

Research Article

Comparison of Texture Features Used for Classification of Life Stages of Malaria Parasite

Vinayak K. Bairagi¹ and Kshipra C. Charpe²

¹*E & TC Engineering Department, AISSMS Institute of Information Technology, Pune 411001, India*

²*Electronics Engineering Department, AISSMS Institute of Information Technology, Pune 411001, India*

Correspondence should be addressed to V. K. Bairagi; vbairagi@yahoo.co.in

Received 30 September 2015; Revised 7 April 2016; Accepted 18 April 2016

Academic Editor: Guowei Wei

Copyright © 2016 V. K. Bairagi and K. C. Charpe. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Malaria is a vector borne disease widely occurring at equatorial region. Even after decades of campaigning of malaria control, still today it is high mortality causing disease due to improper and late diagnosis. To prevent number of people getting affected by malaria, the diagnosis should be in early stage and accurate. This paper presents an automatic method for diagnosis of malaria parasite in the blood images. Image processing techniques are used for diagnosis of malaria parasite and to detect their stages. The diagnosis of parasite stages is done using features like statistical features and textural features of malaria parasite in blood images. This paper gives a comparison of the textural based features individually used and used in group together. The comparison is made by considering the accuracy, sensitivity, and specificity of the features for the same images in database.

1. Introduction

An infected female *Anopheles* mosquito is the host, which spread the malaria parasite [1]. Malaria is a highly occurring, vector borne disease in equatorial region. World Health Organization (WHO) has given the analysis report for the year 2013 as 300–500 billion people were the victims of malaria [1, 2]. There is a need for saving lives by making early diagnosis and medication. Nowadays in the biomedical field, all the disease diagnosis methods are turning into computerized automatic way. In this paper, diagnosis of malaria parasite in the blood images is made computerized by using the proposed method which works with image processing.

There are many clinical methods which can be used for malaria diagnosis which are peripheral blood smear (PBS), quantitative buffy coat (QBC), rapid diagnosis test (RDT), Polymerase Chain Reaction (PCR), and Third Harmonic Generation (THG) [3, 4]. Among these entire tests, the PBS is widely used and has limitations of human resistance and time requirement. These limitations can be overcome by making the system automated. The blood images are processed and

then the diagnosis is done. The major problem in processing the blood images is the segmentation of red blood cells (RBC) from the background; a proper separation can help to achieve high accuracy in detecting the presence of plasmodium and helps in classification of life cycle stages of malaria parasite.

The life cycle stages of malaria parasite are differentiated into three different stages, that is, trophozoites, schizocytes, and gametocytes [3]. Each one of them is characterized by unique features such as shape, size, and texture. To identify and classify these stages, the above features will play important role while working with image processing. Features are the key way to differentiate and diagnose the parasite stages.

The organization of this paper is as follows. Section 2 gives the detailed methodology about the work done and all the different image processing techniques used stepwise, while Section 3 provides all the comparison results obtained in presented study and lastly Section 4 gives the discussion followed by conclusion part of the complete research work carried out.

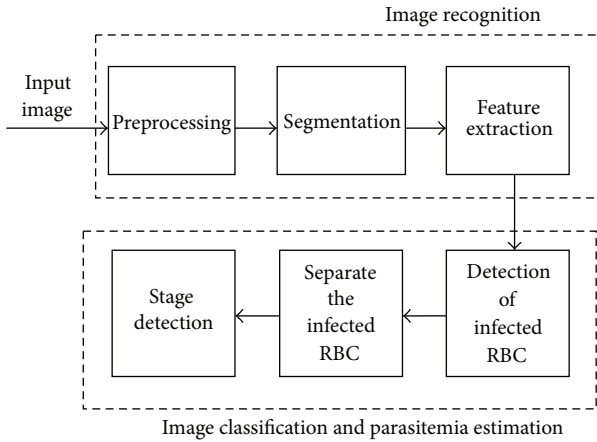


FIGURE 1: Block diagram of the proposed method.

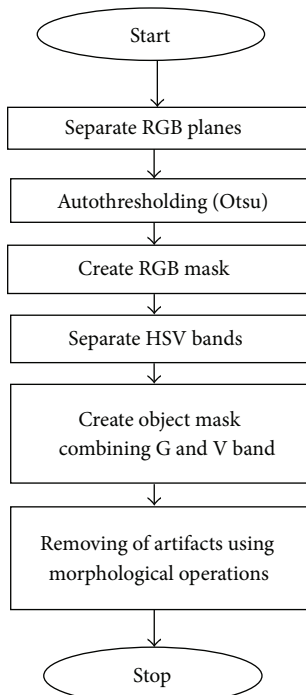


FIGURE 2: Flowchart for preprocessing.

