International Journal of Emerging Engineering Research and Technology
Volume 2, Issue 4, July 2014, PP 107-112
ISSN 2349-4395 (Print) & ISSN 2349-4409 (Online)

Automatic Identification of Malaria Parasites using Image Processing

Miss. Shruti Annaldas
M.E.(ENTC) NBNSCOE, Solapur
shruti_annaldas@yahoo.com

Prof. S.S.Shirgan
Dept. of ENTC NBNSCOE, Solapur
ssshirgan@gmail.com

Prof. V.R.Marathe
Dept. of ENTC NBNSCOE, Solapur
vks.marathe@gmail.com

Abstract: *Objective of the paper is to develop an image processing algorithm to automate the diagnosis of malaria on thin blood smears. The image classification system could positively identify malaria parasites present, and differentiate the species of malaria. Morphological and novel threshold selection techniques can be used to identify erythrocytes (red blood cells) and possible parasites present on microscopic slides. Image features based on color, texture and the geometry of the cells and parasites will be generated and studied. The extracted features could be properly classified to distinguish between true and false positiveness and then to diagnose the species of the infection. The sensitivity and positive predictive value is measured.*

Keywords: *Support Vector Machine, Neural network, malaria parasites.*

1. INTRODUCTION

The important thing in human life is its life and health. So to make it secure and to protect for different type of diseases using modern technology here I have develop certain algorithm which will help full in identifying serious diseases like Malaria. Identification of malaria at early stage will be helpful as its effect increases drastically and cause great harm to human life. The malaria is due to imbalance (increase) of amount of Malaria parasites in the patient's blood and an indicator for the degree of infection. Malaria is caused by a blood parasite named Plasmodium spp. It affects at least 200 to 300 million people every year and causes an estimated 3 million deaths per annum. Diagnosis and medication of it is necessary [1],[2],[3].

So for medication should start at proper time is very important to identify the diseases very fast and accurate. So to achieve this I have developed an algorithm which will very helpful for identifying the diseases fast and accurate which will give accuracy about 96.72% and work efficiently and easy to use. In this technique I have use the blood cell images to find out whether the patient is malaria affected or not. For that here I have used the statistical characteristics of image like (Skewness, Standard deviation, kurtosis and Energy) which will overcome the problem of not clearly visible boundaries of cells.

For the classification here I implemented three algorithms which on by discussed latter and have different advantages over increase in performance. The classification techniques utilized are as follows. Neural Network, Support Vector Machine.

2. METHODOLOGY

2.1 Features or parameters Description

Since the chosen features affect the classifier performance, selection of feature which is to be used in a specific data classification problem is as important as the classifier itself [5]. The features which give predominant difference between normal and infected cells are identified and used for training purpose. The selected features are color and statistical based.

1. Phase of Image (PHI).
2. Mean Value of Green Plane
3. Skewness

4. Kurtosis.
5. Standard Deviations.
6. Energy.

The above parameters are used for feature extraction. The statistical features use gray level histogram and saturation histogram of the pixels in the image and based on such analysis, the mean value; angular second momentum, Skewness, Standard deviation, Kurtosis are treated as the features [6] and calculated using above equations. The description of parameters is as follows.

2.1.1. Phase of Image (PHI)

The method is based on the Fourier Transform; The discrete Fourier transforms of the images will transform the image into frequency domain from which we can easily calculate the phase of image phase:

Phase of FT:

$$\emptyset\big(F(u)\big) = \tan^{-1}\left(\frac{I(u)}{R(u)}\right)$$

The advantage of this method is that the discrete Fourier transforms and its inverse can be performed using the fast Fourier transform, which is much faster than correlation for large images.

2.1.2. Benefits

Unlike many spatial-domain algorithms, the phase correlation method is resilient to noise, occlusions, and other defects typical of medical or satellite images. The method can be extended to determine rotation and scaling differences between two images by first converting the images to log-polar coordinates. Due to properties of the Fourier transform, the rotation and scaling parameters can be determined in a manner invariant to translation.

2.1.3. Mean of Green Plane (M_g)

The planes of malaria image are separated and the mean of Green plane is taken.

$$M\_g = \frac{1}{(MNX)} \sum_{(x,y)=0}^{(m,n)} (f(x,y))$$

2.1.4. Skewness

Skewness is a measure of the asymmetry of the data around the sample mean. If skewness is negative, the data are spread out more to the left of the mean than to the right. If skewness is positive, the data are spread out more to the right. The skewness of the normal distribution (or any perfectly symmetric distribution) is zero. The skewness of a distribution is defined as

$$(\text{Skewness}) = \frac{E(x-\mu)^3}{\sigma^3}$$

Where b is the mean of x, σ is the standard deviation of x, and E (t) represents the expected value of the quantity L.

