ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 2 Issue 11, November 2013

Comparative Study of Malaria Parasite Detection using Euclidean Distance Classifier & SVM

Ms. Snehal Suryawanshi, Prof. V. V. Dixit

***Abstract*—Even though there is tremendous computerization in the field of Cell Segmentation, Feature Extraction and Classification. Still it has become necessary to improve techniques being used, so as to get better Accuracy, Sensitivity and PPV (Positive Predictive Value). This paper presents enhanced technique for Malaria Parasite Detection, where cell segmentation process consists of various steps such as image binarization using Poisson's distribution based Minimum Error Thresholding, followed by Morphological Opening for the purpose of refinement. Seed point localization is done by multiscale LoG filter. Since frequency and orientation representations of Gabor filter are same as that of human visual system, it is used for feature extraction. Two algorithms are compared in this paper in order to get superior classification. Results show that SVM (Support Vector machines) gives better accuracy of 93.33% than that of Euclidean Distance Classifier which is 80%.**

***Index Terms*— Image processing, Biomedical Engineering, Euclidean distance classifier, SVM.**

I. INTRODUCTION

Malaria is major Public health issue, caused by female Anopheles mosquito. According to WHO, there are 300 to 500 million clinical cases of malaria each year resulting in 1.5 to 2.7 million deaths [1]. As automation is increasing rapidly, it has become part and parcel even in medical field to have system which can identify the Malaria parasite within the blood cell images automatically, so as to remove manual error with higher speed and reliability.

A number of methods have been proposed for automatic parasite detection in Giemsa stained blood films based on different approaches. These approaches include pixel-based parasite detection [6], detection based on morphological processing of segmented parasites [2, 3], or detection by extracting image features from the segmented cells [4].

***Manuscript received Nov, 2013*.**

***Ms. Snehal R Suryawanshi** B.E. (E&TC), M.E Signal Processing (Pursuing) from Sinhgad college of Engineering, Pune, M. S. India, has teaching experience of 1.4 years in Pune University M.S. India. Ph. No. 91-9665013518 .*

***Prof. V. V. Dixit ,** Associate professor, Department of E&TC., Sinhgad college of Engineering, Pune M.S. India, Pursuing his Phd in Image processing, has working experience of 18 years in Pune University M.S. India.*

II. SYSTEM ARCHITECTURE

With the help of image processing techniques [14], we have designed a system which is capable detect the presence of malaria parasite within the blood cell images by comparing the test image features with the training dataset and classification is done to check whether the human host is suffering from Malaria. The system block diagram is shown in Fig. 1.

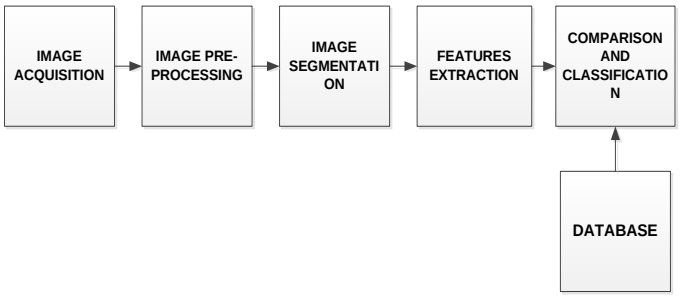


Fig. 1 Block Diagram of Malaria Parasite Detection

A. Image Acquisition

The input images of Giemsa stained blood smears are selected from the Public Health Image Library [5]. Images are of different shape and sizes. Images show high variations in intensity, contrast color tone, etc.

B. Image Pre-processing

The pre-processing block is designed, to remove unwanted effects from the image and to adjust the image as necessary for further processing

The microscopic input image is converted from RGB to gray scale to reduce the processing time. RGB to gray conversion is done by averaging all the three components i.e. R, G and B which results in gray scale, shown as,

$$G = \frac{R+G+B}{3}$$

C. . Image Segmentation

Poisson's distribution based Minimum error thresholding is used for binarization of the image giving an optimal

threshold value [7], [13]. The optimal threshold is obtained by,

$$t^* = \arg\min\{\mu - P_0(t)(\ln P_0(t) + \mu_0(t)\ln\mu_0(t)) - P_1(t)(\ln P_1(t) + \mu_1(t)\ln\mu_1(t))\}$$

Where, P_o and P_1 are the prior probabilities of foreground and background with the mean μ.