2. Methodology

The method of detection of malaria parasite is mainly subdivided into two parts first as the image recognition and second as the image classification and parasitemia estimation. The proposed method follows the stepwise work as shown in Figure 1, as block diagram.

The first block consists of the three basic steps of image processing, that is, preprocessing, segmentation, and feature extraction, while in the image classification and detection of infected red blood cells that is called parasitemia separation of infected RBC and the stage detection are carried out.

In this work, the malaria images of three different stages, schizont, trophozoite, and gametocyte stages, have been captured from the blood smears. First a database was created

using images from the online database like Center for Disease Control (CDC) and Prevention and from Public Health Image Library [5] consisting of both types of samples parasitic and nonparasitic blood samples. Some images were added into the database which was obtained from Smt. Kashibai Navale General Hospital and Research Centre, Pune, consisting of both malaria positive blood samples and the normal blood samples images. The oil immersion images are taken from the digital camera with 100x zoom. Altogether, 100-image database was created among which 50 images are used for training stages and the remaining 50 images for testing stage were used.

2.1. Preprocessing. Preprocessing is done to make the images easy to be processed further as it removes the noise and other unwanted artifact from the image under observation. The steps followed in preprocessing can be explained as in Figure 2 flowchart.

Initially the RGB image is converted into the gray image as shown in Figure 3. The basic Red, Green, and Blue (RGB) planes of the input image are separated so as to create object mask for segmentation. The RGB plane separation is done as shown in Figure 5. As there are intensity variations all over the image, to take histogram and work on intensity factor, thresholding is the key requirement. Therefore, autothresholding is done using Otsu's method [6]. An object mask is the image which is used as the standard sample to segment the region of interest in the input image and hence it should contain the parasitic part in the image. Therefore, considering all these aspects, the Hue, Saturation, and Value (HSV) bands are created, as shown in Figure 4. Combined GV band is obtained and named as object mask. This object mask is used for further process because it gives good quality image containing all the RBC and parasites clearly as shown in Figure 6. Finally, the morphological operations like closing and opening are performed on the object mask for removing the unwanted artifacts [7].

2.2. Segmentation. Segmentation is the process carried out to obtain the area of interest from the complete image [8]. As the malaria parasite affects the red blood cells [3, 4], the region of interest is the red blood cells in the microscopic images of blood sample. In this work, the segmentation is done to separate each red blood cell. These segmented cells are further processed and then the infected red blood cells are identified. The segmentation technique used in this work is watershed transform with the distance transform. The main reason of using watershed transform is to detect the overlapped red blood cells. The principal concept of watershed is to change the input image into another image whose catchment basins are the objects or regions which are required to be segmented or identified.

The segmentation is carried out as follows:

- (i) Distance transform is calculated using Euclidean distance formula:

$$\text{Distance} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}, \quad (1)$$

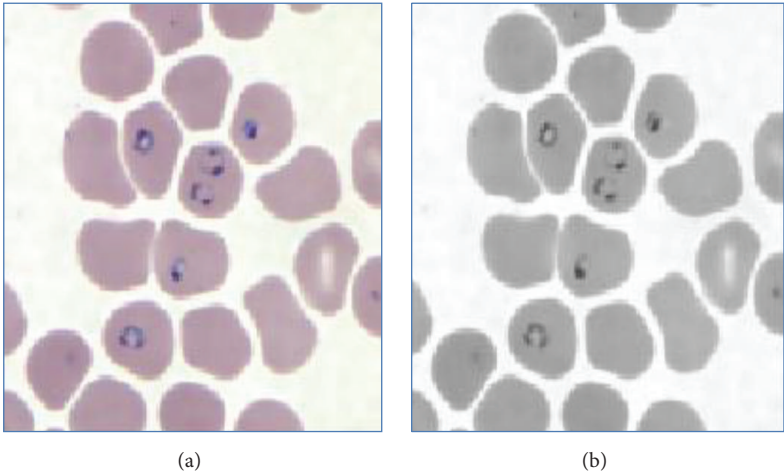


FIGURE 3: RGB image to gray scale image. (a) RGB image. (b) Gray image.

