electronics and Electrical Communication Engineering Department, Indian Institute of Technology, Kharagpur – 721302, West Bengal, India and
Bengal Engineering and Science University, Botanic Garden, Howrah – 711103, West Bengal, India
E-mail: ajoy_ray2004@yahoo.com

Abstract: This paper aims at introducing a textural pattern analysis approach to *Plasmodium vivax* (*P. vivax*) detection from Leishman stained thin blood film. This scheme follows retrospective study design protocol where patients were selected at random in the clinic. The scheme consists of four stages – artefacts reduction, fuzzy divergence-based segmentation of *P. vivax* infected region(s) and normal erythrocytes, textural feature extraction using grey level co-occurrence matrix and fractal dimension, finally classification. Here, we have extracted seven features, out of which five are statistically significant in discriminating textures between malaria and normal classes based on light microscopic blood images at 100× resolutions. Finally, Bayesian and support vector machine-based classifiers are trained and validated with 100 cases and 100 control subjects. In effect, it is hereby observed that the significant textural features lead to discriminate *P. vivax* with 95% and 98% accuracies for SVM and Bayesian classifiers respectively. Results are studied and compared.

Keywords: parasitaemia; erythrocyte; *Plasmodium vivax*; fuzzy divergence; SVM; Bayesian; sensitivity; specificity.


Biographical notes: Madhumala Ghosh received her BTech in Electronics and Telecommunication Engineering from WBUT, West Bengal, India in the year 2005. Then she received her MTech in Intelligent Automation and Robotics.
from Jadavpur University, West Bengal, India in 2008. She is now pursuing her PhD from School of Medical Science and Technology, IIT, Kharagpur in the field of pathological image processing.

Dev Kumar Das received his BTech in Biomedical Engineering from WBUT, West Bengal, India, in 2008. He is currently an MS student at the School of Medical Science and Technology, Indian Institute of Technology, Kharagpur, India. His research interests include pathological imaging and pathological image analysis.

Chandan Chakraborty is currently an Assistant Professor of School of Medical Science and Technology, IIT Kharagpur. He completed his BSc and MSc in Applied Statistics and PhD in Statistical Pattern Recognition. He has been awarded by President of India for his work in 2007. He is an IEEE member. He has authored over 80 research papers in refereed journals, books, and conference proceedings. His current research interests include pattern recognition for medical imaging, biostatistics, and medical image analytics and informatics.

Ajoy Kumar Ray is connected with Bengal Engineering and Science University, Shibpur as Vice Chancellor. Earlier to this he was a Professor of Electronics and Electrical Communication Engineering and Head, School of Medical Science and Technology at IIT Kharagpur. He is co-inventor of six US patents jointly with Intel Corporation.

1 Introduction

Malaria is one of the most serious worldwide health problems causes 1.5–2.7 million of deaths per year (Raviraja et al., 2008). The worldwide annual economic trouble of malaria, including spend on prevention and treatment as well as loss in efficiency due to sickness, is estimated at US$800 million in 1995 (Raviraja et al., 2008). In India, approximately 30% to 40% of people are affected by *Plasmodium vivax* that is reported in National Vector Borne Disease Control Program (NVBDCP) data in the year of 2010. So, fast and accurate diagnosis is required to facilitate prompt treatment to control malaria. In today’s diagnostic paradigm, microscopic imaging technology has immense contributions in generating fruitful medical images, which essentially become the basis for medical experts to make better decisions. In practice, pathologists visualise the abnormalities if any, in the images through microscope based on their knowledge from the view point of intensity, morphology, texture, etc. based features. Usually, small-scale differences in the features are overlooked by human eyes especially for the border-line diagnostic scenario.

