IJCSN International Journal of Computer Science and Network, Volume 4, Issue 2, April 2015
ISSN (Online) : 2277-5420 www.IJCSN.org
Impact Factor: 0.417

Malaria Disease Identification and Detection Using Different Classifiers

[1] Ashok M. Sutkar, [2] Marathe N. V.

[1] Walchand Collage of Engineering, An Autonomous Institute, E & TC, Maharashtra, India

[2] Walchand Collage of Engineering, An Autonomous Institute, E & TC, Maharashtra, India

Abstract - Malaria is an serious infectious disease which is mainly diagnosed by visual microscopical evaluation of Giemsa stained blood smears. As it poses a serious global health problem, automation of the evaluation process is of high importance. The propose a set of features for distinguishing between non-infected red blood cells and cells infected by malaria parasites and evaluate the performance of these features on the set of red blood cells from the created database. The developed graphical user interface provides all tools necessary for creating a database of red blood cells. This approach proved to deliver good results on images with various qualitative characteristics resulting in only occasional over-segmented cells. The main part of this work is devoted to the extraction of features from the red blood cell images that could be used for distinguishing between infected and non-infected red blood cells. We propose a set of features based on shape, intensity, and texture and evaluate the performance of these features on the red blood cell samples from the created database using receiver operating characteristics. The results have shown that some of the features could be successfully used for malaria detection.

Keywords - *RBC Component, Microscopic images, Parasites, Feature Extraction, NN Classifier, SVM Classifier, ANFIS classifier.*

1. Introduction

Malaria is a serious global disease and a leading cause of morbidity and mortality in tropical and sub-tropical countries. It affects between 350 and 500 million people and causes more than 1 million deaths every year. Yet, malaria is both preventable and curable. Rapid and accurate diagnosis which enables prompt treatment is an essential requirement to control the disease. The most widely used technique for determining the development stage of the malaria disease is visual microscopical evaluation of Giemsa stained blood smears. This process consists of manually counting the infected red blood cells against the number of red blood cells in a slide. The manual analysis of slides is, however, time-consuming, laborious, and requires a trained operator. Moreover, the accuracy of the final diagnosis ultimately depends on the skill and experience of the technician and the time spent studying each slide and it has been observed that the agreement rates among the clinical experts for the diagnosis are surprisingly low. In this context, the development of a mechanism that automates the process of evaluation, quantification and classification in thin blood slides becomes a high priority and the aim of this work was to contribute to improvement upon malaria microscopy diagnosis by removing the reliance on the performance of a human operator for diagnostic accuracy. A number of methods have been proposed for automatic parasite detection in Giemsa stained blood films based on different approaches.

In this work, we propose a set of features and evaluate the performance for a general problem of distinguishing between infected and non-infected red blood cells. Some of these features have already been used in other works but some of them may be new for the problem of malaria parasites detection. Exact definition of these features is provided, including description of the parameters controlling the generation of the transformed images and description of the preprocessing steps performed. Individual sets of features are evaluated on a created dataset of red blood cell samples using ROC curves for different parameters controlling the feature extraction. Evaluation is followed by a discussion on the effects of different preprocessing techniques and possible utilization of these features for more specific problems of distinguishing between different types of malaria parasites.

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 blood of patient's which indicates the degree of its infection. Plasmodium spp. Is a prominent blood parasite which causes malaria. It is nothing but recognition of Plasmodium spp and blood sample visual detection The staining process slightly colorizes the red blood cells (RBCs) but highlights Plasmodium spp parasites, artifacts, and white blood cells (WBC). RBCs in pink color and Giemsa stains nuclei, chromatin in blue tone. It has been shown in several field studies that manual microscopy is not a reliable screening method when performed by non-experts.

Malaria parasites host in RBCs when it enter in blood stream. Manual counting of parasitemia is time consuming and tedious and need experts. 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.

2. Methodology

System architecture used for Malaria parasite detection involves following steps: Image Acquisition, Image Pre-processing, Feature Extraction, Database, Classification and Result. General block diagram of malaria detection system is shown in Figure1.

