



Computer Aided Malarial Diagnosis for JSB Stained White Light Images Using Neural Networks

Sriram.R*
CSIR-CSIO, Chennai
India

Meenalochani Chandar
CSIR-CSIO, Chennai
India

Kota Srinivas
CSIR-CSIO, Chennai
India

Abstract— This paper focuses on development of sensitive malarial detection system for images of (JSB) stained thick blood slides acquired from conventional light microscopes. Malaria is a life-threatening disease caused by parasites that are transmitted to people through the bites of infected mosquitoes. Light microscopy enables the visualization of malarial parasites in a thick or thin smear of the patient's blood. Automation of the evaluation process in the diagnosis of malaria is of high importance. The proposed system describes the computerized method of image analysis involving three main phases: pre-processing, where the images are corrected for luminance and transformed to a constant color space. A histogram based image segmentation processing where the maximum artefacts and over stained objects are avoided. Finally, Feature extraction along with a multi-layer, feedforward, backpropagation neural network was employed for classifying the objects as parasite/wbc. The proposed method achieves the 91% of sensitivity, 85% of specificity with positive prediction rate 88%.

Keywords— Jaswant-Singh-Bhattacharji (JSB) Stain, Malaria, Microscopic images, Feature Extraction, Artificial Neural network.

I. INTRODUCTION

About 3.3 billion people, half of the world's population, are at risk of malaria [1]. Malaria is caused by the Plasmodium parasites. There are four parasite species that cause malaria in humans: Plasmodium falciparum, Plasmodium vivax, Plasmodium malariae, Plasmodium ovale. P.Vivax and P.falciparum are the most widely distributed and the most common species causing Malaria in India [2]. WHO reports the number of patients tested by microscopic examination has increased to 171 million in 2011 and the global total is dominated by India, which accounted for over 108 million slide examinations in 2011. Unfortunately visual analysis is time consuming, difficult to reproduce and subjective, which results in large inter and intra – observer variability [3]. In regards to these contexts there is a high priority and demand towards automated computerized detection of malarial parasite. The objective of the work is to add to the betterment upon malaria diagnosis by not completely relying on the human operator for diagnostic accuracy.

The conventional staining method includes Leishman/ Giemsa stain (gold standard), fields stain and JSB stain [4]. In India 40% of cases are as result of P.vivax malarial infection. 44.3% are due to P.falciparum malarial infection. 10-15% is due to mixed malarial infection [2, 5]. The microscopic slide examination of JSB stain thick smear blood films are predominantly used as the primary method for decision making. Recent studies reports JSB stain is fairly a rapid method and 92% sensitive with 100% specificity in discriminating the parasite/infected cells and normal cells. Series of publications has been out for computerized detection of malarial parasite with different staining protocols. This work emphasises on detecting parasite using JSB (thick) stain blood films. In Fig.1 the processing phases for analysis of the images are presented.

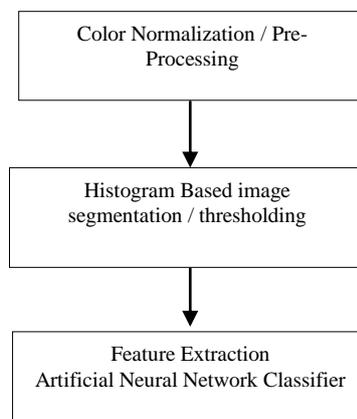


Fig1. Processing phases for Image Analysis

II. MATERIALS AND METHODS

A. IMAGE DATA COLLECTION:

The JSB stain is a rapid staining method for the detection of malarial parasites. This stain is superior to the Field's stain as the parasites stain clearer and the morphology of the parasites is visible even in thick smear. The JSB stain constitutes of JSB solution 1(methylene blue (Medicinal) 0.5 gm, sulphuric acid (H₂SO₄) 1% 3.0 ml, potassium dichromate (K₂Cr₂O₇) 0.5 gm, disodium hydrogen phosphate dihydrate (Na₂H PO₄ 2 H₂O) 3.5 gm, distilled water 500 cc) and Solution 2 (eosine 1.0 gm, distilled water 500 cc). The preparation of stain procedure was followed as recommended National Vector Borne Disease Control Programme (NVBDCP). Chromatin (part of the parasite nucleus) is usually round in shape and stains deep red. Cytoplasm occurs in a number of forms, from a ring shape to a totally irregular shape. It always stains blue, although the shade of blue may vary between the malaria species [5].