In order to circumvent this problem, it is more worthwhile to develop computer-assisted automated screening scheme for automatically characterising the abnormalities, especially in complicated cases where experts fail to take decision. In doing this, microscopic information needs to be analysed quantitatively maintaining the biological integrity in the system. Here, an attempt has been made to develop an automated pattern analyser to detect the malaria parasite (*Plasmodium vivax*) which helps to reduce the time efficiency as well as inter and intra-observer variations.
Some works are reported in the literatures. Tek et al. (2006) introduce automatic detection of malaria parasite based on colour histogram. Diaz et al. (2009) show the quantification and classification of *Plasmodium falciparum* infected erythrocyte. Ross et al. (2006) used the morphological and novel thresholding selection technique for identify erythrocyte and possible parasite. Makkapati and Rao (2009) segment the malaria parasite in HSV colour space. Raviraja et al. (2008) detect the red blood cells that are infected by malaria parasites using statistical-based approach. Toha and Ngah (2007) segment the malaria parasite using grey level thresholding. Garcia (2004) approaches automatic malaria infected cell counting. Halim et al. (2006) estimate parasitemia based on pattern matching with parameter optimisation and cross-validation against the expected biological characteristics, red blood cells are determined. This paper aims at introducing an automated approach to parasitemia detection from ‘Leishman’ stained thin blood film infected with *P. vivax* using fuzzy divergence-based thresholding techniques. The proposed scheme is a combination of four different stages: a preprocessing step for correcting the noise, a segmentation step which is done by threshold selection using fuzzy divergence, textural feature extraction and classification which identifies normal and infected erythrocytes from their textural features.

2 Material and methods

2.1 Human blood smear preparation

Retrospective study design is here followed in selecting the subjects from the population at the Department of Hematology, Midnapur Medical College and Hospital and Medipath Laboratory, Midnapur, West Bengal. The blood samples from there malaria infected patients are collected and processed as follows. Thin blood smear is prepared on a clean and disinfected slide and stained with Leishman for visualising different cellular counterparts. In the laboratory, firstly 5–7 drops Leishman is applied to the slide with the specimen. After 5 minutes, 10–12 drops of a buffer solution (pH 6.8) is added and mixed with the stain, then the specimen is left staying for 20–30 minutes, then washed off with the distilled water and dry the washed slide.

2.2 Microscopic imaging and acquisition

The images are optically grabbed by Leica Observer (Leica DM750, Leica Microsystems (Switzerland) limited) under 100× oil objective (NA 1.515) having 0.064 μm resolution and stored in the database server through the web-based interface.

2.3 Preprocessing

The input images are the microphotographs of Leishman’s stained slides from malaria blood smears. For study, 150 image frames of size (2,048 × 1,536) are taken. The grabbed image is converted to CMYK space out of which C channel is selected and filtered with median filter to reduce the noise. CMYK colour space is chosen because the segmentation of parasite gives better results in that domain.
2.4 Fuzzy divergence approach to parasite segmentation

In pathological evaluation, malaria parasite is usually recognised by the experts because of significant textural variation of the infected erythrocytes. Towards this, segmentation of parasite from infected erythrocytes is necessary. However, the selection of the threshold is very difficult in segmenting the parasite. As a whole blood smear image consists of erythrocytes, leukocytes and platelets. In this paper, our focus is to segment the malaria parasite of *Plasmodium vivax* from not only the background but from infected erythrocyte also using fuzzy divergence minimisation approach based on Cauchy membership function. Our approach is one such approach that highlights as a sound scheme to segment parasite distinctly and provides a comparison among two different stages of infection of malaria parasite, i.e., trophozoites and schizonts, gametocyte. There are various ways of choosing the threshold of an image. Basically, thresholding processes are based on the image histogram (Gonzalez and Woods, 2009). The region between two successive peaks is the region for searching the minimum value which will be set as the threshold point for thresholding the object from background. If there are more valleys (multilevel histogram) succeeding and preceding peaks of each valley are noted and accordingly the search regions are selected. And hence for unimodal case linear search is employed. Thresholding of image can be done in two ways viz. fuzzy threshold selection (Chaira and Ray, 2004) and non-fuzzy threshold selection. In case of crisp image, the pixels are precise and regions are well defined. But while threshold is selected from the histogram, deep valleys of histogram cannot be located properly. Thus, object may not be separated appropriately. But in fuzzy image this problem can be overcome during thresholding.
2.5 Fuzzy image and its membership function

A grey image is a combination of grey values. Let an image $A$ of size $M \times N$ and having $L$ grey levels, be defined as a combination of grey levels and its membership value $\mu(f_{ij})$ of $(i,j)^{th}$ pixel,

$$A = \{f_{ij}, \mu(f_{ij})\}, \forall f_{ij} \in A$$

(1)

where $0 \leq \mu(f_{ij}) \leq 1$. The grade of membership $\mu(f_{ij})$ maps the pixel grey level $f_{ij}$ to positive real numbers in the interval $[0, 1]$. In an image, the count of a particular grey level $f$ denotes the number of occurrences in the image. For a particular given threshold value $t$, the object and background is separated distinctly. The average grey level of background and object region are given by the following relation,

$$\mu_0 = \frac{\sum_{f=0}^{t} f \cdot \text{count}(f)}{\sum_{f=0}^{L-1} \text{count}(f)}, \quad \mu_1 = \frac{\sum_{f=t+1}^{L-1} f \cdot \text{count}(f)}{\sum_{f=t+1}^{L-1} \text{count}(f)}$$

(2)

The membership values of pixels have been determined using Cauchy distribution which is described below.

