IJCAT International Journal of Computing and Technology, Volume 1, Issue 6, July 2014
ISSN: 2348 - 6090
www.IJCAT.org

Malaria Disease Identification and Analysis Using Image Processing

[1] Sneha Chavan, [2] Dr. Manoj Nagmode

[1, 2] Dept. of Electronics & Telecommunication, University of Pune,
MIT College of engineering, Kotharud
Pune, Maharashtra, India.

Abstract - Malaria is an infectious disease. According to World Health Organization (WHO) it affects one million people face deaths each year. There are various methods to analyses malaria using manual microscopy is considered to be However it requires manual assessment, this diagnostic method is dispose to human error and time consuming, even in experienced hands. The this study is to improve an unsupervised, sensitive and sturdy which lessens human trust as well as reduces material cost and is, so, more method compatible in applying diagnostic criteria.

***Keywords* - RBC Component, Microscopic images, Parasites, Feature Extraction, SVM Classifier, ANN Classifier.**

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 blood of patient's which indicates the degree of its infection. Plasmodium spp. Is a prominent blood parasite which causes malaria. It makes an adverse effect approximately 200 to 300 million people every year and nearly 3 million people face fatal death every year. Keen analysis and appropriate medicines of it is very essential. So for proper treatment and medicines should be started on time. It is very important for patient's health point of view. Giemsa staining is used for detection of malaria disease. 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 two 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.

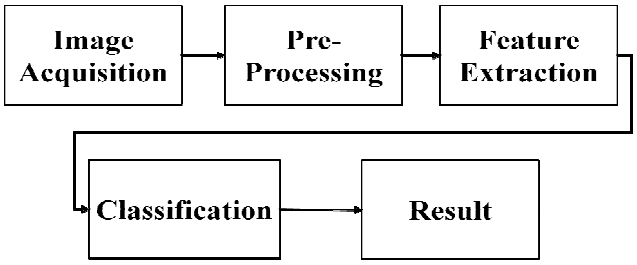
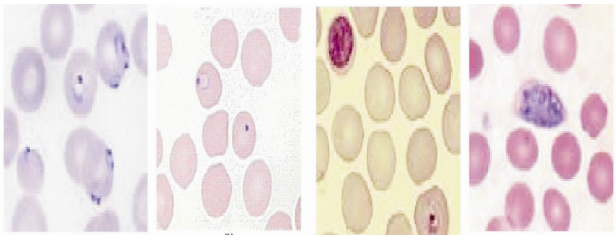


Fig. 1 General block diagram of malaria detection.

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 were initially examined by haematopathologists with expertise in malaria diagnosis. Slide images were acquired using a charge coupled device (CCD) camera with different range of magnification. In-total 50 cases were analysed i.e. 25 positive cases and 25 negative cases obtained from Ashwini hospital 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. The top of this paragraph illustrates a sub-subheading.



(a) (b) (c) (d)

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

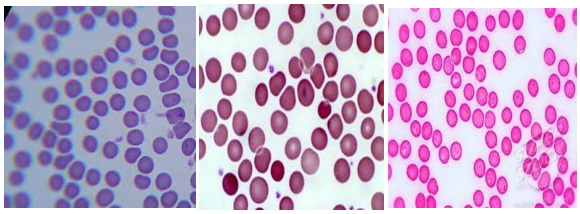


Fig. 3 Non infected input images.

2.2 Image Pre-Processing

The purpose of pre-processing is to remove unwanted objects and noise from the image to facilitate image segmentation into meaningful regions. The steps required to carry out image pre-processing were implemented on low resolution images are as follows: Load colored (RGB) or gray scale image, the colored image is converted to gray scale image. The contrast of the gray scale image is enhanced using local histogram equalization to enhance the visibility of the parasites and RBC.

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:

$$\text{Mean -: } S_M = \overline{b} = \sum_{b=0}^{L-1} bP(b) \quad (1)$$

$$\text{Skewness -: } S_S = \frac{1}{\sigma_b^3}\sum_{b=0}^{L-1}(b-\bar{b})^3 P(b) \quad (2)$$

$$\text{Kurtosis } -\text{: } S_K = \frac{1}{\sigma_b^4}\sum_{b=0}^{L-1}(b-\bar{b})^4 P(b) - 3 \quad (3)$$

$$\text{Standard deviation}-\text{: } S_D = \left[\sum_{b=0}^{L-1}(b-\bar{b})\right]^{-\frac{1}{2}} \quad (4)$$

$$\text{Energy}-\text{: } S_N = \sum_{b=0}^{L-1} P(b)^2 \quad (5)$$

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.