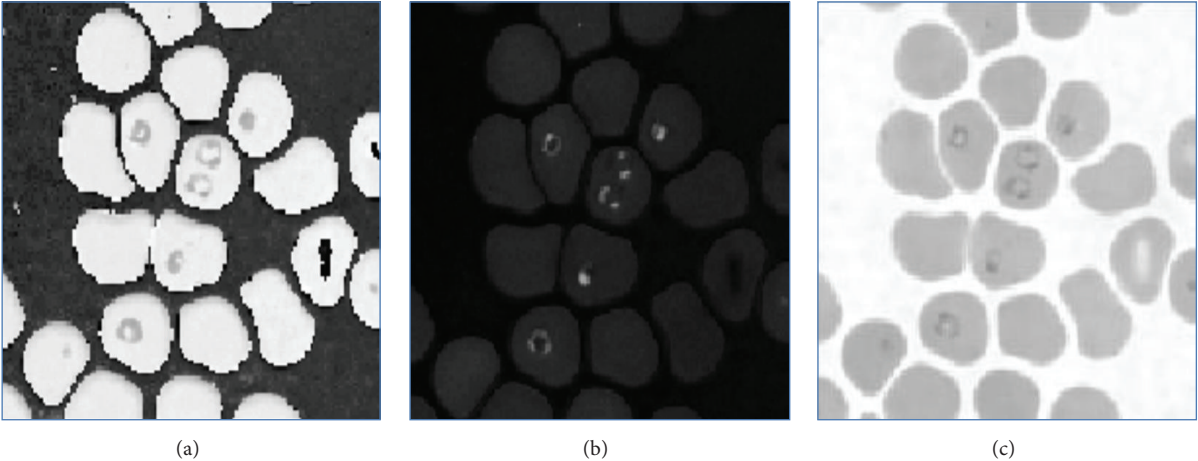


FIGURE 4: Separated HSV bands: (a) H band, (b) S band, and (c) V band.

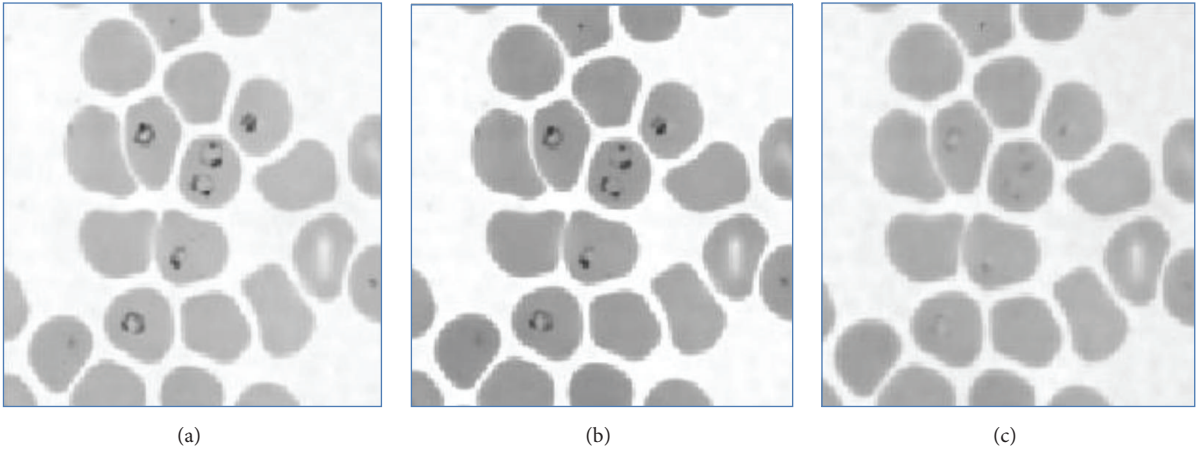


FIGURE 5: Separated RGB bands: (a) R band, (b) G band, and (c) B band.

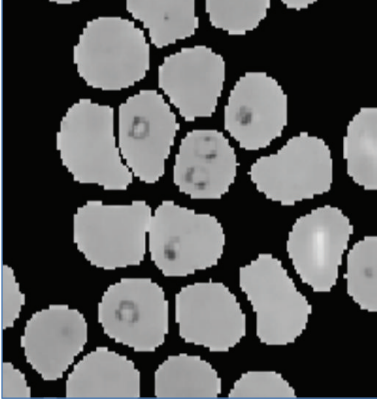


FIGURE 6: Gray scale converted image of object mask.

where “ a ” and “ b ” are the two neighboring pixels between which the distance is calculated and their position can be given as $a(x_1, y_1)$ and $b(x_2, y_2)$, where x_1, x_2 are the x -coordinates and y_1, y_2 are y -coordinates of the two neighboring pixels which are considered. Then, these distance values are calculated for all the pixels in an image and a distance matrix is formed. This is used as an input for the watershed transform.

- (ii) After calculating the distance between each pixel, it creates matrix of these values called label matrix.
- (iii) Label matrix is then given as an input to watershed and the image is obtained as shown in Figures 7(a) and 7(b).
- (iv) This obtained image is segmented using the major and minor axis criteria as shown in Figure 7(c).

