Enhanced Identification of Malaria Parasite using Different Classification Algorithms in Thick Film Blood Images

Miss. Shruti Annaldas[1], Prof. S.S.Shirgan[2], Prof. V.R.Marathe[3]
[1]M.E.(ENTC) NBNSCOE, Solapur, [2]Dept. of ENTC NBNSCOE, Solapur, [3]Dept. of ENTC NBNSCOE, Solapur,
[1]shruti_annaldas@yahoo.com, [2] ssshirgan@gmail.co,m[3]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, ANFIS, 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, Adaptive Neuro fuzzy Interface System.

2. Literature Survey

An literature survey summarizes all the relevant literature researched during the course of this project. It presents certain approaches used by many researchers for classification. It also compares the performance of all classifier with other common classifier with same parameters. Finally the best parameters and classifier combination is discussed.

[2] Illustrate a technique for identifying the malaria for blood cell images. This paper involves the counting of Blood cell using an adaptive OTSU thresholding technique. Which use to segment the image and separate the RBC and WBC for Counting? The paper also considers the area of cells to declare severity. The paper uses SVM as Classifier for declaring the result of whether the patient is affected by Malaria or Not. The proposed automated method of segmentation and classification of cell is simple. An approach is proposed to detect red blood cells with consecutive classification into parasite infected and normal cells for estimation of parasitemia. The extraction of red blood cells achieves a reliable performance and the actual classification of infected cells. Sensitivity of system is 93.12%, and Specificity is 93.17%. Shape based and statistical features are generated for classification. The features are selected for recognition of two classes only. This approach leads to the high specialization of each classifier and results in an overall increase in accuracy.

Makkapati and Rao [6] explored the segmentation for HSV color space. A scheme presented in [4] is based on HSV color space that segments Red Blood Cells and parasites by detecting dominant hue range and by calculating optimal saturation thresholds. Methods those are less computation-intensive than existing approaches are presented to remove artifacts. The scheme is evaluated using images taken from Leishman-stained blood smears. Sensitivity of the scheme was found to be 83%. The method operates in HSV space and is dynamic in the sense that relevant thresholds are determined from the statistics of the given image rather than keeping them fixed for all images. Schemes determine optimal saturation thresholds to segment RBCs and chromatin dots that are robust with respect to the color variability encountered. The work in [4] illustrates the use of color image processing techniques.

Raviraja and et al. [7] introduces a blood image processing for detecting and classifying malarial parasites in images of Giemsa stained blood slides, in order to evaluate the parasitamia of the blood. To detect the red blood cells that are infected by malarial parasites, statistical based approach is used. To separate automatically the parasites (trophozoites, schizonts and gametocytes) from the rest of an infected blood

image, color, shape and size information are used and later the image is compared with infected images after transformation of image by scaling, shaping to reconstruct

the image. The images returned are statistically analysed and compare to generate a mathematical base. Also the evaluation of the size and shape of the nuclei of the parasite is also considered.

Ruberto et al. [8] introduces morphological approach to cell image segmentation more accurate than the normal watershed based algorithm. The used non-flat disk-shape structuring element enhanced the roundness and compactness to improving the accuracy of normal watershed based algorithm whereas flat disk-shape structuring element to separate overlapping cells. These methods make use of knowledge of the RBC structure that is not used in existing watershed based algorithm.

Sadeghian et al. [9] demonstrated a framework for segmenting white blood cells using digital image processing. This grey level image processing scheme has divided into two parts, first, nucleus segmentation based on morphological analysis, and then cytoplasm segmentation is based on pixel-intensity thresholding (Zack thresholding).

In [8] a scheme based on RGB color space that segments Red Blood Cells and parasites by detecting dominant hue range and by calculating optimal saturation thresholds is presented. Methods that are less computation intensive than existing approaches are proposed to remove artifacts. The scheme is evaluated using images taken from Leishman-stained blood smears. Sensitivity of the scheme is found to be 83%. Automated image analysis-based software. Malaria Count. for parasitemia determination, i.e. for quantitative evaluation of the level of parasites in the blood, has been described in [8]. The presented system is based on the detection of edges representing cell and parasite boundaries. The described technique includes a pre-processing step, edge detection step, edge linking, clump splitting, and parasite detection.

3. METHODOLOGY

3.1 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.

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:

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

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.

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. Mean of Green Plane (M_g)

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

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

3. 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

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

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.

4. 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

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

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.

5. 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

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

6. Energy

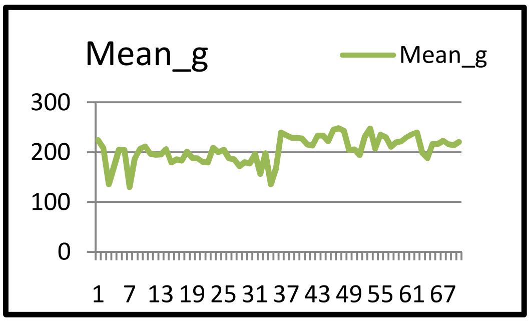
The Energy is derived by using Gray Level Co-occurrence Matrix (GLCM). The GLCM computes the matrix depending upon our design and with required resolution factor. Then it gives energy by Squaring and summing the elements of GLCM.

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.



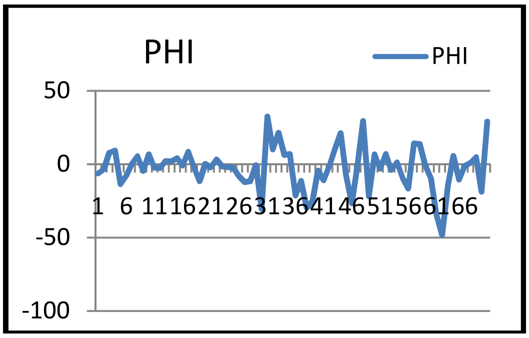


Figure (1): Showing graph of features with respect to images.