2.1.5. Kurtosis

Kurtosis is a measure of how outlier-prone a distribution is. The kurtosis of the normal distribution is 3. Distributions that are more outlier-prone than the normal distribution have kurtosis greater than 3; distributions that are less outlier-prone have kurtosis less than 3.The kurtosis of a distribution is defined as

$$K(\text{Kurtosis}) = \frac{E(x-\mu)^4}{\sigma^4}$$

Where μ is the mean of x, σ is the standard deviation of x, and $E(t)$ represents the expected value of the quantity t.

2.1.6. Standard Deviations

Standard Deviations normalizes by n-1 where n is the sample size. The result Y is the square root of an unbiased estimator of the variance of the population from which X is drawn, as long as X consists of independent, identically distributed samples. The standard deviation is

$$\text{Standard Deviation} = \left[\frac{1}{N}\sum_{i=1}^{n}(x-\mu)^2\right]^{1/2}$$

2.1.7. Energy

The Energy is derived by using Gray Level Co-occurrence Matrix (GLCM). The GLCM computes the matrix depending upon our design and with required resolution factor. Then it gives energy by Squaring and summing the elements of GLCM.

Energy returns the sum of squared elements in the GLCM. Range = [0 1]. Energy is 1 for a constant image

$$\text{Energy} = \sum_{i=1}^{n}(x-\mu)^2$$

Energy is also known as uniformity, uniformity of energy, and angular second moment.The above features are calculated for the affected and not affected malaria images and dataset is use to train the classifiers

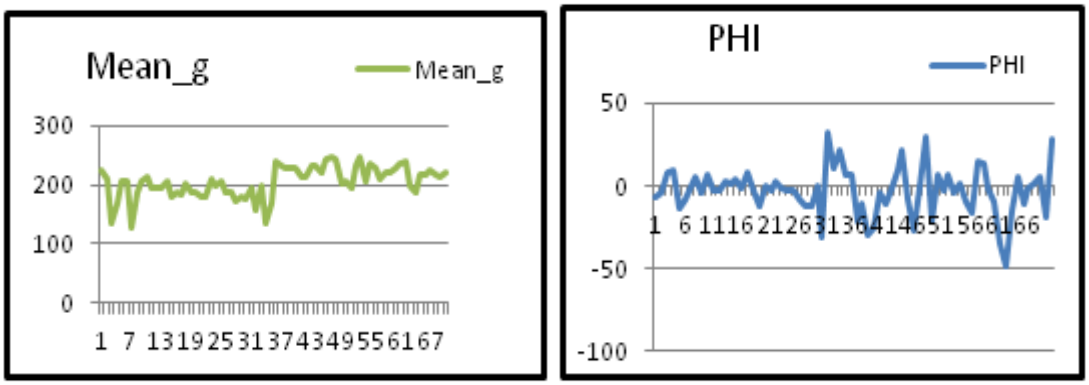

Fig1. *Showing graph of features with respect to images*

2.2 Classification Technique

For the classification three different classifiers utilized. So that we compare the performance of classifier and to decide which classifier gives best results. Here utilized classifiers are as follows

1. Neural Network
2. Support Vector Machine

2.2.1. Neural Network

ANNs are popular machine learning algorithms that are in a wide use in recent years. Multilayer Perception (MLP) is the basic form of ANN, which is a neural network that updates the weights through back propagation during the training. Pattern recognition Network (PRN) and Convolution Neural Network (CoNN) are the other variations in neural networks, which are recently, became popular in texture classification [10, 12].Pattern recognition Network (PRN) is derived from Radial Basis Function (RBF) Network and it has parallel distributed processor that has a natural tendency for storing experiential knowledge. It is predominantly a classifier that maps any input pattern to a number of classifications and can be forced into a more general function approximate. A PNN is an implementation of a statistical algorithm called kernel discriminate analysis in which the operations are organized into a multilayered feed forward network having four layers such as Input layer, Pattern layer, Summation layer, and output layer. Fig.2 demonstrates the architecture of PNN classifier considering a general example of BP and Pulse acting as an input vectors [4, 9].