2.3. Feature Extraction

2.3.1. Statistical Features. The staining of the slides varies in the color acquired by the RBC, in different colors like blue, purplish, pink, and so forth [4]. Hence the color intensity varies, therefore that can be counted into the feature list, and this variation can be plot on histogram of each image. The histogram is studied using the statistical values of color variation of the image on the basis of different staining. Therefore, the histogram based statistical features are studied in this research. The variance feature gives the variation of the intensity values in image, while the skewness gives the symmetric distribution of those intensity values. There are high peaks observed in the infected cells as compared to the normal cells, and hence the kurtosis values are efficient for detection of infected red blood cells. The statistical features are calculated as described below.

If image function is $I(x, y)$ of two space variables x and y , $x = 0, 1, 2, \dots, N - 1$ and $y = 0, 1, 2, \dots, M - 1$, then the discrete values of function $I(x, y)$ can be taken in terms of the intensity values defined by “ k ” as $k = 0, 1, 2, \dots, G - 1$, where G represents the total number of intensity levels in the image. All these features are calculated using the histogram

of the image $h(k)$ and probability density function $p(k)$. Here $p(k)$ is the ratio of all the intensity values $h(k)$ and the total number of pixels in the image. The features are calculated as follows:

$$\begin{aligned} \text{Mean: } \mu &= \sum_{k=0}^{G-1} k p(k), \\ \text{Variance: } \sigma^2 &= \sum_{k=0}^{G-1} (k - 1)^2 p(k), \\ \text{Skewness: } \mu_3 &= \sigma^{-3} \sum_{k=0}^{G-1} (k - \mu)^3 p(k), \\ \text{Kurtosis: } \mu_4 &= \sigma^{-4} \sum_{k=0}^{G-1} (k - \mu)^4 p(k) - 3. \end{aligned} \quad (2)$$

See [9].

2.3.2. Texture Based Features. In second stage, that is, schizonts, there is growth of granular spotting, that is, the growth of cytoplasm known as matured trophozoites. These cytoplasm dots create a particular texture which differentiates it from other two stages more prominently. Hence, the texture features are included in the feature extraction list. The infected cells are correlated with the background pixels which can be defined by the correlation values calculated and the energy level of those infected cells defines them with higher energy values as compared to normal cell values.

The texture based features are calculated using Gray-Level Cooccurrence Matrix (GLCM) from the input. The GLCM calculates the occurrence of a pixel with gray-level value “ i ” horizontally adjacent to a pixel with the value “ j .” Every element can be mentioned by (i, j) in GLCM which specifies the number of times that the pixel with a value “ i ” occurred horizontally adjacent to a pixel with value “ j .”

These texture based features can be calculated and used as follows:

- (i) *Contrast* is the value which returns the measure of intensity contrast between a pixel and the neighboring pixel over the complete image. The standard range of these values for general images is $[0, (\text{size}(\text{GLCM}, 1) - 1^2)]$ [10]:

$$\text{Contrast} = \sum_{i,j} |i - j|^2 p(i, j). \quad (3)$$

- (ii) *Correlation* is the measure of how pixel is correlated to the neighboring pixel in the complete image. The standard range of these values for general images is $[-1, 1]$ [10]:

$$\text{Correlation} = \sum_{i,j} \frac{(i - \mu_i)(j - \mu_j) p(i, j)}{\sigma_i \sigma_j}. \quad (4)$$

TABLE 1: The values obtained while training dataset for stage I.

Image number	Stage I: trophozoites						
	Statistical features				Texture based features		
	Mean	Variance	Skewness	Kurtosis	Contrast	Energy	Correlation
T1	0.950	0.006	-2.052	-25.114	0.059	0.843	0.859
T2	0.951	0.007	-3.339	-38.326	0.053	0.868	0.906
T3	0.961	0.006	-5.414	-68.288	0.054	0.938	0.894
T4	0.937	0.013	-2.647	-22.71	0.092	0.792	0.918
T5	0.946	0.013	-3.281	-28.756	0.097	0.846	0.913