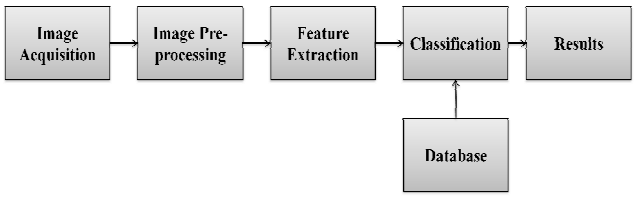


Fig -1: Malaria detection system

2.1 Image Acquisition

Thin blood film images were obtained from laboratory. The samples obtained mostly had low number of parasites in early stages (rings) of their life cycle. The samples were stained using a fast Giemsa protocol to highlight the parasites and Slide images were acquired using a charge coupled device (CCD) camera with different range of magnification. In total 140 cases were analyzed i.e. 50 positive case, 50 negative cases, 20 other cases and 20 invalid cases obtained from laboratories. Some input images are shown in Figure 2 which is malaria parasites in blood sample or infected by malaria & Figure 3 shows normal blood sample or not infected by malaria.

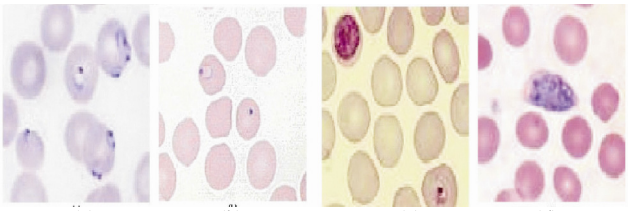


Fig -2: (a) P. Falciparum (b) P. Vivax (c) P. Malariae (d) P. Ovale.

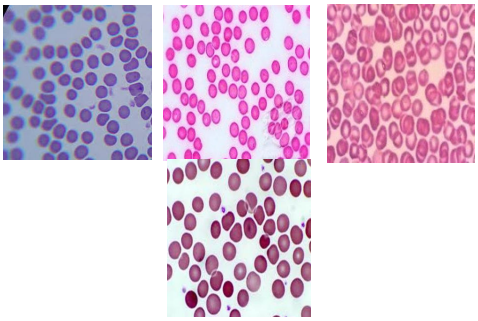


Fig -3: Non-infected input images

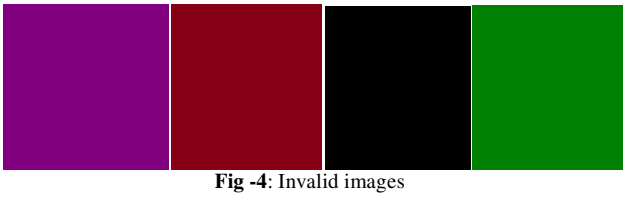


Fig -4: Invalid images

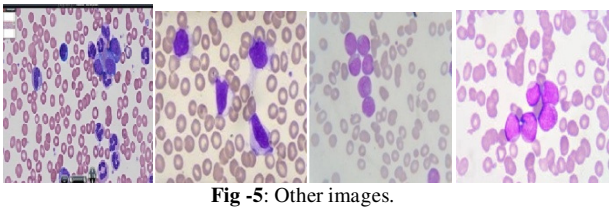


Fig -5: Other images.

2.2 Image Pre-Processing

Image analysis usually starts with a pre-processing stage, which includes operations such as noise reduction. In microscopic image processing, it is usually necessary to perform high degree of noise reduction in an image before performing higher-level processing steps, such as edge detection. The median filter is a non-linear digital filtering technique, often used to remove noise from images or other signals. The idea is to examine a sample of the input and decide if it is representative of the signal. This is performed using a window consisting of an odd number of samples. The values in the window are sorted into numerical order; the median value, the sample in the center of the window, is selected as the output. The oldest

sample is discarded, a new sample acquired, and the calculation repeats.

The median filter considers each pixel in the image in turn and looks at its nearby neighbors to decide whether or not it is representative of its surroundings. Instead of simply replacing the pixel value with the mean of neighboring pixel values, it replaces it with the median of those values. The median is calculated by first sorting all the pixel values from the surrounding neighborhood into numerical order and then replacing the pixel being considered with the middle pixel value. (If the neighborhood under consideration contains an even number of pixels, the average of the two middle pixel values is used.)