Table 1: Feature Extracted values

Images	Phase	Mean	Skewness	Kurtosis	Standard Deviation	Energy
1	-14.152	191.54	-0.5826	2.3798	0.0822	0.3070
2	-5.1976	187.54	-0.6091	2.3606	0.1237	0.2101
3	21.173	233.85	-1.2430	4.2072	0.0620	0.5844
4	-7.7442	161.88	-0.0469	1.7473	0.1016	0.2447
5	6.3109	216.59	-0.5131	3.0235	0.0752	0.4382
6	-1.6942	223.05	-0.8626	3.0001	0.0864	0.4865
7	-11.376	206.67	-0.0013	2.5529	0.0932	0.2890
8	-3.4268	195.54	-0.5136	3.0233	0.1924	0.3148
9	-25.160	229.09	-0.9576	3.2825	0.0763	0.5010
10	21.298	198.04	-0.6258	2.1234	0.1279	0.2553
11	-2.8634	133.91	-0.4669	1.8473	0.2356	0.1547
12	-32.003	218.82	-1.0028	3.6523	0.1115	0.4520
13	6.4918	172.63	-0.4922	1.8365	0.2564	0.2800
14	13.877	184.33	-0.7551	2.4295	0.1910	0.2645
15	95.774	222.70	-1.6234	6.4954	0.1408	0.4526
16	62.521	222.38	-0.9417	3.5287	0.1167	0.4094
17	16.859	150.16	-0.1509	2.8820	0.0899	0.2670
18	46.368	218.83	-0.8857	3.9442	0.0984	0.4035
19	17.722	154.51	-0.2869	2.7399	0.1641	0.1299
20	11.738	134.64	-0.4183	1.9369	0.1212	0.2819

Above table 1 shows the extracted features using formulas. There are first ten images gives the infected images value and another ten images gives the non infected input images value. Using this value we can create database for classification purpose.

2.4 Classification

The classification techniques utilized are as follows:
1. Support Vector Machine.
2. Artificial Neural Network.

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.

2.4.1 Support Vector Machine

SVM has been used for the regression and classification of cells. The SVM is a very powerful solution to the classification problems. The main advantage of the SVM network used as a classifier is its extremely powerful learning procedure, , leading to the global minimum of the defined error function and very good generalization ability. Linear SVM is a linear discriminative classifier working on the principle of maximum margin between two classes. The decision function of the N-dimensional input vector x for K-dimensional feature space (K>N) is defined as D(x) = wT (x) +b through the use of function (x). Where b as the bias weight, w as the weight vector of network w=[w1 , w2, ..., wk]T, (x) =[1(x), 2(x), . . . , K(x)]. All values of weights have been arranged in decreasing order and only the most important have been selected for each pair of classes and then used in the final classification system.

The learning of the SVM network working in the classification mode is aimed at the maximization of the separation margin between two classes. Simple classification algorithm is proposed that classifies points by assigning them to the closer of two parallel planes (in input or feature space). Standard support vector machines (SVMs), which assign points to one of two half spaces. SVM classifier is used for classification of normal and infected cells. Machine Learning is considered as a subfield of Artificial Intelligence and it is concerned with the development of techniques and methods which enable the computer to learn. In simple terms development of algorithms which enable the machine to learn and perform tasks and activities. Machine learning overlaps with statistics in many ways. Over the period of time many techniques and methodologies were developed for machine learning tasks.