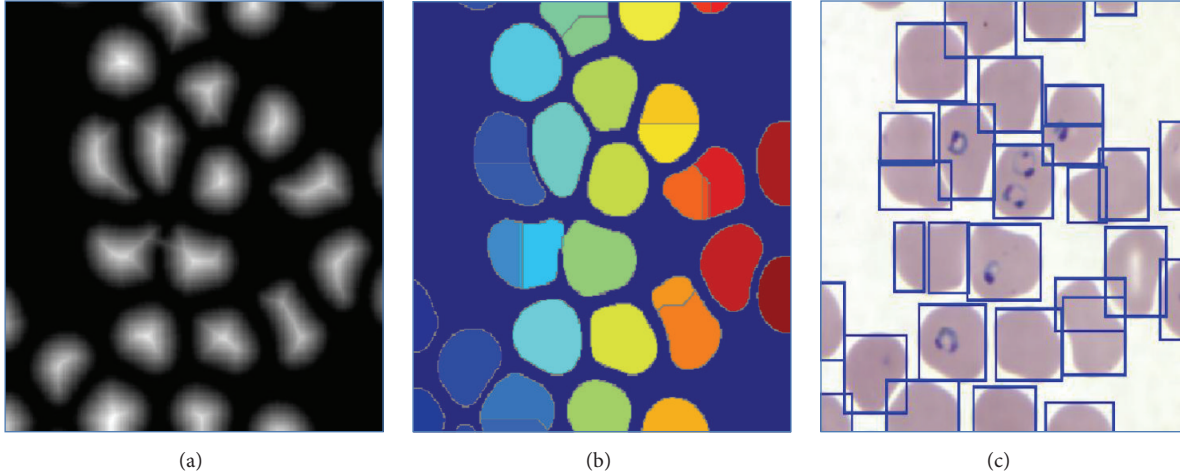


FIGURE 7: Output obtained after watershed transform (a), distance transform (b), and watershed output (c) after applying major and minor axis criteria.

- (iii) *Energy* is the sum of all squared elements in the GLCM matrix. The standard range of these values for general images is [0, 1] [10]:

$$\text{Energy} = \sum_{i,j} p(i, j)^2, \quad (5)$$

where the Gray-Level Cooccurrence Matrix (GLCM) is calculated first and then the properties, like contrast, correlation, and energy, are calculated.

The different samples of all three stages of malaria parasite are taken for the training of the Support Vector Machine classifier as shown in Figure 8. The three images in (a) are of third stage as gametocytes while (b) consists of second stage, that is, the schizonts, and (c) consists of the first stage trophozoites image samples which are used for training purpose.

The different values for all seven features are given to the SVM for training the stages. The images used are labeled according to the stages name as the trophozoites are named as T1, T2, T3, . . . , and so forth, while schizonts are named as S1, S2, S3, . . . , and so forth, and lastly third-stage gametocytes are named as G1, G2, G3, . . . , and so forth. The values of features are as shown in Tables 1, 2, and 3 for each stage I, II, III, respectively.

2.4. Image Classification and Parasitemia Estimation. Classification is done using the feature matrix created by calculating

all the feature values for each RBC. This feature matrix of trained data is compared with the feature matrix for the image taken as an input. Then, the class is decided accordingly on the comparison results and displayed. The Support Vector Machine (SVM) is for the classification. SVM is a supervised classifier which requires trained dataset and can be used for multistage classification using multi-SVM.

Based on the classification results obtained in the work, the quality of the result is to be decided by calculating as follows:

$$\begin{aligned} \text{Accuracy} &= \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})}, \\ \text{Sensitivity} &= \frac{\text{TP}}{(\text{TP} + \text{FN})}, \\ \text{Specificity} &= \frac{\text{TN}}{(\text{FP} + \text{TN})} \end{aligned} \quad (6)$$

(see [11]), where TP is true positive (infected sample correctly classified), TN is true negative (normal sample correctly classified), FP is false positive (normal sample misclassified), and FN is false negative (infected sample misclassified).

All the values of calculation done for the above values are as shown in Table 4 with the elapse time requirement. The algorithm is executed on hardware platform of Laptop, model: Lenovo Win7PC, Core i5 @250 GHz, RAM: 8 GB, 64-bit OS. From this, it is observed that the time required for calculation is proportional to the number of RBC in

TABLE 2: The values obtained while training dataset for stage II.

Image number	Stage II: schizonts				Texture based features		
	Statistical features				Contrast	Energy	Correlation
Mean	Variance	Skewness	Kurtosis				
S1	0.961	0.004	-1.055	-15.137	0.0269	0.766	0.885
S2	0.772	0.107	-0.308	-0.940	0.397	0.203	0.937
S3	0.812	0.085	-0.449	-1.539	0.377	0.272	0.929
S4	0.785	0.100	-0.359	-1.133	0.381	0.228	0.938
S5	0.621	0.230	0.015	0.0317	1.144	0.106	0.885

TABLE 3: The values obtained while training dataset for stage III.

Image number	Stage III: gametocytes				Texture based features		
	Statistical features				Contrast	Energy	Correlation
Mean	Variance	Skewness	Kurtosis				
G1	0.835	0.055	-0.194	-0.825	0.242	0.267	0.915
G2	0.974	0.001	-1.758	-41.813	0.008	0.926	0.887
G3	0.830	0.056	-0.157	-0.661	0.188	0.256	0.933
G4	0.897	0.043	-1.347	-6.491	0.201	0.544	0.942
G5	0.945	0.013	-2.679	-22.707	0.203	0.753	0.823

TABLE 4: Calculations of total RBC, infected RBC, and the TP, TN, FP, and FN values considering all the seven features.