Images were acquired using microscope system - Leica DM1000 which is interfaced to a Leica DFC 295 camera using IEEE 1394. The slides are examined under oil immersion with 1000x magnification maintaining a constant image size of 640X480 pixels.

B. IMAGE PROCESSING:

An image from JSB stained sample (thick) is prone to differ widely in the foreground / background color due to several conditions. This may be due to difference in the light source or filters, cameras, slide preparation. In order to have an analysis towards constant color characteristics, the images are normalized. In this work, the gray level normalization is incorporated through which a constant gray value of the image is maintained which does not change to different conditions. In a diagonal model, an image of unknown illumination I_u can be simply transformed to the known illuminant space I_k by multiplying pixel values with a diagonal matrix ($I_k r g b(x) = M I_u r g b(x)$). Where $\mu I r g b$ are the mean for channels r,g,b[6]. The constant grey values for each channels was assumed to be 255 (the maximum possible value) which is similar to colourless transparent pixel color. Further the normalized image is transformed to LAB color space. This color space is chosen because the L layer of the image has the image intensity as one of its component which ensures the contrast enhancement and equalization in more efficient way compared to other color spaces. The processed LAB color space image is converted back to RGB color space. The corrected RGB image is segmented using histogram based thresholding operation. This step ensures the removal of noise and artefacts to major extent without missing the infected cells. Since the protocol is dominant towards other color components, the threshold is applied on the green component of the RGB image. The architecture of the pre-processing phase is presented in Fig2. The regional maxima and minima were used as markers and thresholded images were reconstructed in order to avoid objects that are artefacts. The objects in the reconstructed image are labelled.