Further the same image is refined using morphological opening [10], which removes unwanted noise and gives the clear view of the parasites. Opening is basically erosion followed by dilation,

$$A \circ B = (A \ominus B) \oplus B$$

Foreground is finally extracted from the background by masking process. It is executed by inverting the morphological output and then multiplying with the gray scale image as;

Foreground Extraction = ((255-Output of Opening)/255) x Gray-scale image

D. Feature Extraction

Since the foreground is being extracted next step is to localize seed point so as to extract them. The localization process is carried out by Multiscale LoG (Laplacian of Gaussian) filter [7], [11]. LoG assigns individual marker per cell and then extracts single cell from the cluster of cells. Edge detection can be done effectively. Boundaries obtained by this process are robust. Multiscale LoG filter is,

$$LoG(x, y; \sigma) = \frac{\partial^2 G(x, y; \sigma)}{\partial x^2} + \frac{\partial^2 G(x, y; \sigma)}{\partial x^2}$$

All the localized seed points are labeled by using binary object labeling technique. Then we extract the first type of features which is cellular dimension, given by;

$$Cellular\ Dimension = \frac{sum\ of\ labels}{Number\ of\ elements\ in\ the\ Array\ of\ Image}$$

The other feature is being obtained by Gabor filtering process [12]. Gabor filter extracts the seed points using equation given below,

$$G_{\sigma,\psi,\theta}(x, y) = g_\sigma(x, y) \cdot \exp(j2\pi\psi(x\cos\theta + y\sin\theta))$$

E. Comparison and Classification

This is the most challenging part; system results are completely reliable on this block. Since it is medical based project, it directly affects on the human life. Hence classification becomes of high risk. Comparison of the test image is being done with the trained dataset in order to achieve our objective of classifying whether the image is infected by parasite or not. Classification is done by using two different classifiers viz., Euclidean distance classifier and Support Vector machine (SVM) classifier. Basically Euclidean distance [8] is the distance between two points.

$$D(x,y) = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2 + \cdots + (y_n - x_n)^2}$$

While SVM achieves, classification by realizing a linear or non-linear separation surface in the input space [9].

$$W \bullet X + b = 0$$

Performance of both the classifier is different accuracy wise. Thus a classification model is created in the form of hyper plane, having two different features on each axis. Here the features being considered are cellular dimension and the intensity feature.

III. FLOW OF THE SYSTEM

The detailed flow of the system is shown in the Fig. 2. It elaborates the process of each and every block and gives pictorial representation of the whole system.

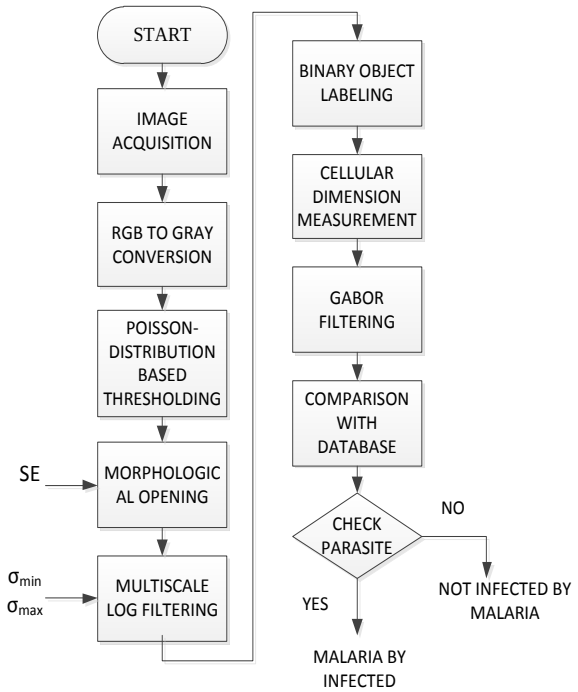


Fig. 2 Flow chart outlining the main steps of the proposed system.

IV. RESULT AND DISCUSSION

Blood cell analysis can be easily done through image processing. Different diseases lead to different changes in blood. These changes in cell images of the blood can guide to detect the illness. Though further analysis of the patient should be done manually for confirmation of the diseases, initial diagnosis can be performed using image processing. In this project the training set is made of 60 images while for testing we use around 30 images in the database. When tested with a selected set of different images other than that used for

training the Euclidean distance classifier and SVM was able to categorize it accordingly.

In order to compare the performance of the two classifiers (SVM and Euclidean distance classifier) which we have used in our system design can be explained with the help of some parameters such as, Accuracy, Sensitivity and PPV (Positive Predictive Value). Formulae for the same are given below,

$$\text{Accuracy} = (TP+TN) / (TP+TN+FP+FN)$$

$$\text{Sensitivity} = (TP) / (TP+FN)$$

$$\text{PPV} = (TP) / (TP+FP)$$

Where, TP: True positive, TN: True negative, FP: False positive, FN: False negative.