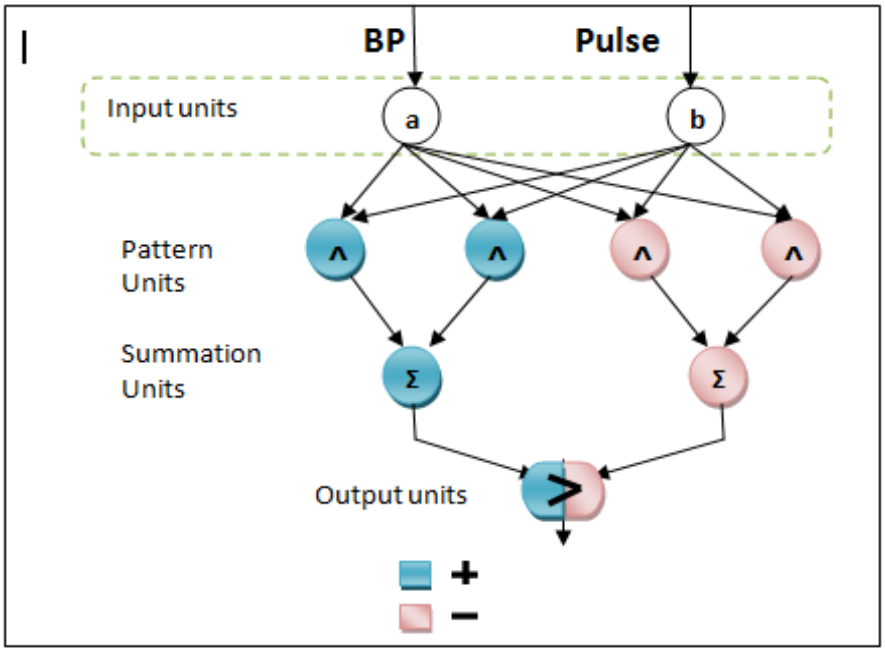

Fig2. *Architecture of Pattern neural network*

2.2.2. Support Vector Machine

Support vector machine (SVM) is a non-linear classifier, which is a newer trend in machine learning algorithm and is popularly used in many pattern recognition problems, including texture classification. In SVM, the input data is non-linearly mapped to linearly separated data in some high dimensional space providing good classification performance. SVM maximizes the marginal distance between different1 classes. The division of classes is carried out with different kernels.SVM is designed to work with only two classes by determining the hyper plane to divide two classes. This is done by maximizing the margin from the hyper plane to the two classes. The samples closest to the margin that were selected to determine the hyper plane is known as support vectors.

Fig.3 Support vector machine Multiclass classification is also applicable and is basically built up by various two class SVMs to solve the problem, either by using one-versus-all or one versus-one. The winning class is then determined by the highest output function or the maximum votes respectively. This leads the multiclass SVM to perform slower than the MLPs. The main advantage of SVM is its simple geometric1 interpretation and a sparse solution. Unlike neural networks, the computational complexity of SVMs does not depend on the dimensionality of the input space. One of the drawbacks of the SVM is the large number of support vectors used from the training set to perform classification task. However, SVM is still considered to be powerful classifier, soon to be replacing the ANNs.

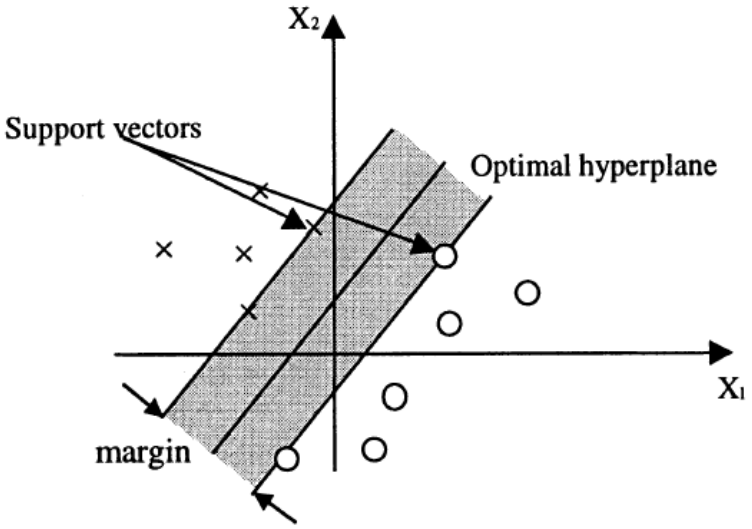

Fig3. *Shows the support vector machines concept*

3. RESULTS

The performance of classifier is defined by the feature used to train the classifier. The results for the experiment are given in table (1) and (2). For the malaria images database the result obtained are as follows.