3.2 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
3. Adaptive Neuro Fuzzy Network

3.2,1. 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].

a) *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 approximator. 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].

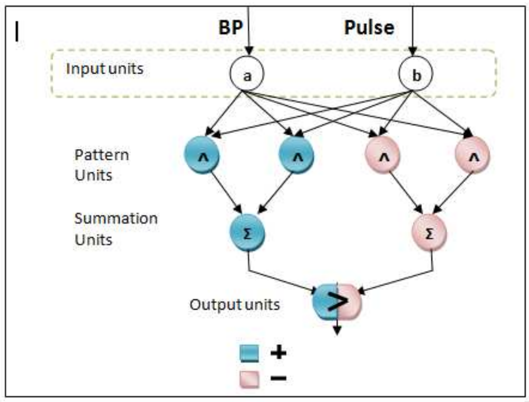


Fig.2 Architecture of Pattern neural network .

3.2.2 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.

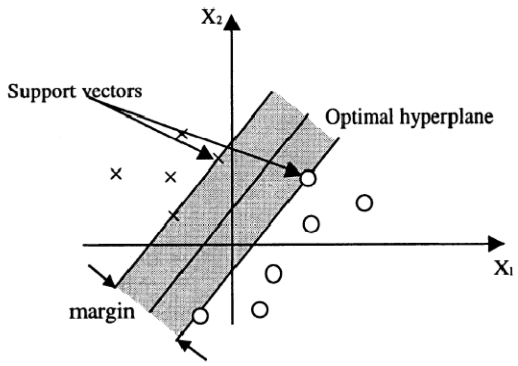


Fig.3 shows the support vector machines concept

Fig.3 Support vector machine Multiclass classification is also applicable and is1 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.

3.2.3. Adaptive Neuro Fuzzy interface System Architecture of ANFIS

The ANFIS is a fuzzy Sugeno model put in the framework of adaptive systems to facilitate learning and adaptation. Such framework makes the ANFIS modeling more systematic and less reliant on expert knowledge. To present the ANFIS architecture, two fuzzy if-then rules based on a first order Sugeno model are considered:

Rule 1: If (x is $A1$) and (y is $B1$) then ($f1 = p1x + q1y + r1$)

Rule 2: If (x is $A2$) and (y is $B2$) then ($f2 = p2x + q2y + r2$)

where x and y are the inputs, A_i and B_i are the fuzzy sets, f_i are the outputs within the fuzzy region specified by the fuzzy rule, p_i, q_i and r_i are the design parameters that are determined during the training process. The ANFIS architecture to implement these two rules is shown in Fig. 4, in which a circle indicates a fixed node, whereas a square indicates an adaptive node. In the first layer, all the nodes are adaptive nodes. The outputs of layer 1 are the fuzzy membership grade of the inputs, which are given by:

$\mathbf{O_i^{1}} = \mu_{\mathbf{Ai}}(x) \qquad \mathbf{i=1, 2}$

$\mathbf{O_i^{1}} = \mu_{\mathbf{Bi-12}}(y) \qquad \mathbf{i=3, 4}$

where μAi (x), $\mu Bi{-}2$ (y) can adopt any fuzzy membership function. For example, if the bell shaped membership.

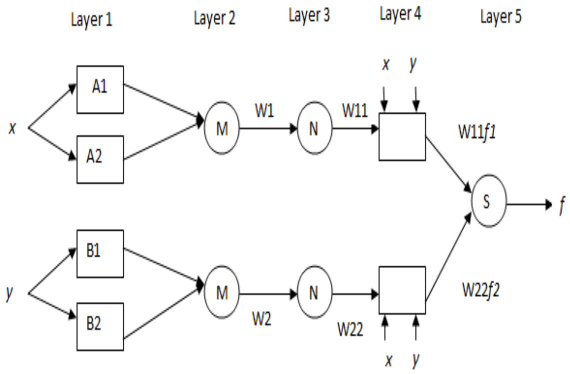


Fig .4 ANFIS Architecture

The output of the ANFIS is calculated by employing the consequent parameters found in the forward pass. The output error is used to adapt the premise parameters by means of a standard back propagation algorithm. It has been proven that this hybrid algorithm is highly efficient in training the ANFIS

4. RESULT

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

Method	Neural Network	ANFIS	SVM
Accuracy	78.53%	89.63%	98.25 %

Table 1. Shows accuracy of algorithm for different methods of classification

5. CONCLUSION AND FUTURE SCOPE

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. The experimental results indicate that method work good with Neural Network better with ANFIS and finally best by using SVM as classifier.

Future scope: We can further increase the performance by creating new set feature which can be well optimized with classifier and which gives best results.

1.1 Improving Performance

- By reducing the redundancy in the dataset.
- By increasing System complexity as using Human interface like the current temperature of patient while running this algorithm.

2. Pre-processing in Feature Extraction

- Pre-processing can be very helpful while extracting features from the blood cell images like removing Noise etc

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] Pallavi T. S., *Detection of Malarial Parasite in Blood Using Image Processing* International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 10, April 2013

[4] Amit Kumar, Choudhary A., Tembhare P. U., Pote C. R., *Enhanced Identification of Malarial Infected Objects using Otsu Algorithm from Thin Smear Digital Images*, International Journal of Latest Research in Science and Technology Vol.1,Issue 2 :Page No159-163 ,July-August(2012)

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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

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

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

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