TABLE I. PERFORMANCE MEASURE FOR VARIOUS PARAMETERS BY USING EUCLIDEAN DISTANCE CLASSIFIER.

Blood Images	Number of Test Images	Malaria	Non-Malaria
Malaria	15	12(TP)	03(FN)
Non-Malaria	15	02(FP)	13(TN)

TABLE II. PERFORMANCE MEASURE FOR VARIOUS PARAMETERS BY USING SVM.

Blood Images	Number of Test Images	Malaria	Non-Malaria
Malaria	15	14(TP)	01(FN)
Non-Malaria	15	01(FP)	14(TN)

Hence from the above given two table accuracy, sensitivity and PPV of SVM is far better than that of Euclidean distance classifier.

TABLE III. PERFORMANCE COMPARISON BETWEEN TWO CLASSIFIER

Parameters	Euclidean Distance Classifier	SVM Classifier
Accuracy	83.33%	93.33%
Sensitivity	80%	93.33%
PPV	85.71%	93.33%

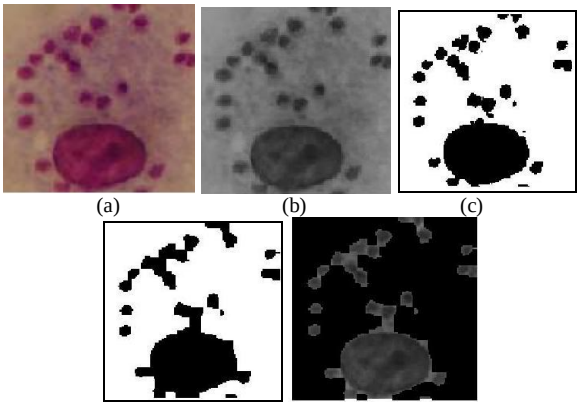
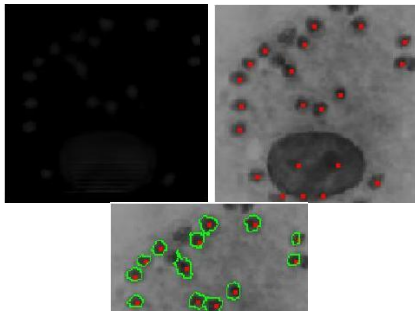


Fig. 3 Steps of system architecture: (a) Original image, (b) Gray scale image, Segmentation: (c) PD thresholded image, (d) Morphologically operated SI, (e) Extracted foreground.



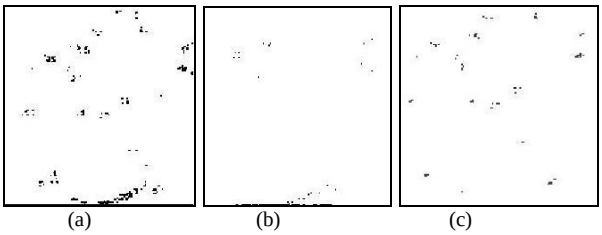


Fig. 5 Steps for feature extraction: (a) Gabor features (OR logic), (b) Gabor features (AND logic), (c) Intensity Feature points.

All above steps in fig. 3, fig. 4 and fig. 5 are performed on each and every test image and finally whichever features are extracted those are compared with the dataset. Thus as you know comparison is done with the help of two classifiers viz. Euclidean distance classifier and SVM. Both the classifiers create hyper plane consisting of two axes, class I on x-axis and class II on y-axis. The training images are the class I Malaria parasitic images and class II are the Non-Malaria parasitic images. The hyper planes of both the algorithm are shown as follows in fig. 6;