2.3 Feature Extraction

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. The features which give predominant difference between normal and infected cells are identified and used for training purpose. The selected features are geometrical, color and statistical based. The mathematical morphology provides an approach to the processing of image based on shape. The set of parameters corresponds to the geometrical features are as follows:

1. Phase -: It computes the complex vector.
2. Mean -: Mean is average or mean value of array.
Here, mean is in r, g, b plane.

$$S_M = \overline{\mathrm{b}} = \sum_{b=0}^{L-1} \mathrm{bP(b)} \qquad (1)$$

3. Skewness -: The coefficient of Skewness is a measure for the degree of symmetry in the variable distribution. It's normal distribution is zero.

$$S_S = \frac{1}{\sigma_b^3} \sum_{b=0}^{L-1} \left(\mathrm{b} - \bar{\mathrm{b}}\right)^3 \mathrm{P(b)} \qquad (2)$$

4. Kurtosis -: The coefficient of Kurtosis is a measure for the degree of peakedness/flatness in the variable distribution. Here, kurtosis normal distribution is 3.

$$S_K = \frac{1}{\sigma_b^4} \sum_{b=0}^{L-1} \left(\mathrm{b} - \bar{\mathrm{b}}\right)^4 \mathrm{P(b)} - 3 \qquad (3)$$

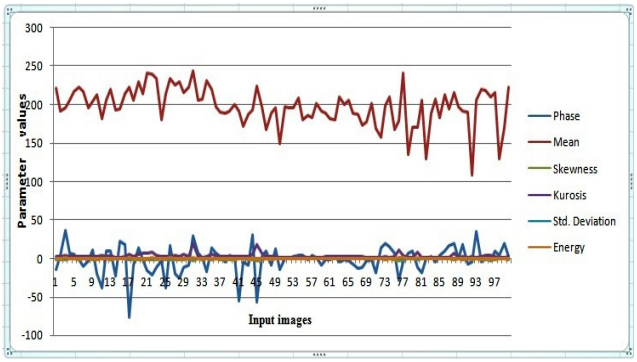
5. Standard Deviation -: Standard Deviations normalize by n-1 where n is the sample size.

$$S_D = \left[\sum_{b=0}^{L-1} (\mathrm{b} - \bar{\mathrm{b}})\right]^{-\frac{1}{2}} \qquad (4)$$

6. Energy -: The Energy is derived by using Gray Level Co-occurrence Matrix (GLCM). Energy is 1 for a constant image.

$$S_N = \sum_{b=0}^{L-1} \mathrm{P(b)}^2 \qquad (5)$$

P(b) is the first-order histogram estimate, Parameter b is the pixel amplitude value, b is the mean of x, σ is the standard deviation of x. L is the upper limit of the quantized amplitude level. 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 and calculated using above equations and then plot the graph.



Graph 1 : Parameter Values

2.4 Classification

The classification techniques utilized are as follows:
1. Neural Network.
2. Support Vector Machine.
3. Adaptive Neuro-Fuzzy Interface System.

These techniques will describe in detail later with their performance aspect. The developed algorithm gives good accuracy with Neural Network and Better with Support Vector Machine and Adaptive Neuro-Fuzzy Interface System

.

1. Neural Network:

Artificial Neural Network (ANN) has been employed together with image processing techniques to automate the assessment of these blood disorders using the morphological features of erythrocytes in the blood. Prior to training, the first necessary step is to preprocess the giemsa stained blood sample images acquired from using a high resolution digital camera mounted on a microscope.

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons.

Neural networks consist of simple elements & they work in parallel. Biological nervous systems inspire these elements. There is cordial relationship between elements which decide the network function. For performing a specific, you can train a neural network by adjusting the values of the connections (weights) between elements. We should adjust and train the neural networks, and give a particular input so that it will leads to a specific target output. Training has been given to Neural networks to perform complex functions in various fields, such as classification, pattern recognition, identification, and control systems and vision, speech. Training can be given to neural networks to solve problems which are troublesome to human beings or conventional computers. ANNs are computational networks which attempt to simulate the networks of neurons.

This simulation is neuron by neuron simulation. A neural network system consists of many simple processing elements with operate in parallel and whose function is decided by connection strengths, network structure, and the processing performed at computing elements modes.