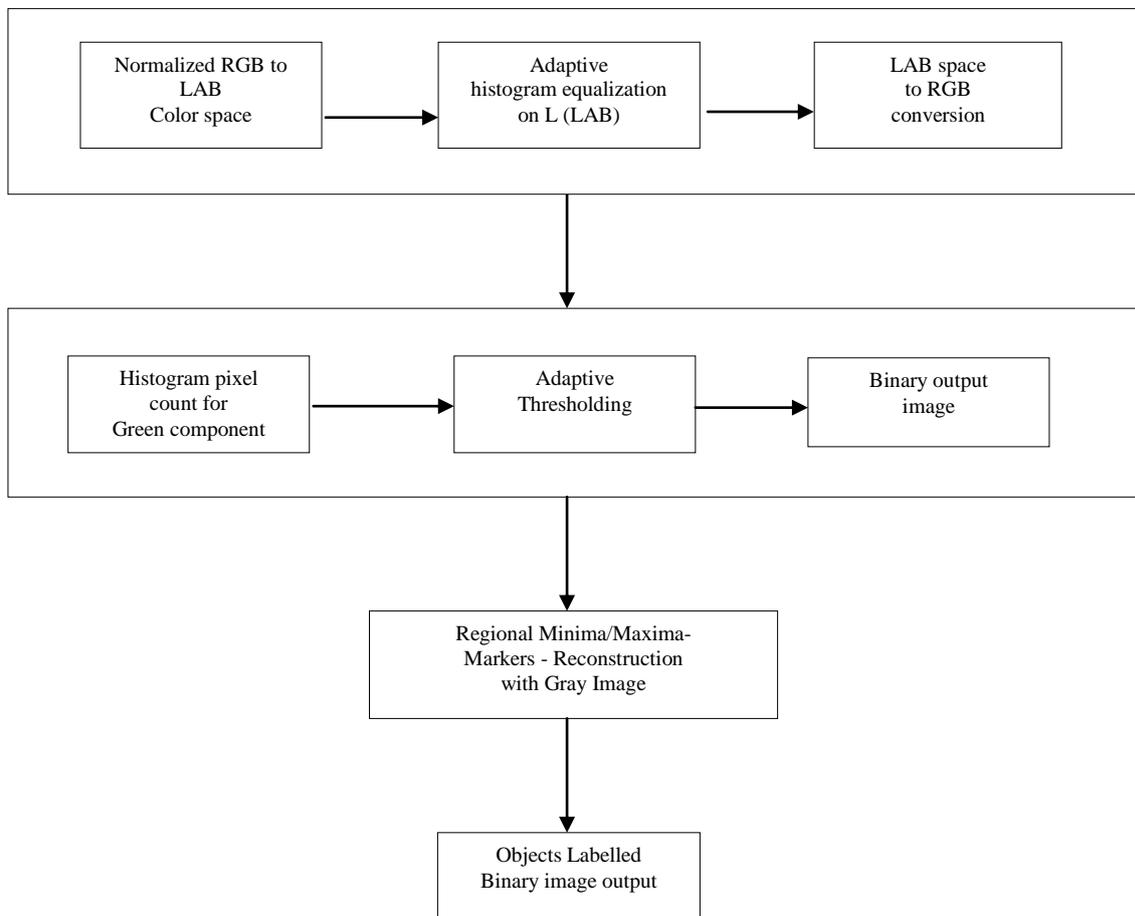


Fig2. Architecture of Image processing

The above process of object detection is called for both the normalized image and the original input RGB image. The detection using normalized image outputs a binary image which ensures the reconstruction of cells that are of interest and with very minimal artefacts which tends to be appearing as cells or due to intensity factors. This image is used as marker and the object detection using original input RGB is used as the mask image. A general reconstruction is performed between the mask and the morphologically (disk of constant radius) eroded marker image. The reconstructed image is added with the marker image in order to retain the original structure of the cells.

C. FEATURE EXTRACTION

A total of 60 samples were used for training. Each samples had number of normal and infected cells along with artefacts. The objects extracted from these samples are Parasites, WBC and artefacts. In order to classify the detected objects, twenty three image features were extracted from the detected objects for training the system. The feature includes intensity based Histogram features and shape measurement features. These features are extracted for different channel of color spaces namely gray, hue, saturation and luminosity (standard deviation).

1. First Order Statistical Features / Histogram Features:

The histogram counts and the bin locations are *pixelcounts* and *bin* (256) respectively. The first order features are defined by the following equations,

$$\text{Mean}=M=\frac{\text{sum}(\text{bin} \cdot \text{pixelcounts})}{\text{Total Number of pixels}}$$

$$\text{Variance}=V=\frac{\text{sum}((\text{bin}-M)^2 \cdot \text{pixelcounts})}{\text{Total Number of pixels}-1}$$

$$\text{Standard Deviation}=SD=\sqrt{V}$$

$$\text{Skewness}=sk=\frac{\text{sum}((\text{bin}-M)^3 \cdot \text{pixelcounts})}{\text{Total Number of pixels}-1}$$

$$\text{kurtosis}=kr=\frac{\text{sum}((\text{bin}-M)^4 \cdot \text{pixelcounts})}{\text{Total Number of pixels}-1}$$

$$\text{Fifth standard moment}=M5=\frac{\text{sum}((\text{bin}-M)^5 \cdot \text{pixelcounts})}{\text{Total Number of pixels}-1}$$

$$\text{sixth standard moment}=M6=\frac{\text{sum}((\text{bin}-M)^6 \cdot \text{pixelcounts})}{\text{Total Number of pixels}-1}$$

2. Shape Measurement Features:

Since these features are independent of color spaces, the following equations were directly applied to the binary mask image. Shape measurements can detect the changes in the size. The advantage of shape measurements is straightforward interpretation of the calculated feature values.

$$\text{Convexity}(R)=\frac{\text{Perimeter}(\text{ConvexHull}(R))}{\text{Perimeter}(R)}$$

$$\text{Compactness}(R)=\frac{2\sqrt{\pi}\text{Area}(R)}{\text{Perimeter}(R)}$$

$$\text{FormFactor}(R)=\frac{4\pi\text{Area}(R)}{\text{Perimeter}(R)^2}$$

$$\text{Area}(R)=\text{Total Number of pixels}(R)$$

III. CLASSIFICATION USING ANN

The main aim of the work is to discriminate the parasite cells from other normal cells and artefacts. A multi-layer feedforward Scaled conjugate gradient backpropagation (SCG) neural network was implemented for classifying the detected objects. The network architecture shown in Fig.3 includes 20 hidden layers and 3 output layers which are fully connected. The choice of training function was made as SCG as they are a general purpose second order technique that helps in minimizing the goal functions of several variables, while other standard backpropagation uses only the first

derivatives. SCG has been shown to be considerably faster than standard backpropagation and than other conjugate gradient methods [7].