Image number	Total RBC count		Infected RBC count		TP	TN	FP	FN	Elapsed time
	Manual	Algorithm	Manual	Algorithm					
Ts-1	28	29	4	5	4	24	1	0	33.94
Ts-2	11	10	1	1	1	10	0	0	21.96
Ts-3	30	33	3	3	3	27	0	0	34.19
Ts-4	14	15	1	1	1	14	0	0	23.97
Ts-5	10	10	4	4	4	6	0	0	21.57
Ts-6	10	10	3	3	3	7	0	0	20.82
Ts-7	11	11	3	4	3	8	1	0	21.00
Ts-8	7	7	1	1	1	6	0	0	17.15
Ts-9	12	11	3	3	3	8	0	0	20.72
Ts-10	12	12	1	2	1	11	1	0	21.36
Ts-11	22	25	3	4	3	23	1	0	30.84
Ts-12	22	25	3	5	2	23	1	1	30.69
Ts-13	8	10	5	5	5	5	0	0	19.69
Ts-14	15	16	3	3	3	13	0	0	23.95
Ts-15	36	37	1	1	1	36	0	0	31.83
Total	248	261	39	45	38	221	5	1	24.91

the image. All the images used for testing were labeled as Ts-1, Ts-2, Ts-3, . . . , and so forth, where Ts stands for testing sample and the number of sample followed by this.

Some of the testing samples are shown in Figure 9 where in (a) there are first-stage sample images, that is, trophozoites, and (b) consists of second-stage sample images, that is, schizonts, and lastly (c) consists of all the third-stage parasite samples, that is, gametocytes.

3. Experimental Results

The complete process is performed as mentioned step-wise above by considering all the statistical features and

texture based features and then obtained the results in terms of accuracy; sensitivity and specificity are calculated in percentage. Secondly the statistical features were kept constant and texture features were considered one by one individually and calculated the values for accuracy, sensitivity, and specificity for the same database as mentioned before.

Table 5 gives all the results obtained after calculating the values of accuracy, sensitivity, and specificity for individual texture based features and altogether with the statistical features this comparison is shown in a graphical way in Figure 10.

TABLE 5: Comparison of individual texture based features in terms of accuracy, sensitivity, and specificity.

	Contrast%	Energy%	Correlation%	Altogether%
Accuracy	85.1	90.55	91.8	97.7
Sensitivity	47.7	85.36	75	97.4
Specificity	93.7	91.66	95.3	97.7

TABLE 6: Summary of comparison with other works.

Sr. number	Author	Reference number	Features used	Accuracy
1	Ross et al.	[12]	Morphological (erosion, dilation)	85.3%
2	Tek et al.	[13]	Color and cell geometry (Hough transform for circle detection)	83%
3	Das et al.	[8]	Morphological (opening, closing)	88.77%
4	Kareem	[14, 15]	Edge filtering (Sobel, Canny)	86%
5	Savkare and Narote	[9]	Statistical features (area, mean, and variance)	94%
6	Alomari et al.	[16]	Geometry (modified randomized circle detection)	95%
7	Presented system	[17]	Texture based and statistical based	97.7%

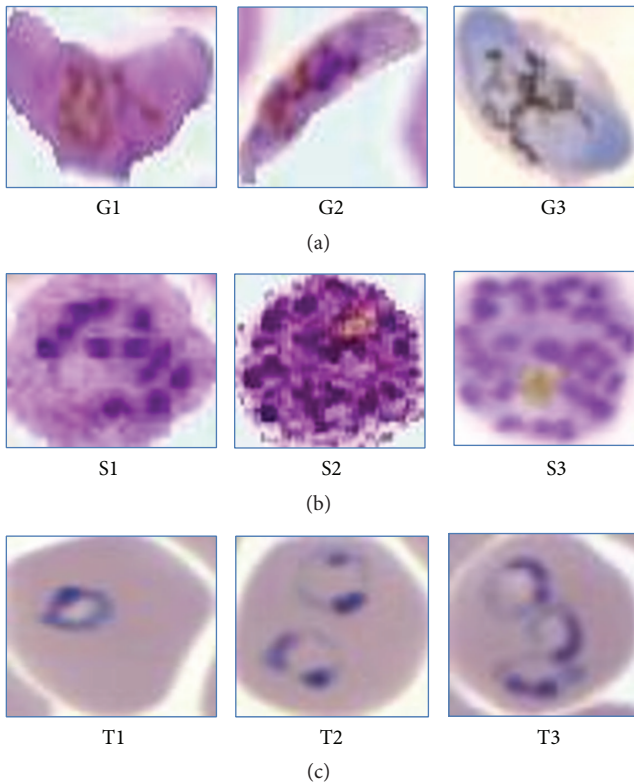


FIGURE 8: Samples of training images.