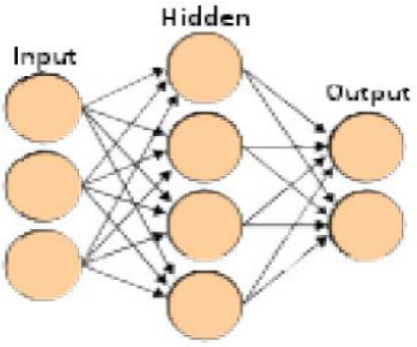


Fig -6: Artificial neural network interconnected groups.

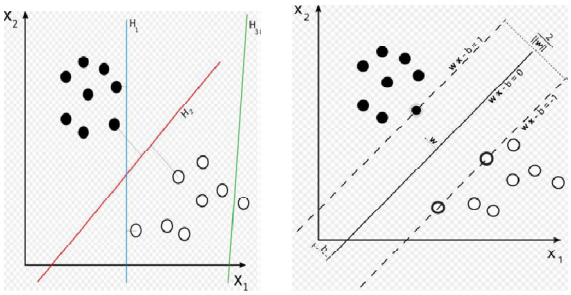
The terminology of artificial neural network has developed from biological model of brain. The ANN processes information in parallel with a large number of processing elements called neurons and uses large interconnected networks of simple and non linear units. Neural networks consist of the connected cells: The neurons. The neuron receives impulses from either input cells or other neurons and performs some kind of transformation of the input and transmits the outcome to other neurons or to output cells. The neural networks are build from layers of neurons connected so that one layer receives input from the preceding layer of neurons and passes the output to the subsequent layer. ANN includes three groups or layers of units such as input, hidden and output. They are interconnected with one another. The input units give raw information which is fed the network.

The activity of input unit and the weights on the connections between the input and the hidden units determine the activity of each hidden unit. The behavior of the output units relies on the weight between output unit and activity of the hidden units. It is interesting as the hidden units are free to build their own representations of the weights between the hidden units and input units determine when hidden unit is active by making proper modification of these weights, a hidden unit can select what it represents below figure 6.

2. Support Vector Machine:

Support vector machine (SVM), which is based on Statistical Learning Theory (SLT), has shown much better performance than most other existing machine learning methods which are based on traditional statistics. Support vector machine is widely used for data analyzing and pattern recognizing. The algorithm was invented by Vladimir Vapnik and the current standard incarnation was proposed by Corinna Cortes and Vladimir Vapnik.

SVM is based on structural risk minimization (SRM) principle rather than empirical risk minimization (ERM) principle which is employed by conventional neural networks. It is originated from statistical learning theory (STL) which aims to learn patterns from a small sample set and has some attractive features such as generalization and high empirical performance. Classifying data has been one of the major parts in machine learning. The idea of support vector machine is to create a hyper plane in between data sets to indicate which class it belongs to. The challenge is to train the machine to understand structure from data and mapping with the right class label, for the best result, the hyper plane has the largest distance to the nearest training data points of any class.



(a) (b)

Fig -7: Hyper Plane

(As we can see from figure 5, H3 does not separate the two classes while H1 separate the two class with a small margin, only H2 gives a maximum margin between two classes, therefor it's the right hyper plane used by support vector machine)

However, instead define a function for the hyper plane itself; we define the margin in between the two classes. From figure 5, we can see that the position of our hyper plane is depend on the value of W.

3. Adaptive Neuro-Fuzzy Interface System

Artificial neural network and fuzzy logic are used in ANFIS' architecture. ANFIS consists of if–then rules and couples of input–output. Also for ANFIS training, learning algorithms of neural network are used. A Neuro-fuzzy approach as a combination of neural networks and fuzzy logic has been introduced to overcome the individual weaknesses and to offer more appealing features.

Neuro-fuzzy systems Soft computing methods that combine in various ways neural networks and fuzzy concepts.