Three target output classes namely Parasite, WBC and noise were created. Input class had 117 objects with 23 features each. The overall collected data for training consisted of $117 \times 23 \times 3 = 351 \times 23$. The network uses these data as the prior knowledge and hence the network can only be as accurate as the data that are used in to train the network. The dataset contains the input matrix (image features) and a target matrix (trueness of a class). Secondly the data has to be pre-processed and divided into subsets before the training process begins. The standard practice involves the normalization of inputs in order to avoid the slower training process and to prevent the transfer function being saturated. For this reason both the inputs and the targets are normalized. There are several optimization techniques such as mapping to [0, 1], normalizing inputs and targets to have zero mean and unity variance, etc. [7]. Network used here for the classification uses a most common pre and post processing function mapminmax which would map the minimum to maximum and normalize the data to fall between [0, 1].[7].

Before training, the input data is divided into 3 subsets.

- 1 Training set (where the gradient and weights are updated)
- 2 Validation set (the weights and biases are saved at minimum validation error)
- 3 Testing set (for comparison)

The random division of input data is done with 75% of 351 for training and 15% of 351 each for validation and testing. However this division set can be modified after the training process and the network can be reinitialized, as the network results provides the index of the data which has been used for training, validation and testing respectively.

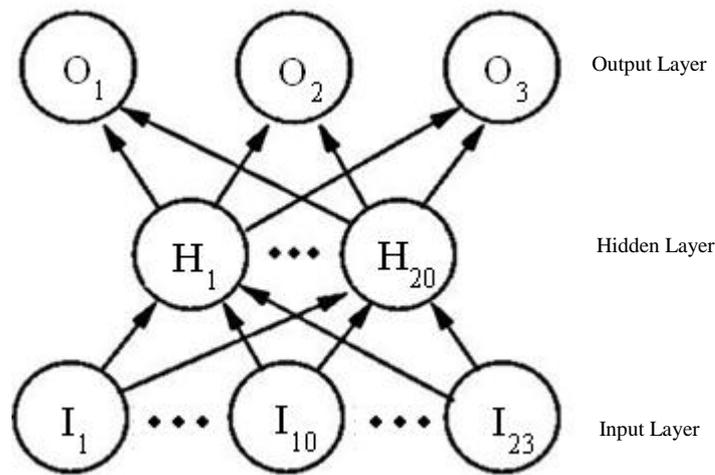


Fig.3 Neural Network Architecture

The network has been trained in batch mode with mean squared normalized error as the performance function. To terminate the training process two criteria were used, the minimum performance value and the validation check. Minimum performance value was set to 0.002 [(mean of variance of targets)/100]. The number of validation check was set to 10 which represent the number of successive iterations that the validation performance fails to decrease.

IV. RESULTS

The performance of the generated network was measured through confusion matrix which is presented in Fig.4. The diagonal cells show the number of classes that were classified correctly. And the off diagonal cells show the misclassified cases. The blue cell in the bottom right shows the total percentage of correctly classified cases (in green) and the total percentage of misclassified cases (in red). The percentage of false negatives, false positives, and true positives for the class and out of class is given in table 1.

TABLE I
NETWORK CLASSIFICATION RATES FOR THREE CLASSES

Class	False Negative	False Positive	True Positive
Parasites	8.5470	15.3846	91.4530
WBC	9.4017	5.1282	90.5983
Noise/Artefacts	13.6752	11.1111	86.3248

The proposed method has sensitivity of 91% which indicates the ability to detect the parasites correctly, with the positive prediction value to 88% indicating the success towards leaving out the non-infected cells. The specificity which relates to the identification of negative results was 85.9%. This initial study was performed with the limited number of

available positive and negative samples. Total number of samples used in this study was 110 out of which 60 were used for training the network and 50 samples were used for testing the performance of the network.