2.6 Cauchy distribution

Cauchy distribution is defined as

$$f(x, x_0, \gamma) = \frac{1}{\pi \gamma \left[1 + \left(\frac{x - x_0}{\gamma}\right)^2\right]}, \quad x > x_0, \gamma > 0.$$  

(3)

where $\gamma$ is the scale parameter, $x_0$ is the location parameter. When $\gamma = 1$, the membership function $\mu(f_{ij})$ of the object region may be computed as,

$$\mu(f_{ij}) = B \left[\frac{1}{1 + \left(f_{ij} - \mu_1\right)^2}\right]$$

(4)

where $\mu_1 =$average grey level of object region when $f_{ij} > t$.

In a similar way, membership function $\mu(f_{ij})$ of the background region may be computed as,

$$\mu(f_{ij}) = B \left[\frac{1}{1 + \left(f_{ij} - \mu_0\right)^2}\right]$$

(5)

where $\mu_0 =$average grey level of background region when $f_{ij} \leq t$. Here, ‘$B$’ is chosen as $c / \pi$. 
Where $\mu_0$ indicates average grey level of background region when $f_{ij} \leq t$. It may be pointed out that in the membership function, the constant ‘c’ has been taken to ensure membership value of the grey level feasible in the range [0, 1]. Here, ‘c’ may be chosen as, $c = 1 / (f_{\text{max}} - f_{\text{min}})$, where $f_{\text{min}}$ and $f_{\text{max}}$ are the minimum and maximum grey level in the test image respectively. For computing the membership function, the absolute value of the distance between the mean of the region to which a pixel belongs and the grey level of that pixel has been considered.

2.7 Fuzzy divergence method

Fuzzy divergence was proposed by Fan and Xie (1999) where they mentioned the fuzzy exponential entropy by using a single row vector. Chaira and Ray (2003) extended that concept of Fan and Xie (1999) to an image which may be represented by a matrix for image thresholding. For two images, $A$ and $B$ of size $M \times M$ with ‘$L$’ distinct grey levels, the amount of information of $\mu_A(a_{ij})$ and $\mu_B(b_{ij})$ of images $A$ and $B$ at a particular point $(i,j)$ in the image, is depicted as,

$$\frac{\mu_A(a_{ij})}{\mu_B(b_{ij})} = e^{\mu_A(a_{ij})-\mu_B(b_{ij})}$$

(6)

where $\mu_A(a_{ij})$ and $\mu_B(b_{ij})$ are the membership values of the $(i, j)^{th}$ pixels in the images $A$ and $B$, respectively. $i, j = 0, 1, 2, \ldots, (M - 1)$. The discrimination of image $A$ against image $B$ is defined as

$$D_I(A, B) = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} \left[1 - \left(1 - \mu_A(a_{ij})\right)e^{\mu_A(a_{ij})-\mu_B(b_{ij})} - \mu_A(a_{ij})e^{\mu_A(a_{ij})-\mu_B(b_{ij})}\right]$$

(7)
Quantitative characterisation of Plasmodium vivax in infected erythrocytes

In a similar manner, the discrimination of image $B$ against image $A$ is

$$D_2(A, B) = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} \left[ 1 - (1 - \mu_A(b_{ij}))e^{\mu_A(b_{ij}) - \mu_B(a_{ij})} - \mu_B(b_{ij})e^{\mu_A(a_{ij}) - \mu_B(b_{ij})} \right]$$

(8)

Therefore, the total fuzzy divergence between image $A$ and $B$ is derived from equations (15) and (16),

$$D(A, B) = D_1(A, B) + D_2(A, B)$$

(9)

i.e.,

$$D(A, B) = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} \left[ 2 - (1 - \mu_A(a_{ij}) + \mu_B(b_{ij}))e^{\mu_A(a_{ij}) - \mu_B(b_{ij})} \right]$$

(10)

where $i, j = 0, 1, 2, \ldots, (M - 1), \mu_A(a_{ij})$ and $\mu_B(b_{ij})$ denote the membership value of the pixel where $a_{ij}$ and $b_{ij}$ are the $(i, j)^{th}$ pixel of the image $A$ and $B$.