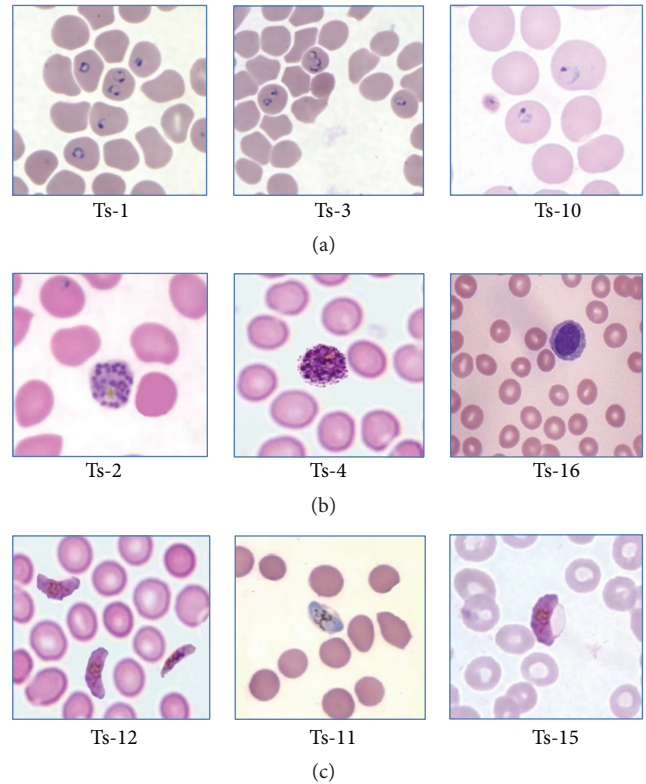


FIGURE 9: Sample images from test database used.

4. Results and Discussion

From the observation done in Table 5 for accuracy, sensitivity, and specificity for considering all features together and for individual texture based features, it can be seen that if all the features are used together, then the high diagnosis accuracy can be achieved. If all the features are used together, the SVM gets trained very efficiently and the classification of the parasitic and nonparasitic samples and the stages determination are done with higher accuracy of about 97%.

This means the diagnosing of parasite and the detection of stage are more reliable if all the features are used together. Among three texture based features, namely, contrast, energy, and correlation, when only energy features are considered for classification, it gives more accuracy, sensitivity, and specificity up to 90.55%, 85.36%, and 91.66%, respectively. The high accuracy achieved by presented system can be compared with other methods in the literature as shown in Table 6.

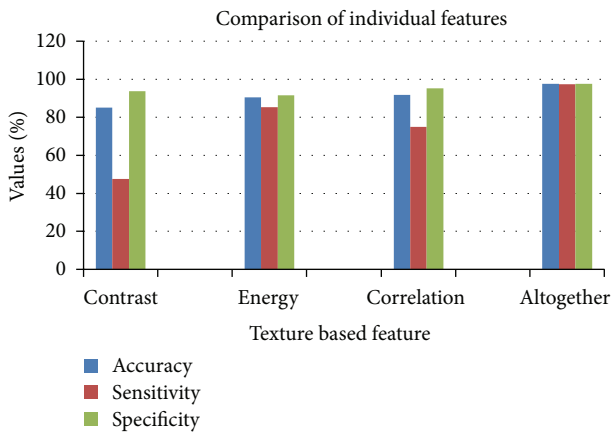


FIGURE 10: Comparison of individual texture based features in terms of accuracy, sensitivity, and specificity.

5. Conclusion

As the number of people infected by malaria in tropical region is high, there is special need of early diagnosis of malaria parasite, for its prevention against getting into next stage and medication of a person. This paper has presented automated diagnosis of malaria parasite in blood images. The diagnosis is done by using features like statistical and textural based features which has enhanced the diagnosis accuracy of the presented system. Additionally the presented method is also helpful in detecting stages of malaria parasite. The segmentation is done using watershed algorithm using distance transform which counts the overlapped RBC in the blood images. The system achieved high accuracy of 97.7% for automatic diagnosis of malaria parasite in blood images.