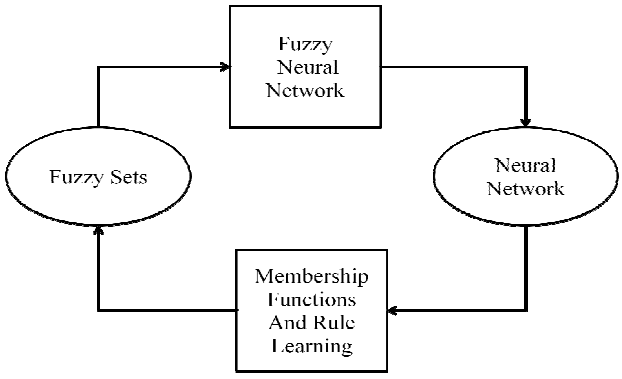


Figure 8: conceptual View of ANFIS

ANN – nervous system – low level perceptive and signal integration and as illustrated in ANN bit

Fuzzy part – represents the emergent "higher level" reasoning aspects "Fuzzification" of neural networks. Endowing of fuzzy system with neural learning features. Co-operative-neural algorithm adapt fuzzy systems

• Off-line – adaptation
• On-line – algorithms are used to adapt as the system operates.

Adaptive Network-based Fuzzy Inference System is Neuro-fuzzy system that can identify parameters by using supervised learning methods and works adaptively it uses Sugeno-type fuzzy system with learning capabilities.

First order model –:
IF x is A_1 AND y is B_1 THEN f_1=p_1x+q_1+r_1
IF x is A_2 AND y is B_2 THEN f_2=p_2x+q_2+r_2

The reasoning mechanism for this model is:
F=$\frac{w_1 f_1 + w_2 f_2}{w_1 + w_2}$ = $\overline{w_1} + \overline{w_2}$

The example illustrates how ANFIS declare the output or how the inputs era utilized to make proper output. Following are the certain fundamentals used by ANFIS

I. Learning algorithm is a hybrid supervised method based on gradient descent and Least-squares.
II. Forward phase: signals travel up to layer 4 and the relevant parameters are fitted by least squares.
III. Backward phase: the error signals travel backward and the premise parameters are updated as in backpropagation.
IV. Fuzzy toolbox Matlab.
V. Mackey-Glass prediction/excellent non-linear fitting and generalization / less parameters and training time is comparable with ANN methods.

Since a wide class of fuzzy controllers can be transformed into equivalent adaptive networks, ANFIS can be used for building intelligent controllers that is, controllers that can reason with simple fuzzy inference and that are able to learn from experience in the ANN style.

Flowchart

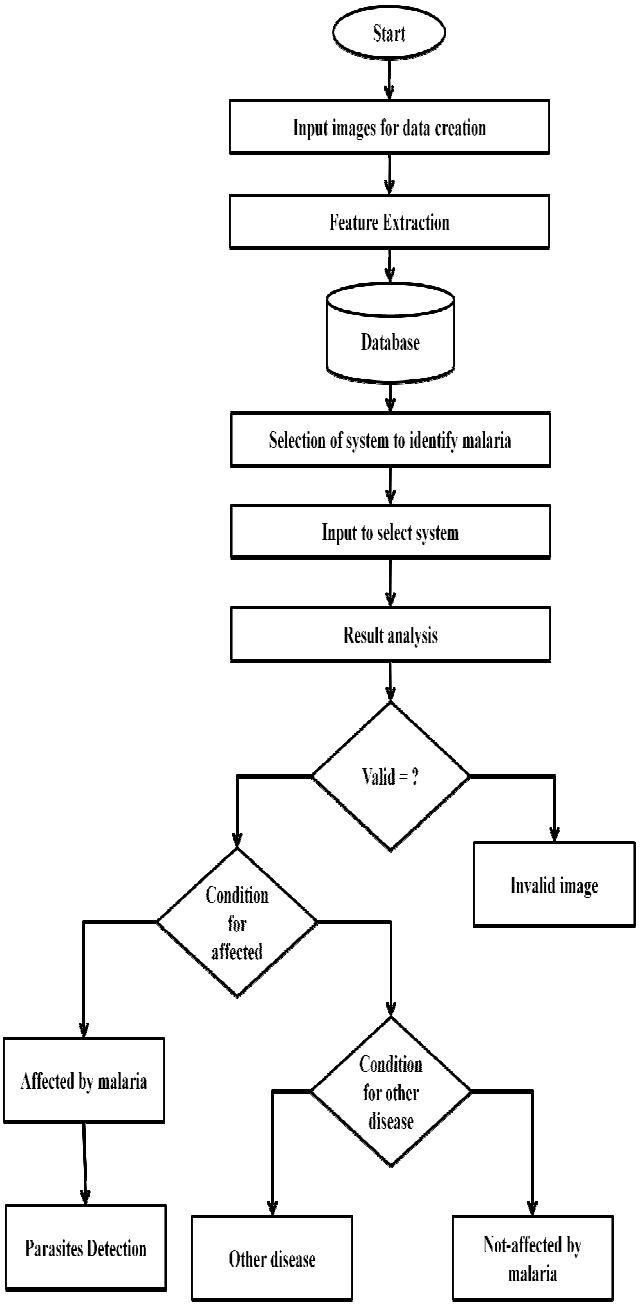
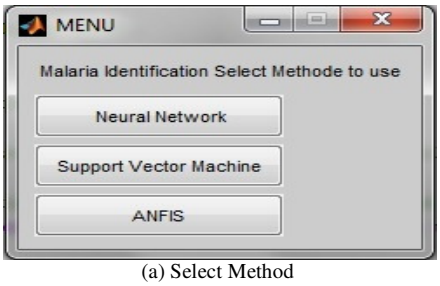