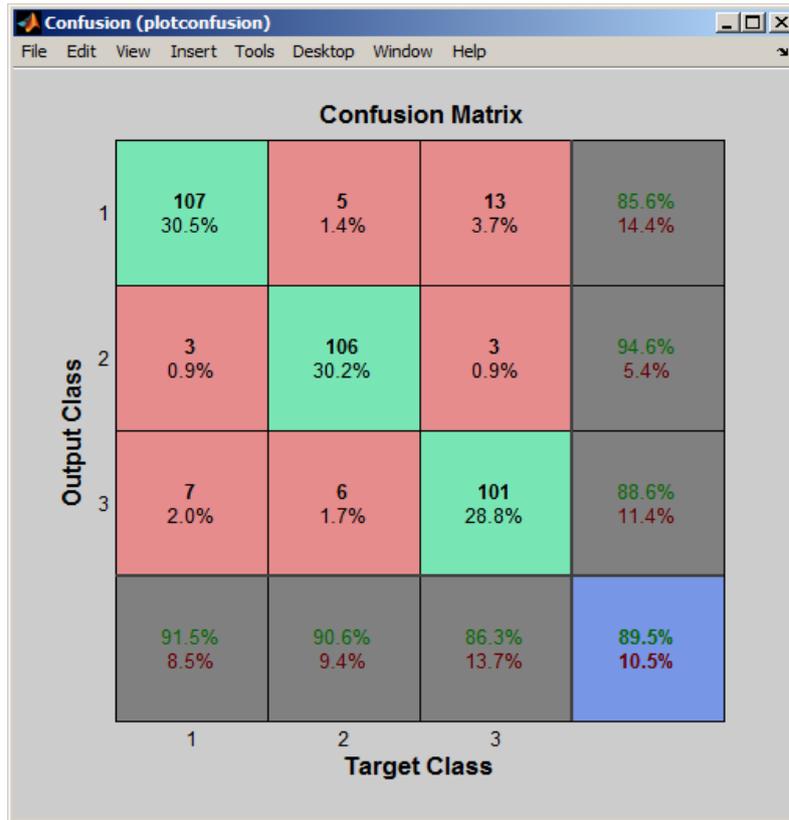


Fig.4 Confusion Matrix (Generated using Matlab)