2.8 Threshold selection

For thresholding, the membership value is calculated using equation (10). The membership values of the threshold image are compared with an ideally threshold image. An ideally threshold image is defined as the image whose object and background regions should totally belong to the respective regions. In such situation, the membership values for ideally threshold image of each pixel belong to the object/background region should be equal to one. Hence the above equation (10) becomes,

$$D(A, B) = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} \left[ 2 - (2 - \mu_A(a_{ij}))e^{\mu_A(a_{ij}) - \mu_A(a_{ij})} - \mu_A(a_{ij})e^{1 - \mu_A(a_{ij})} \right]$$

(11)

The membership value of test image $A$ should lie closer to that of the ideally threshold image $B$ for better thresholding. When the pixel presents in the object/background region, the membership should contribute more to the corresponding object/background region. Henceforth, in that way, the divergence value of each pixel is calculated for whole image and corresponding grey level is noted. The grey value corresponding to the minimum divergence is chosen as threshold initially for segmenting the object (parasite) and background (rest of the image) regions. In fact, the minimum divergence value indicates the maximum belongingness of each object pixel to the object region and each background pixel to the background region. Instead, we revise the threshold obtained at initial stage by taking 5.5 times of the grey value to the same. The segmented results are visualised in the result section. The proposed scheme is pictorially described in the following block diagram (Figure 3).
2.9 Textural feature generation

The most important part of pattern recognition in image processing is feature extraction. When the input data to an algorithm is too large to be processed and it is suspected to be redundant then the input data will be transformed into a reduced representation set of features. And a feature is a set of data or parameter that redefines the input. Transforming the input data into the set of features is called feature extraction. If the extracted features are chosen carefully then it is expected that the features set will extract the relevant information from the input data in order to perform the desired classification task using this reduced representation instead of the full size input. In this approach, six different textural features are extracted from the segmented infected erythrocyte and from normal erythrocyte. The features are average intensity, average contrast, smoothness, skewness, uniformity and entropy (Gonzalez and Woods, 2009). Another textural feature considered
here is fractal dimension which helps to quantify the complexity and irregularity of these objects (Pavlopoulos et al., 2004).

If we consider the grey scale profile as the third dimension along with two dimensions of image, the variation in this profile gives the roughness or textural changes of the virtual surface consisting of the epithelium. This can be estimated by fractal dimension. Any surface ‘A’ in Euclidean n-space is self-similar or self-affine if A is the union of N_r distinct (non-overlapping) copies of itself scaled up or down by a factor of r. Mathematically fractal dimension is defined by

\[ D = \frac{\log N_r}{\log(1/r)} \]  

There are different approaches to find fractal dimension. One of them is differential box counting (DBC). Straightforward implementation of this requires O(N^2) time complexity. So we have considered modified DBC with sequential algorithm (Mandelbrot, 1982; Biswas et al., 1998). Sequential algorithm takes grey scale image as its input and the grid size is in the power of 2 for computation efficiently. The initial grid size (s) is 2 and finds the maximum and minimum intensity for each box (2 × 2). The summation of difference between maximum and minimum intensities gives the N and r can be found by

\[ r = \frac{S}{M} \]  

where M is the minimum size of the image. Grid size increases to double and maximum size reduces to half and above procedure repeats until the maximum size remains larger than 2. Linear regression model to fit the line from plot \( \log(N) \rightarrow \log(1/r) \) and the slope gives the fractal dimension as mentioned below which is used to analyse the texture of the epithelium.

\[ \log N_r = D\log(1/r) \]  

2.10 Statistical analysis

Box-Whisker’s plot (Rastogi, 2008) (also known as a box-and-whisker diagram) is a convenient way of graphically depicting groups of numerical data through five statistical means: the smallest observation (sample minimum), lower quartile, median, upper quartile, and largest observation (sample maximum). It may also indicate which observations, if any, might be considered outliers. It is useful to display differences between populations without making any assumptions of the underlying statistical distribution, i.e., they are non-parametric. The spacings between the different parts of the box indicate the degree of dispersion (spread) and skewness in the data, and identify outliers. Through box plots the variability between the median of samples can be compared.