In another words, Support Vector Machine (SVM) is a regression and classification prediction tool that uses machine learning theory to enhance predictive accuracy while automatically avoiding over-fit to the data. SVM is more accurate than sophisticated neural network due to pixel map as input though neural network has elaborated features in a handwriting recognition task. It is also being used for many applications, such as face analysis, hand writing analysis, and so forth, especially for regression and pattern classification based applications. SVM has many promising features such as better empirical performance. For conventional neural networks the Structural Risk Minimization (SRM) principle, which has been shown to be superior to traditional Empirical Risk Minimization (ERM) principle, ERM minimizes the error on the training data where as SRM minimizes an upper bound on the expected risk. Using the difference between SRM and ERM, SVM achieves the goal in statistical learning.

Why SVM?

Firstly working with neural networks for supervised and unsupervised learning showed good results while used for such learning applications. MLP's uses feed forward and recurrent networks. Multilayer perceptron (MLP) properties include universal approximation of continuous nonlinear functions and include learning with input-output patterns and also involve advanced network architectures with multiple inputs and outputs.

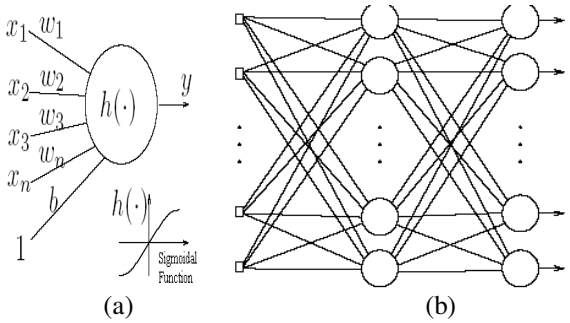


Figure 4:a] Simple Neural Networ b] Multilayer Perceptron.

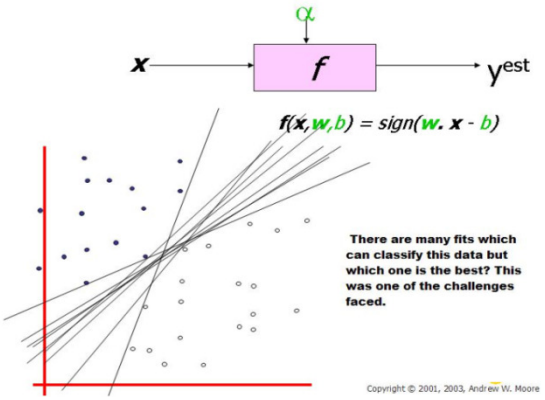


Fig. 5 : Hyperplanes.

These are simple visualizations just to have a overview as how neural network looks like. There can be some issues noticed. Some of them are having many local minima and also finding how many neurons might be needed for a task is another issue which determines whether optimality of that NN is reached. Another thing to note is that even if the neural network solutions used tends to converge, this may not result in a unique solution. Now let us look at another example where we plot the data and try to classify it and we see that there are many hyper planes which can classify it. But which one is better? Here we see that there are many hyper planes which can be fit in to classify the data but which one is the best is the right or correct solution. The need for SVM arises. Note the legend is not described as

they are sample plotting to make understand the concepts involved.

From above illustration, there are many linear classifiers (hyper planes) that separate the data. However only one of these achieves maximum separation. The reason we need it is because if we use a hyper plane to classify, it might end up closer to one set of datasets compared to others and we do not want this to happen and thus we see that the concept of maximum margin classifier or hyper plane as an apparent solution. The next illustration gives the maximum margin classifier example which provides a solution to the above mentioned problem. Note the legend is not described as they are sample plotting to make understand the concepts involved.

Expression for Maximum margin is given as

$$\text{margin} \equiv \arg\min_{\mathbf{x}\in D} d(\mathbf{x}) = \arg\min_{\mathbf{x}\in D} \frac{|\mathbf{x}\cdot\mathbf{w}+b|}{\sqrt{\sum_{i=1}^{d} w_i^2}}$$

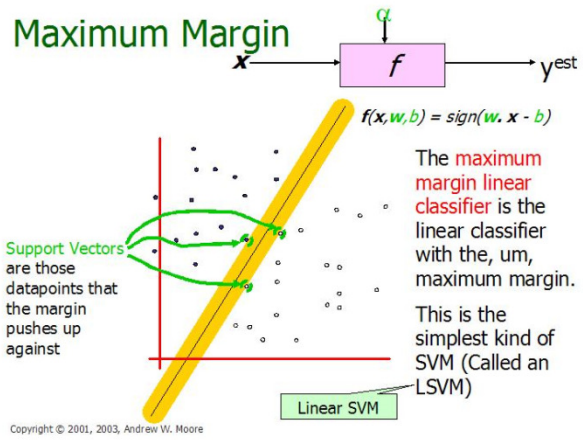


Fig. 6 Illustration of Linear SVM.