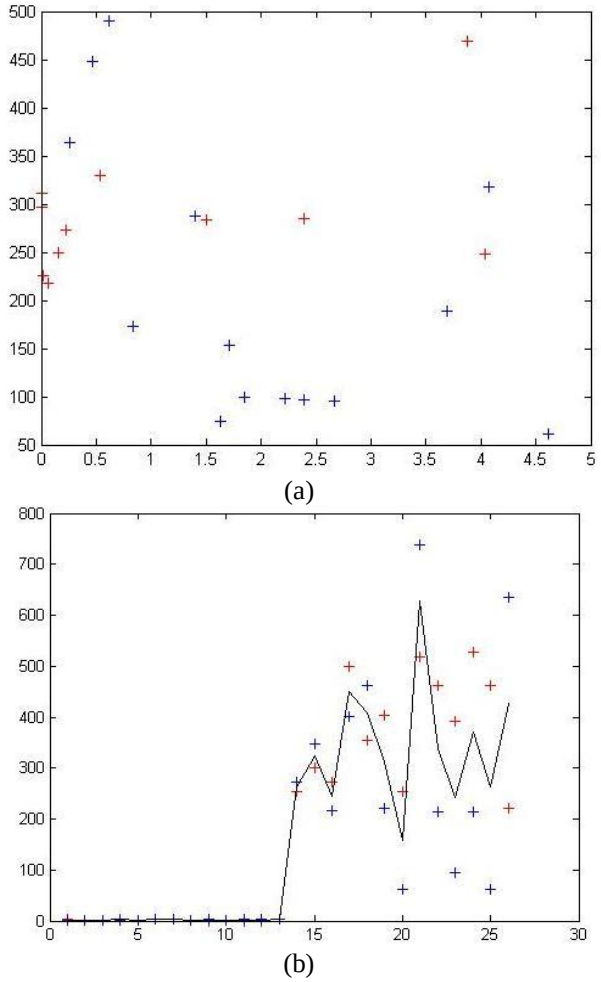


Fig. 6 Hyper plane: (a) Euclidean distance classifier, (b) SVM classifier.

V. CONCLUSION

There are many possible classification scenarios in the problem of malaria diagnosis. In this project, we focused mainly on general evaluation of the features and we tested their intensity features and cellular dimension with trained dataset on the problem of distinguishing between infected and non-infected red blood cells. The detection of infected cells is the primary task and usually has to be carried out before any further analysis can be performed. The results have shown that very good discrimination between the two classes can be obtained using intensity features obtained by Gabor filtering. As two techniques are used, if comparison is done between the SVM gives 93.33% accuracy while the Euclidean distance classifier has 83.33% accuracy.

REFERENCES

[1] The Malaria Control Programme, World Malaria Report, World Health Organization, 2008.

[2] Di Ruberto, C., et al: *"Morphological Image Processing for Evaluating Malaria Disease"*. IWVF4, 2001, pp. 739-748

[3] Di Ruberto, C., et al: *"Analysis of infected blood cell images using morphological operators"*. Image and Vision Computing 20(2).2002, pp. 133-146.

[4] Ross, N.E., Pritchard, C.J., Rubin, D.M., Duse, A. G.: *"Automated Image Processing method for the diagnosis and classification of Malaria on thin blood smears"*. Medical and Biological Engineering and Computing 44, 2006, pp. 427-436.

[5] Centres for Disease Control and Prevention: Public Health Image Library [online]. 2005 www: http://phil.cdc.gov/phil/ home.asp.

[6] Diaz, G., Gonzalez, F., Romero, E., *"Infected Cell Identification in thin Blood Images Based on Color Pixel Classification: Comparison and Analysis"*. Springer Berlin, 2007, pp. 812-821.

[7] Yousef Al. K., Wiem L. W. Lee, B. R.: *"Improved Automatic Detection and Segmentation of Cell Nuclei in Histopathology Images"*. IEEE Transaction on Biomedical Engineering, Vol. 57, No. 4, pp.841-852, 2010.

[8] O. Culsenaire and B. Macq: *"Fast Euclidean Distance Transformation by propagation"*, Computer Vision and Image Understanding. Vol. 76, No. 2, Nov., pp. 163-172, 1999.

[9] C. J. C. Burges: *"A Tutorial on Support Vector Machines for Pattern Recognition"*. Data Mining and Knowledge Discovery. 2 (2), pp. 121-168, 1998.

[10] Soille, P.: *"Morphological Image Analysis: Principles and Applications"*. Springer-Verlag, Heidelberg, 2003.

[11] T. Lindeberg: "*Feature detection with automatic scale selection"*, Int. J. Comput. Vis., vol. 30, no. 2, pp. 79–116, Nov. 1998.

[12] Javlerk movellan: *"Tutorial on Gabor Filters"*, Copyright C 1996, 2002.

[13] J. L. Fan: "*Notes on Poisson distribution-based minimum error thresholding"*, Pattern Recognit. Lett., vol. 19, no. 5–6, pp. 425–431, Apr. 1998.

[14] Gonzalez, R.C., Woods, R.E., Eddins, S.L., *"Digital Image Processing Using, MATLAB"*. Prentice Hall, 2004.

Ms. Snehal R Suryawanshi B.E. (E&TC), M.E Signal Processing (Pursuing) from Sinhgad college of Engineering, Pune, M. S.India, has teaching experience of 1.4 years in Pune University M.S. India

Prof. V. V. Dixit Associate professor, Department of E&TC., Sinhgad college of Engineering, Pune M.S. India, Pursuing his Phd in Image processing, has working experience of 18 years in Pune University M.S. India.