Kernel density (KS) (Duda et al., 2007) computes a probability density estimate of the sample in the chosen textural features (e.g., avg. intensity, contrast, skewness, fractal, etc.). And compute a function vector of density values evaluated at the points in specified feature vector. The estimate is based on a normal kernel function, using a window
parameter (‘width’) that is a function of the number of points in \( x \). The density is evaluated at 100 equally spaced points that cover the range of the data in feature vector.

Another statistical approach is \( t \) test analysis which is used to compare means of a particular feature between the groups, i.e., \( n_1 \) is the number of sample of normal erythrocyte and \( n_2 \) is the number of infected erythrocytes. The paired \( t \)-test is actually a test which shows whether the difference between these two classes is 0 or not. Here, the degrees of freedom is \( n - 1 \) (where \( n = n_1 + n_2 \)). The test statistic is \( t \) with \( n - 1 \) degrees of freedom. If the \( p \)-value associated with \( t \) is low (< 0.05), there is evidence to reject the null hypothesis. Thus, we should have evidence that there is a difference in means across the paired classes of samples.

### 2.11 Parasitaemia detection

The parasitemia detection has become eventually a two class pattern classification problem for the classes viz. healthy and malaria infected erythrocytes. In this study, we have been attempted two well established statistical learning techniques, Bayes’ classifier and support vector machines to differentiate normal and infected erythrocytes.

#### 2.11.1 Bayes’ classifier

A Bayes classifier is a simple probabilistic classifier based on applying Bayes’ theorem (from Bayesian statistics) with strong (naive) independence assumptions. Bayes classifier express that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature and classifier efficiency depends upon all properties of that particular feature which are independently contribute to the probability. This classifier is trained with supervised learning and its parameter estimation can be based on maximum likelihood scheme. Given a set of variables, \( X = \{x_1, x_2, \ldots, x_d\} \), we want to constructing the posterior probability for the event \( C_j \) of \( j \)th class, among a set of possible classes \( C = \{C_1, C_2\} \). In a more familiar language, \( X \) is the predictors and \( C \) is the set of categorical levels present in the dependent variable. Using Bayes rule,

\[
p(C_j | x_1, x_2, \ldots, x_d) \propto p(x_1, x_2, \ldots, x_d | C_j) p(C_j)
\]

where \( p(C_j | x_1, x_2, \ldots, x_d) \), is the posterior probability of class membership, i.e., the probability that \( x \) belongs to \( C_j \). Since Naive Bayes assumes that the conditional probabilities of the independent variables are statistically independent we can decompose the likelihood to a product of terms,

\[
p(X | C_j) \propto \prod_{k=1}^{2} p(x_k | C_j)
\]

and rewrite the posterior as,

\[
p(C_j | X) \propto p(C_j) \prod_{k=1}^{2} p(x_k | C_j)
\]

Using Bayes rule above, we label a new case \( X \) with a class level \( C_j \) that achieves the highest posterior probability.
2.11.2 SVM classifier

In this study, the diagnostic problem is designed based on four textural features (average intensity, skewness, uniformity and entropy) and another feature, i.e., fractal calculation for normal and infected erythrocyte classification, which may be treated as a two-class pattern classification problem. In fact, these five features lead to a pattern by which normal and infected erythrocytes are classified using appropriate SVM. Let us denote a feature vector (termed as pattern) by \( \mathbf{x} = (x_1, x_2) \) where \( x_1 \) and \( x_2 \) represent here eccentricity and compactness respectively; and its class label by \( y \) such that \( y = \{1, -1\} \). Therefore, consider the problem of separating the set of \( n \) training patterns belonging to two classes

\[
(x_i, y_i), \quad x_i \in \mathbb{R}^n, \quad y_i = \{1, -1\}, \quad i = 1, 2, \ldots, n. \tag{18}
\]

A decision function \( g(x) \) that can correctly classify an input pattern \( x \) that is not necessarily from the training set. A linear SVM is used to classify datasets which are linearly separable. The SVM linear classifier tries to maximise the margin between the separating hyperplane. The patterns lying on the maximal margins are called support vectors (see Figure 3). Such a hyperplane with maximum margin is called maximum margin hyperplane (Duda et al., 2007). In case of linear SVM, the discriminant function is of the form

\[
g(x) = w^T x + b \tag{19}
\]

such that \( g(x) \geq 0 \) for \( y_i = +1 \) and \( g(x) < 0 \) for \( y_i = -1 \).