The above illustration is the maximum linear classifier with the maximum range. In this context it is an example of a simple linear SVM classifier. Another interesting question is why maximum margin? There are some good explanations which include better empirical performance. Another reason is that even if we've made a small error in the location of the boundary this gives us least chance of causing a misclassification. The other advantage would be avoiding local minima and better classification. Now we try to express the SVM mathematically and for this tutorial we try to present a linear SVM. The goals of SVM are separating the data with hyper plane and extend this to non-linear boundaries using kernel trick. For calculating the SVM we see that the goal is to correctly classify all the data. For mathematical calculations we have,

[a] If Yi= +1;

[b] If Yi= -1; $wxi + b \leqslant 1$

[c] For all i; $yi\ (wi + b) \geqslant 1$

In this equation x is a vector point and w is weight and is also a vector. So to separate the data [a] should always be greater than zero. Among all possible hyper planes, SVM selects the one where the distance of hyper plane is as large as possible. If the training data is good and every test vector is located in radius r from training vector. Now if the chosen hyper plane is located at the farthest possible from the data. This desired hyper plane which maximizes the margin also bisects the lines between closest points on convex hull of the two datasets. Thus we have [a], [b] & [c].

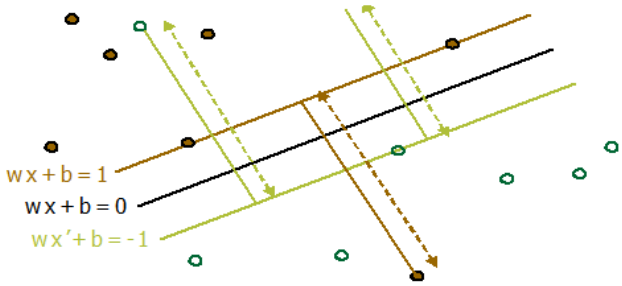


Fig. 7: Representation of Hyper planes.

Distance of closest point on hyper plane to origin can be found by maximizing the x as x is on the hyper plane. Similarly for the other side points we have a similar scenario. Thus solving and subtracting the two distances we get the summed distance from the separating hyperplane to nearest points. Maximum Margin = M = 2 / ||w||

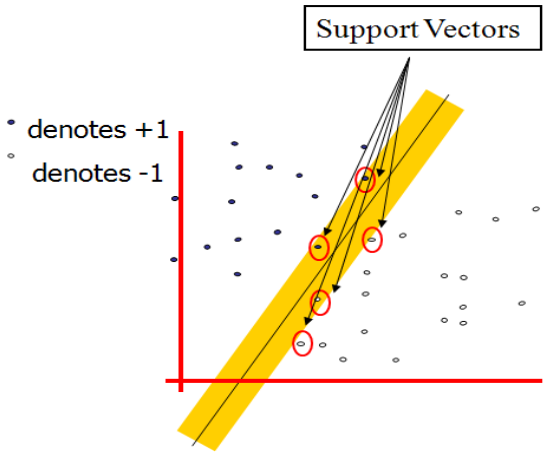


Fig. 8 : Representation of Support.

Now maximizing the margin is same as minimum. Now we have a quadratic optimization problem and we need to solve for w and b. To solve this we need to optimize the quadratic function with linear constraints. The solution

involves constructing a dual problem and where a Langlier's multiplier αi is associated. We need to find w and b such that Φ (w) =½ |w'||w| is minimized;

And for all {(xi, yi)}: yi (w * xi + b) ≥ 1.

Now solving: we get that w =Σαi * xi; b= yk- w *xk for any xk such that αk¹ 0

Now the classifying function will have the following form:

$$f(x) = \Sigma \alpha i\ yi\ xi\ * x + b.$$

2.4.2 Artificial Neural Network

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.

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 built 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.