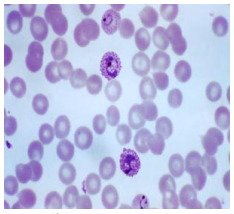
Fig. 9 Flowchart

2.5 Results

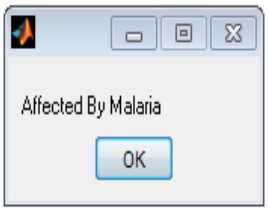
We took 140 images from different laboratories and did the testing. The aim is to distinguish between negative and positive cases of malaria using thin or thick smear blood slide images. It does require minimum supervision of human interference and it enhances the speed of whole process of diagnosis. Final Results of Malaria Disease Identification Using GUI



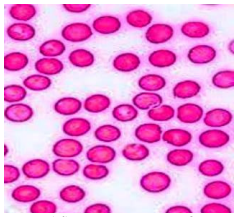
(a) Select Method



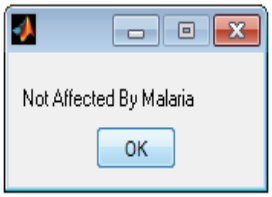
(b) Input Image



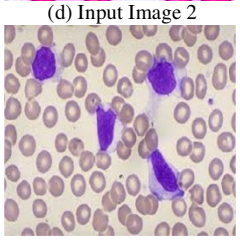
(c) Affected by malaria



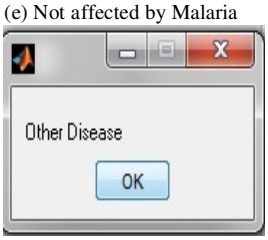
(d) Input Image 2



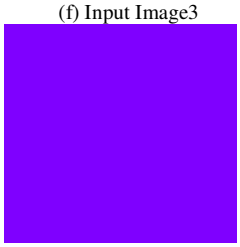
(e) Not affected by Malaria



(f) Input Image3

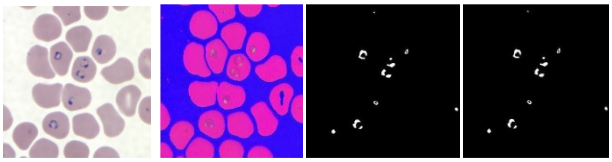


(g) Other disease

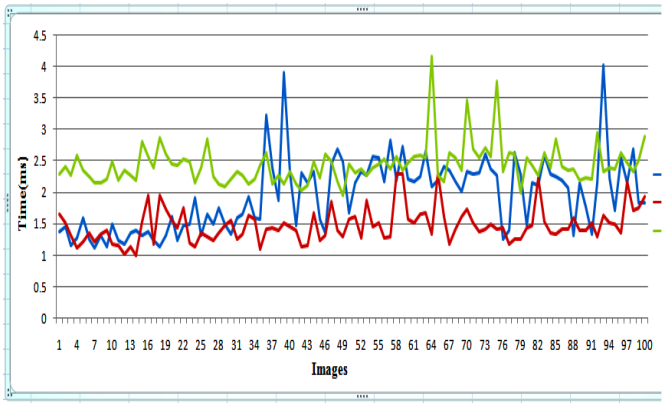


(h) Input Image 4

(i) Invalid Image



(j)Input image (k)HSV image (l)Processed image (m)Detected parasites
Figure 9: (a) selection of classifier (b) Input Image (c) Affected by malaria(d) Input Image 2 (e) Not affected by Malaria (f) Input Image3 (g) Other disease (h) Input Image 4 (i) Invalid Image (j)Detected parasites (k)HSV image (l)Input image (m)Processed image.