To solve quadratic optimisation problem one must find the saddle point of the Lagrange function

\[
L_p(w, b, \alpha) = \frac{1}{2} \| w \|^2 - \sum_{i=1}^n \alpha_i \{ y_i (w^T x_i + b) - 1 \}
\]

where the \( \alpha_i \) denotes Lagrange multipliers, hence \( \alpha_i > 0 \). The search for an optimal saddle point is necessary because \( L_p \) must be minimised with respect to the primal variables \( w \) and \( b \) and maximised with respect to the dual variable \( \alpha \). By differentiating with respect to \( w \) and \( b \) and introducing the Karush-Kuhn-Tucker (KKT) (Huang et al., 2008) conditions for the optimum constrained function, then is transformed to the dual Lagrange \( L_D(\alpha) \)

\[
\begin{align*}
L_D(\alpha) &= \left[ \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i^T x_j \right] \\
\text{subject to} \quad &\alpha_i \geq 0, \quad i = 1, 2, \ldots, n \\
&\sum_{i=1}^n \alpha_i y_i = 0
\end{align*}
\tag{21}
\]

2.12 Performance measure

The performance of the classifier is measured based on the classification accuracy, sensitivity and specificity. These measures are calculated using a matrix called confusion matrix defined in Table 1.
Table 1: Confusion matrix

<table>
<thead>
<tr>
<th>Actual</th>
<th>Predicted</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive</td>
<td>Negative</td>
</tr>
<tr>
<td>Positive</td>
<td>True positive (TP)</td>
<td>False negative (FN)</td>
</tr>
<tr>
<td>Negative</td>
<td>False positive (FP)</td>
<td>True negative (TN)</td>
</tr>
</tbody>
</table>

The performance measures of the classifier are defined as

\[
\text{Sensitivity} = \frac{TP}{TP + FN} \times 100\% \tag{22}
\]

\[
\text{Specificity} = \frac{TN}{TP + TN} \times 100\% \tag{23}
\]

\[
\text{Classification accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \tag{24}
\]

### 3 Results and discussion

Statistical analysis directs us to formulate the proper conclusions based on the collected feature data space. Basically, statistical analysis conducts us with logical accuracy towards the decision making in classification approach. Here, KS estimation and box plot approach are introduced to guide our decision in characterising normal and infected erythrocytes based on the textural features obtained during process. \(t\)-test analysis proves that there is a significant difference in normal and infected erythrocytes. A general description about the textural features is described in the following Table 1. During the scheme design, we have implemented different colour space united with fuzzy divergence-based approach in segmenting the malaria parasite from the infected erythrocytes. A precise comparative study among various colour space-based segmentation is given in Figure 4.

It is observed from the box whisker’s plot of Figure 5 that there is a significant variation in between the normal and infected erythrocytes in cases of average intensity, skewness, uniformity, entropy and fractal.

In KS density plot, the variation in distribution in the textural features for normal and infected erythrocytes are tested to visualise whether there is any overlapping in data between any pair. It is observed from the following KS density plots (Figure 6) that there is a less overlapping of the distribution in between the normal and infected erythrocytes in cases of average intensity, skewness, uniformity, entropy and fractal. But in other cases box plots does not show any significant variability as overlapping is present between the distribution of normal and infected stages of erythrocytes.
The textural parameters are tested with paired sample $t$-test analysis and the result obtained (Tables 2 and 3) was quite satisfactory in characterising normal and infected erythrocytes (trophozoites, schizonts and gametocytes). It is observed from the ‘$t$’ values that ‘$t$’ values are higher in all other cases of normal and infected erythrocytes except average contrast and smoothness which lead to the rejection of null hypothesis (feature is not significant) at 1% level of significance. For classification, number of samples taken from both the normal and infected erythrocytes is 100 each. And 60 data are trained to the classifier whereas, rest of data are taken for testing. The SVM classifier shows a 95% overall accuracy. This classifier is also trained and tested with the same ratio (3:2) of the total data. In case of Bayesian classifier 96% sensitivity and 99% specificity is obtained.