Table1. *Shows accuracy of algorithm for different methods of classification*

Method	**Neural Network**	**SVM**
Accuracy	78.53%	98.25%

Table2. *Shows no of accurate decision made by an algorithm for different methods of classification*

Type of images	**No. of samples**	**No. of Correctly Identified images**		**%Accuracy**	
		NN	**SVM**	**NN**	**SVM**
Affected	35	30	34	80	90
Not-Affected	35	29	35	77	100

In total, 70 images of blood cells were classified into categories of affected by malaria or unaffected. The performance of system is defined by a classifier used with Parameters or feature set of a database. The result for the experiment is given in tables (1) (2) which show the accuracy of algorithm for different classifier in terms of percentage.

The use of parameter extracted using GLCM and Other method work good with Neural Network better and best by using SVM as classifier.

4. CONCLUSION

This paper addresses how the identification of malaria diseases is possible using image processing by effectively analyzing various parameter of blood cell image by using Phase of Image, Mean Of green plane GLCM as Energy and other like Skewness, Kurtosis, Standard Deviation. The experimental results indicate that the proposed approach is a valuable approach, which can be significantly support an accurate identification of malaria diseases in a little computational effort. There can be mistake in counting manually the number of RBC & WBC (process of Giemsa) as the boundaries are not clearly defined or visible which lead us to the error in wrong decision. So to solve this problem the developed algorithm be more helpful the other techniques. As this system can meet the real time application requirements, so we can easily have the standalone working version of this system.

REFERENCES

[1] F.Sadeghian, Z. Seman and A. R. Ramli, .A Framework for White Blood Cell Segmentation in Microscopic Blood Images Using Digital Image Processing. Biological Procedures Online, vol. 11, no. 1, pp. 196-206, Dec. 2009.

[2] Shiff, C., 2002. Integrated approach for malaria control. Clin. Microbiol. Rev.15, 278–293.

[3] World Health Organization What is malaria? Factssheetno94. http://www.who.int/ mediacentrefactsheetsfs094/en/./factshee. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp 68–73.

[4] V. V. Makkapati and R. M. Rao, .Segmentation of malaria parasites in peripheral blood smear images, Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2009, pp. 1361-1364, Apr. 2009.

[5] S. Raviraja, G. Bajpai and S. Sharma, .Analysis of Detecting the Malaria Parasite Infected Blood Images Using Statistical Based Approach IFMBE Proceedings, 3rd Kuala Lumpur International Conference on Biomedical Engineering 2006, vol. 15, part 12, pp. 502-505, 2007.

[6] C. Di Ruberto, A. Dempster, S. Khan and B. Jarra, .Segmentation of blood images using morphological operators., Proceedings of 15th International Conference on Pattern Recognition Barcelona, Spain, vol. 3, pp. 3401, 2000.

[7] S. S. Savkare , S. P. Narote, "Automatic Detection of Malaria Parasites for Estimating Parasitemia", International Journal of Computer Science and Security (IJCSS), Volume (5), Issue (3), 2011, Page 310-315

[8] Vishnu V. Makkapati and Raghuveer M. Rao .Segmentation of malaria parasites in peripheral blood smear images. IEEE 2009.

[9] S. Raviraja1, Gaurav Bajpai1 and Sharma S .Analysis of Detecting the Malaria Parasite Infected Blood Images Using Statistical Based Approach., Proceedings 15, pp. 502-505, 2007

[10] Sio, W.S.S, et al: Malaria Count: An image analysis-based program for the accurate determination of parasitemia. Journal of Microbiological Methods. 2006.

[11] S.P.Premaratnea, N. D. Karunaweerab, and S. Fernandoc, .A Neural Network Architecture for Automated Recognition of Intracellular Malaria Parasites in Stained Blood Films., 2003. [9] C. J. Janse and P. H. Van Vianen, .Flow cytometry in malaria detection., Methods Cell. Biol. 42 Pt. B: 295.318, 1994.

[12] [C. Pan, X. Yan and C. Zheng, .Recognition of Blood and Bone Marrow Cells using Kernel-based Image Retrieval., IJCSNS International Journal of Computer Science and Network Security, vol.6 no.10, October 2006.

[13] S. K. Lee, C-S. Lo, C-M. Wang and P-C. Chung, .A Computer- Aided Design Mammography Screening System for Detection and Classification of Micro calcifications., International Journal of Medical Informatics, vol. 60, pp. 29-57, 2000

[14] J. Angulo, G. Flandrin, "Automated detection of working area of peripheral blo smears using mathematical morphology", U. S. National Library of Medicine, Analytical Cellular Pathology 25(1), pp 39-47, 2003

[15] T. Markiewicz, S. Osowski, "Data mining techniques for feature selection in blood cell recognition", European Symposium on Artificial Neural Networks, Bruges (Belgium), 26-28 April, pp 407-412, 2006

[16] G. Diaz et al., "A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images", Journal