Graph 2: Computational time

$$Accuracy = (TP+TN) / (TP+FN+FP+TN)$$

$$Sensitivity = TP / (TP + FN)$$

Table 1

	SVM	NN	ANFIS
TP	85	67	79
TN	38	38	38
FP	5	23	11
FN	12	12	12

Table 2: Classifier Comparison

	SVM	NN	ANFIS
Accuracy(%)	87.85	75	83.57
Sensitivity(%)	87.629	84.81	86.81

3. Conclusions

This project addresses how the identification of malaria diseases is possible using image processing by effectively analyzing various parameter of blood cell image by using GLCM as Energy and other like Skewness, Kurtosis, and 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. Support vector machine gives good accuracy as compared to neural network.

References

[1] Yashasvi Purwar, Sirish L Shah, Gwen Clarke, Areej Almugairi and Atis Muehlenbachs, "Automated and unsupervised detection of malarial parasites in microscopic images", Purwar et al. Malaria Journal 2011, 10:3641,Department of Chemical and Materials Engineering, University of Alberta, Edmonton, Canada.

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

[3] Sriram R., Meenalochani Chandar and Kota Srinivas, "Computer Aided Malarial Diagnosis for JSB Stained White Light Images Using Neural Networks", Inte-rnational Journal of Advanced Research in Computer Science and Software Engi-neering, Volume 3, Issue 8, August 2013, CSIR-CSIO, Chennai, India.

[4] S.Abdul Nasir, M.Y.Mashor and Z.Mohamed, "Segmentation Based Approach for Detection of Malaria Parasites Using Moving K- Means Clustering", Department of Medical Microbiology & Parasitology, School of Medical Sciences, Health Campus, Universiti Sains Malaysia, IEEE EMBS International Conference on Biomedical Engineering and Sciences, Langkawi, 17th - 19th December 2012.

[5] Saowaluck Kaewkamnerd, Apichart Intarapanich, Montri Pannarat, Sastra Chaotheing, Chairat Uthaipibull and Sissades Tongsima, "Detection and Classification Device for Malaria Parasites in Thick- Blood Films" The 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications 15-17 September 2011, Prague, Czech Republic.

[6] 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", IEEE Trans.on Applied DSP and VLSI Research Group University of Westminster London, United Kingdom, 2012.

[7] Isha Suwalka, Ankit Sanadhya, Anirudh Mathur and Mahendra S Chouhan, "Identify Malaria Parasite Using Pattern Recognition Technique" Geetanjali Institute of Technical Studies Dabok, Udaipur, India,2011.

[8] D. Anggraini, A. S. Nugroho, C. Pratama, et al., "Automated status identification of microscopic images obtained from malaria thin blood smears," in 2011 International Conference on Electrical Engineering and Informatics (ICEEI), Bandung, Indonesia, 2011.

[9] Matthias Elter, Erik Haßlmeyer and Thorsten Zerfaß Fraunhofer, "Detection of malaria parasites in thick blood films", Institute for Integrated Circuits IIS Image Processing and Medical Engineering Department Am Wolfsmantel 33, 91058 Erlangen, Germany, 33rd Annual International Conference of the IEEE EMBS Boston, Massachusetts USA, August 30 - September 3, 2011.

[10] Amit Kumar, Manjunatha M, J Chatterjee and Ajoy K Ray, "Segmentation of Blood Smear Images using Normalized Cuts for Detection of Malarial Parasites" Dept. of Electronics & Electrical Communication Engineering, Annual IEEE India Conference (INDICON) 2010.A. A. Name, and B. B. Name, Book Title, Place: Press, Year.