Figure 4 Malaria parasite segmentation in different colour spaces (see online version for colours)
Figure 5  Box plot of textural features (see online version for colours)
Quantitative characterisation of Plasmodium vivax in infected erythrocytes

Figure 6  KS density plots of textural features (see online version for colours)
Table 1  Statistical description of normal, trophozoites, schizonts and gametocytes

<table>
<thead>
<tr>
<th>Textural features</th>
<th>Normal</th>
<th>Trophozoites</th>
<th>Schizonts and gametocytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. intensity</td>
<td>116.07 ± 5.45</td>
<td>54.76 ± 15.58</td>
<td>106.20 ± 14.20</td>
</tr>
<tr>
<td>Avg. contrast</td>
<td>65.19 ± 3.78</td>
<td>65.22 ± 3.54</td>
<td>75.38 ± 5.09</td>
</tr>
<tr>
<td>Smoothness</td>
<td>0.06 ± 0.01</td>
<td>0.06 ± 0.01</td>
<td>0.08 ± 0.01</td>
</tr>
<tr>
<td>Skewness</td>
<td>-4.92 ± 0.51</td>
<td>1.84 ± 1.78</td>
<td>-3.88 ± 1.69</td>
</tr>
<tr>
<td>Uniformity</td>
<td>0.08 ± 0.02</td>
<td>0.37 ± 0.13</td>
<td>0.12 ± 0.04</td>
</tr>
<tr>
<td>Entropy</td>
<td>4.56 ± 0.17</td>
<td>3.22 ± 0.87</td>
<td>5.05 ± 0.41</td>
</tr>
<tr>
<td>Fractals</td>
<td>2.29 ± 0.00</td>
<td>2.30 ± 0.01</td>
<td>2.31 ± 0.01</td>
</tr>
</tbody>
</table>

Table 2  Statistical significance of textural features

<table>
<thead>
<tr>
<th>Textural features</th>
<th>Normal vs. trophozoites</th>
<th>Normal vs. schizonts and gametocytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. intensity</td>
<td>&lt; 0.0001*</td>
<td>&lt; 0.0001*</td>
</tr>
<tr>
<td>Avg. contrast</td>
<td>0.0037</td>
<td>0.0186</td>
</tr>
<tr>
<td>Smoothness</td>
<td>0.0037</td>
<td>0.0671</td>
</tr>
<tr>
<td>Skewness</td>
<td>&lt; 0.0001*</td>
<td>&lt; 0.0001*</td>
</tr>
<tr>
<td>Uniformity</td>
<td>&lt; 0.0001*</td>
<td>&lt; 0.0001*</td>
</tr>
<tr>
<td>Entropy</td>
<td>&lt; 0.0001*</td>
<td>&lt; 0.0001*</td>
</tr>
<tr>
<td>Fractals</td>
<td>&lt; 0.0001*</td>
<td>&lt; 0.0001*</td>
</tr>
</tbody>
</table>

Note: *Denotes higher significance

Figure 7  Original image and segmented malaria parasite (see online version for colours)
Quantitative characterisation of Plasmodium vivax in infected erythrocytes

Figure 7  Original image and segmented malaria parasite (continued) (see online version for colours)

<table>
<thead>
<tr>
<th>Original image</th>
<th>Segmented ROI (infected erythrocyte)</th>
<th>Magnified view of ROI</th>
</tr>
</thead>
<tbody>
<tr>
<td>![Image 1]</td>
<td>![Image 2]</td>
<td>![Image 3]</td>
</tr>
<tr>
<td>![Image 7]</td>
<td>![Image 8]</td>
<td>![Image 9]</td>
</tr>
<tr>
<td>![Image 10]</td>
<td>![Image 11]</td>
<td>![Image 12]</td>
</tr>
</tbody>
</table>
4 Conclusions

This study has identified few new dimensions for precise and quantitative characterisation of malaria parasite. Not only that this technique overcomes the detraction of microscopy, which is basically dependent on skill, experience and motivation of pathologist. Here, in this work, we have informed about two different stages (trophozoites, schizonts and gametocytes) of *P. vivax* malaria parasite. In our approach, textural features significantly differentiate healthy and infected erythrocytes at a rate of 95\% (using SVM) and 98\% (using Bayes’) which will in turn afford a proper decision in rapid screening and in critical situation. In our future study, we cordially include the other significant stages of malaria parasite including *Plasmodium falciparum*. Finally, it can be emphasised that this kind of quantitative approach in defining the biological process could be of major interest for researchers as it has depicted how a mathematical technique can track biological divergence due to its unique learning potential.
References