Further this technique can be used for diagnosing different hematological disorders by training the SVM accordingly. The textural properties can be used for enhancing the results of accuracy and sensitivity.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

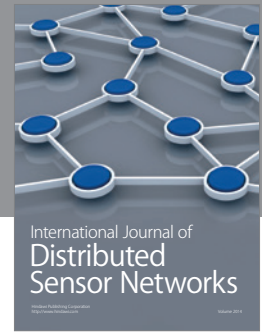
The authors gratefully acknowledge and thank Dr. Saurabh Joshi from Smt. Kashibai Navale General Hospital and Research Centre; he helped them in the formation of database in a proper way and guided them in the diseases related medical terms. The authors would like to thank all authors of the references which have been used.

References

- [1] National Vector Borne Disease Control Program (NVDCP), Directorate General of Health Services Ministry of Health and Family Welfare statistical report on malaria, Delhi, <http://www.nvbdc.gov.in/malaria9.html>.
- [2] World Health Organization (WHO) statistical analysis of Malaria diseases in the year 2013, http://www.who.int/malaria/publications/world_malaria_report_2013/en/.
- [3] World Health Organization, *Basic Malaria Microscopy, Part I Learners Guide*, World Health Organization, Genève, Switzerland, 2nd edition, 2010.
- [4] P. Jain, B. Chakma, S. Patra, and P. Goswami, "Potential biomarkers and their applications for rapid and reliable detection of malaria," *BioMed Research International*, vol. 2014, Article ID 852645, 20 pages, 2014.
- [5] Centres for Disease Control and Prevention: Public Health Image Library, 2005, <http://phil.cdc.gov/phil/home.asp>.
- [6] Y. Purwar, S. L. Shah, G. Clarke, A. Almagairi, and A. Muehlenbachs, "Automated and unsupervised detection of malaria parasites in microscopic images," *Springer-Malaria Journal*, vol. 10, pp. 11–22, 2011.
- [7] K. M. Khatri, V. R. Ratnaparkhe, S. S. Agrawal, and A. S. Bhalchandra, "Image processing approach for malaria parasite identification," in *National Conference on Growth of Technologies in Electronics, Telecom and Computers—India's Perception (GTETC—IP '14)*, IJCA Proceedings, pp. 5–7, May 2014.
- [8] D. Das, M. Ghosh, and C. Chakraborty, "Probabilistic prediction of malaria using morphological and textural information," in *Proceedings of the International Conference on Image Information Processing (ICIIP '11)*, pp. 20–24, Bengal, India, 2011.
- [9] S. S. Savkare and S. P. Narote, "Automatic system for classification of erythrocytes infected with malaria and identification of parasite life stage," in *Proceedings of the 2nd International Conference on Communication, Computing & Security [ICCCS '12]*, vol. 6, pp. 405–410, Elsevier, Mumbai, India, 2012.
- [10] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Pearson Prentice Hall, Upper Saddle River, NJ, USA, 3rd edition, 2008.
- [11] A. S. Abdul-Nasir, M. Y. Mashor, and Z. Mohamed, "Colour image segmentation approach for detection of malaria parasites using various colour models and k-means clustering," *WSEAS Transactions on Biology and Biomedicine*, vol. 10, no. 1, pp. 41–55, 2013.
- [12] N. E. Ross, C. J. Pritchard, D. M. Rubin, and A. G. Dusé, "Automated image processing method for the diagnosis and classification of malaria on thin blood smears," *Medical & Biological Engineering and Computing*, vol. 44, no. 5, pp. 427–436, 2006.
- [13] F. B. Tek, A. G. Dempster, and I. Kale, "Parasite detection and identification for automated thin blood film malaria diagnosis," *Computer Vision and Image Understanding*, vol. 114, no. 1, pp. 21–32, 2010.
- [14] S. Kareem and R. C. S. Morling, "Automated P. falciparum detection system for post-treatment malaria diagnosis using modified annular ring ratio method," in *Proceedings of the 14th International Conference on Modelling and Simulation*, pp. 432–436, IEEE Computer Society, Cambridge, UK, 2014.
- [15] S. Kareem, I. Kale, and R. C. S. Morling, "Automated malaria parasite detection in thin blood films: a hybrid illumination and color constancy insensitive, morphological approach," in *Proceedings of the IEEE Asia Pacific Conference on Circuits and Systems (APCCAS '12)*, pp. 240–243, Kaohsiung, Taiwan, December 2012.
- [16] Y. M. Alomari, S. N. H. Sheikh Abdullah, R. Zaharatul Azma, and K. Omar, "Automatic detection and quantification of WBCs and RBCs using iterative structured circle detection algorithm,"

Computational and Mathematical Methods in Medicine, vol. 2014, Article ID 979302, 17 pages, 2014.

- [17] K. Charpe and V. K. Bairagi, "Automated malaria parasite and there stage detection in microscopic blood images," in *Proceedings of the 9th IEEE Sponsored International Conference on Intelligent Systems & Control*, IEEE, Coimbatore, India, January 2015.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