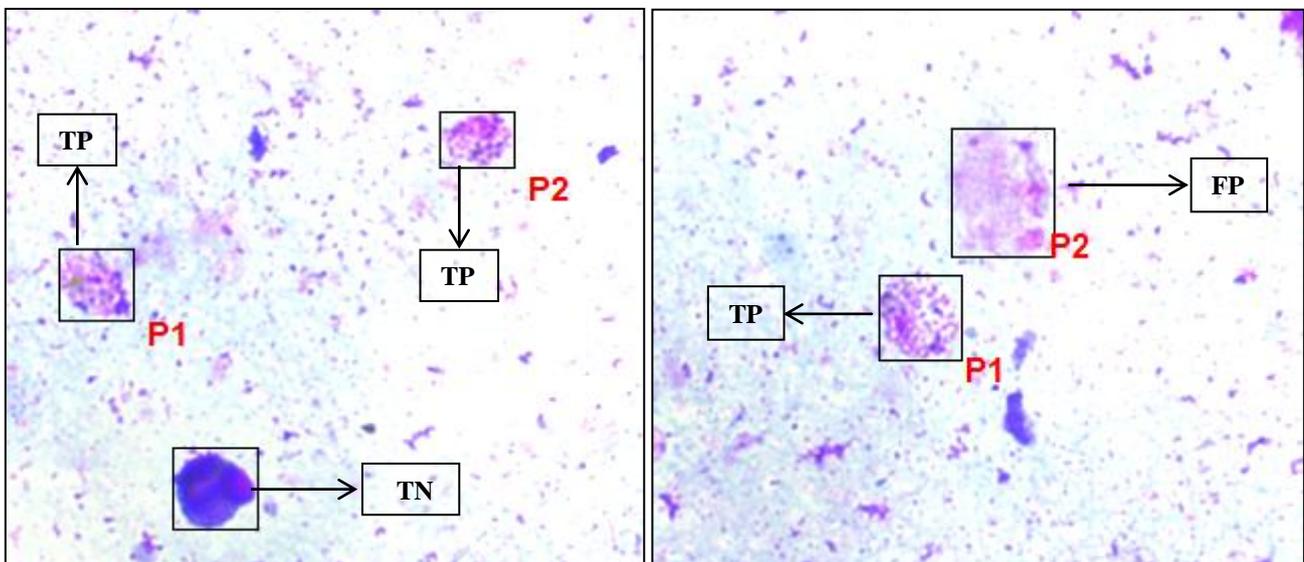


Fig.5 Two different malaria images after classification. P- P.Vivax TP – True positive TN – True negative FP – False Positive

V. CONCLUSION

This work proposes a method that detects parasites in images acquired from peripheral JSB stained (thick) blood samples using light microscope. This computerized method of detection of parasites incorporates the advantage of strengthening of malaria surveillance through which the county can evaluate the malaria control programmes and design effective health policies. Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, with minimal additional efforts, can be used to extract patterns and detect trends that are too complex. In a real diagnosis scenario a blood film from a test case could provide thousands of stained objects. Thus, the diagnostic decision can be made according to the decisions on the total number instead of a single one [8]. Moreover the current and

future study involves in the improvement of efficiency of the proposed method taking complications into consideration. A study has to be carried out in choosing the Feature that would provide more discriminant decision probability between the infected and non-infected cells. A hybrid classifier involving two or more classifiers in the same system can be considered in addressing the decision making.

ACKNOWLEDGEMENT

We thank our director for encouraging us to do this work in our organisation. This work has been carried out under the funds of DST-NHHID project and we also thank Regional Health & Family Welfare, Chennai for providing the necessary samples.

REFERENCES

- [1] World Health Organization *Fact File: 10 facts on malaria* updated on March 2013.
- [2] *World Malaria Report 2012 India* – South East Asia Region.
- [3] Kilian AH, Metzger WG, Mutschelknauss EJ, Kabagambe G. *Reliability of malaria microscopy in epidemiological studies: results of quality control*. Ethiop J Health Dev 2003; 17(3):197–204.
- [4] Samir Pachpute, Chitra Pai, A.D. Urhekar, *A comparative evaluation of conventional staining methods and immunological techniques for the diagnosis of malaria*, Journal of Evolution of Medical and Dental Sciences/ Volume 2/ Issue 7/ February 18, 2013 Page-712 – 724.
- [5] *Laboratory diagnosis of malaria- Operational guidelines*, National vector borne disease control program (NVBDCP).
- [6] Y.M Hirimutugoda, et.al. *Image Analysis system for detection of red cell disorders using Artificial neural networks*, Srilanka Journal of Biomedical Informatics 2010;1(1):35-42.
- [7] Martin Fodsllette Moller. *A scaled conjugate gradient algorithm for fast supervised learning*, Neural Networks, 6:525--533, 1993.
- [8] F.Boray Tek et.al. *Malaria Parasite detection in peripheral blood images*, Med. Imaging Understand. Anal. Conf., Manchester, UK, 2006.
- [9] Magudeeswaran Veluchamy, et al. *Feature Extraction and Classification of blood cells using Artificial Neural Network*, Americal Journal of Applied Sciences 9(5):615-619, 2012.
- [10] Herbert Ramoser, et.al. *Leukocyte segmentation and classification in blood smear images*, Proceedings of the 2005 IEEE, Engineering in Medicine and Biology 27th Annual Conference Shangai, China, September 1-4,2005.
- [11] World Health Organization - *Disease surveillance of malaria control (2012)*.
- [12] Nicholas E.Ross et.al. *Automated image processing method for the diagnosis and classification of malaria on thin blood smears*, Med. Biol. Eng. Comput. (2006) 44:427-436.
- [13] Gloria Diaz et.al. *A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images*, Journal of Biomedical Informatics 42 (2009) 296-307.
- [14] Maombi Edison et.al. *Digital analysis of changes in plasmodium vivax malaria in erythrocytes*, Indian Journal of Experimental Biology, Vol.49, January 2011, pp. 11-15.