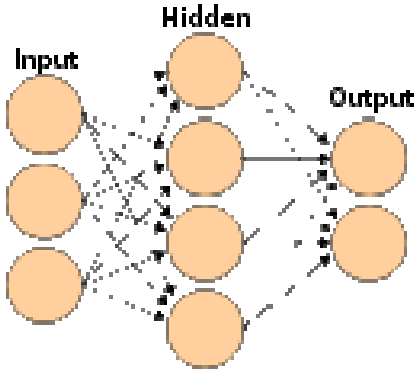


Fig. 9 : Artificial neural network interconnected groups.

The basic principles of the artificial neural network depends on five assumptions, as follows

1. The condition of being active of a neuron (ANN) is all-or-nothing.
2. As a neural tobe excited a certain fixed number of synapses are greater than one within a given interval of neural addition.
3. The synaptic delay is nothing but the significant delay within the neural system.
4. The excitation of the neuron at that time is prevented absolutely by the activity of the inhibitory synapse.
5. The structure of the interconnection network remains intact.

An ANN is typically defined by three types of parameters:

1. The activation function that converts a neuron's weighted input to its output activation.
2. The interconnection pattern between different layers of neurons.
3. The learning process for updating the weights of the interconnections.

For designing the network, follow the standard steps for designing neural networks to solve problems in four application areas: pattern recognition, time series analysis, clustering, and function fitting. The work flow for any of these problems has six primary steps. If the problem occurs in the work flow, the following six primary steps should be followed.

1. Collect data
2. Create the network,
3. Configure the network,

4. Initialize the weights and biases,
5. Train the network,
6. Validate the network and use the network.

The network object is used by the Neural Network software uses to store all of the information that defines a neural network.

2.5 Results

We took 30 to 40 images from 2 independent 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. When we detect malaria parasites, we find approximately 50-88 % specificity and we get 100% sensitivity.

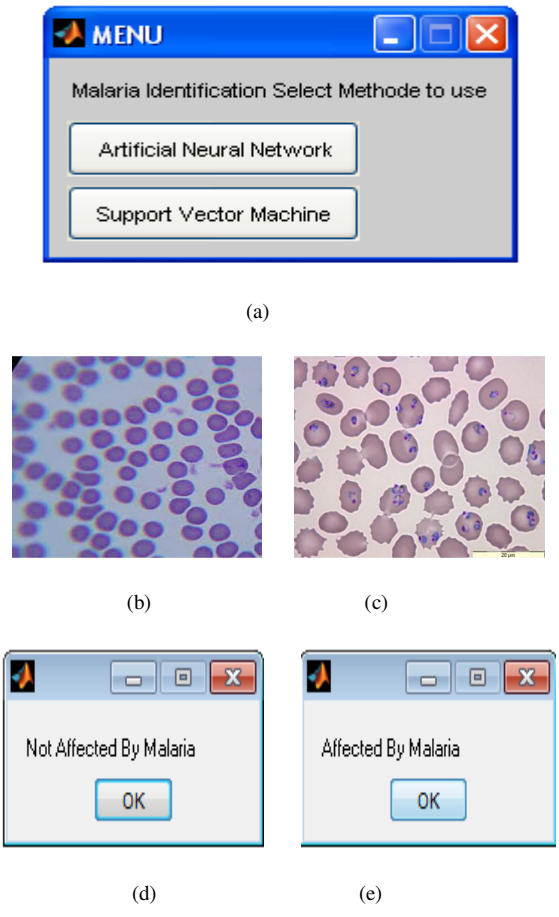


Fig. 10 : (a)Select classifier(b) Non-affected image, (c) Affected Image, (d) Non affected image Output & (e) Affected image output

. Table 2: Computational Time

Algorithms	**Computational Time(sec)**
Artificial Neural Network	6.58
Support Vector Machine	3.852

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.

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 2013.

[3] Amit Kumar, Prof. Arun Choudhary, Prof. P. U. Tembhare and Prof. C. R. Pote, "Enhanced Identification of Malarial Infected Objects using Otsu Algorithm from Thin Smear Digital Images", International Journal of Latest Research in Science and0 Technology ISSN (Online):2278-5299 Vol.1, Issue 2 :Page No159-163, July-August(2012).

[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 Malarial 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] 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.

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