

Comparing the Performance of Different Neural Networks for Malaria Classification

Dipti D. Patankar¹

Department of Electronics and Telecommunication
D.Y. Patil COE, Akurdi
Pune, India

Dilip G. Khairnar²

Department of Electronics and Telecommunication
D.Y. Patil COE, Akurdi
Pune, India

Abstract — Malaria is life threatening global health global health disease. Traditionally, light microscopy is used for malaria detection which is less accurate and time consuming process. To reduce false detection because of human errors, neural networks are used. For this analysis, a database contains 200 Giemsa stained malaria infected and normal RBC blood images. The Size, shape of parasite is extracted from an image and is used to identify the class and to train network. Back Propagation Neural Network (BPNN) with Bayesian Regularization as training algorithm and Naive Bayes classifier are used for malaria classification. The comparison of Bayesian Regularization (BR) and Naive Bayes Classifier technique are based on correct detection, time taken to compute the results, training time and memory utilization. Results suggest that accuracy of Naive bayes classifier is better than bayesian regularization.

Keywords— Back Propagation Neural Network, Bayesian Regularization, Giemsa stained image, Malaria Detection and Classification, Naive Bayes classifier

I. INTRODUCTION

Malaria is caused by the bites of infected anopheles mosquito. It is recorded as one of the most dangerous disease. Malaria may be asymptomatic in nature means symptoms are not seen immediately after mosquito bite, it takes 10 to 15 days. Symptoms of malaria are fever, headache, and vomiting. Human blood has three different kinds of cells RBCs, WBCs and blood platelets. Among these malaria parasites affect red corpuscles. In the human body, the parasites multiply in the liver, and then infect RBC [1].

The infection rate of malaria is around 300 to 500 people per year. Optical microscopy is used for malaria parasite detection with Giemsa stain. Staining is required to highlight the plasmodium, WBC. To make system work well without false detection, plasmodium is analyzed. There are three class of malaria. They are, P. Falciparum, P. Malariae, P. Vivax. Among these types P. Falciparum and P. Vivax are found most common. Fig. 1 shows different types of stained malaria and normal RBCs in the blood. These classes are differentiated by their shape and size [2].

This method is based on BPNN and Naive Bayes Classifier classification problems. Bayesian Regularization is based on the error optimization using Levenberg algorithm. For error approximation it reduces a combination of sum squared errors. So, BR finds the minimum weight so as to produce a network that generalizes error well. It error minimization is called as Bayesian regularization. BPNN is a

multilayer perceptron (MLP) and works well for classification problem.

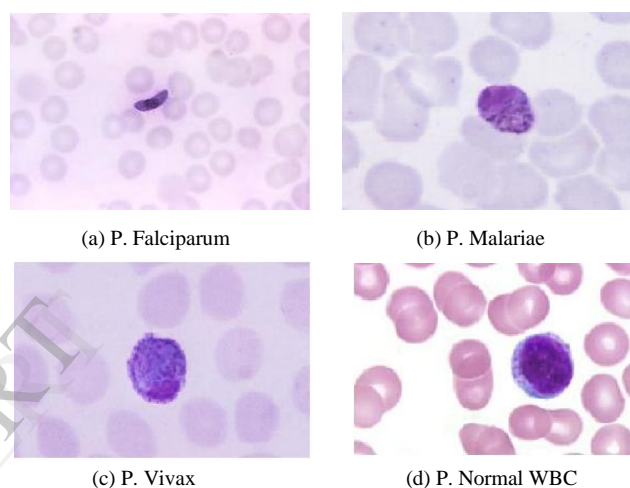


Figure 1: Different types of malaria parasites and normal WBC

Naive bayes classifier follows Bayes probabilistic model. It assumes the presence or absence of particular feature of class independent of other features. Levenberg Marquardt algorithm is used to find mean square error and weights are updated to minimize the error function. Error approximation is possible using Jacobian matrix which given by rate of change of error function with respect to weight and it is less complicated than hessian matrix.

II. RELATED WORK

There are many techniques to detect and classify malaria parasite. These techniques are described in this section. Neural networks and image processing methods help for identification. Light microscopy is time consuming and laborious process. Malaria detection in early stage is not possible with microscopy test. A trained operator is required for testing, thus it increases the cost of test. To make a system automated and improve the accuracy of system neural networks are used.

General method of identification is image acquisition, noise removing, image segmentation, feature extraction and classification. Images are acquired from public health image library with digital camera. Database contains 460 X 307 pixels 90 images. Pre-processing stage is used to remove noise from image. SUSAN edge detector is used for edge and corner detection. SUSAN edge detector gives better

performance than median filter. The next stage identifies infected RBC and parasite from background.

This is done with the help of thresholding [3]. Two threshold levels are selected one for parasite and one for infected RBC. Set of generated features is based on shape, size, color attribute and gray level texture. Final classification is possible with the BPNN. For training 77 images are used for both malaria parasite and non parasite with gradient descent algorithm. The sensitivity of algorithm is measured from PPV. This value is expressed in terms of TP, FP, FN. This process gives much better results than traditional microscopy test.

Gradient descent algorithm minimizes the error to reach at bottom of error function. This error is reduced in small steps. New weights are updated such that error is changed from previous weights. The speed of error correction is dependent on the convergence parameter. If this rate is more, then local minima may not be achieved. Convergence may take a long time if speed control parameter is slow. The error is determined as target value minus actual value.

Post treatment of malaria parasite means to re evaluate the parasite presence in blood for treatment of disease. Parasites are differentiated from each other because of their shape. *P. Falciparum* has elongated shape. This technique eliminates morphological dilation and erosion using ARR method. As the images are stained, this process locates those stained objects and helps to speed up the process. Staining process highlights parasite, WBC and artifacts in image. So, picking artifacts is found to be drawback of the process. By setting high threshold level this can be eliminated [4].

Average intensity and variance of image is computed. The region with less mean intensity and less variance is marked as a WBC and region with greater mean intensity is considered as gametocytes. The software used is MATLAB. The process gives correct result for 18 out of 20 images means 90% accuracy. It avoids the preprocessing stage and complex operations such as dilation and erosion.

Granulometry is referred to the size distribution of cell. Segmentations means to divide the image based on similar characteristics. Size of WBC and RBC is different. Dilation reduces the object boundary and erosion increases the size of highlighted object. The image is first converted to gray scale. There are two different structuring elements ring shape and disk shape. Ring shape element is used to dilate the image and disk shape is used to erode the image [5].

Ratio of average intensity of concentric ring to disk shape is computed. The ratio of ring is between 35 to 70 percent of RBC. Then peak intensities are determined to locate co-ordinates of RBC. The ratio transform gives a blob like structure is produce to locate each cell. These are TIFF format figures and used to count RBC in image. The method can directly work on gray scale so, no need to convert to binary form. The process can be applied for malaria detection.

Another technique for determination of malaria parasite is using ANN and BN. Experiment is carried using MATLAB 7 software. For this database have 580 records of infected, symptomatic and asymptomatic by malaria. These records

include data from laboratory test, previous infections by malaria. For training tansig and purelin functions are used. MATLAB has a neural network and bayesian network tool box. The training algorithm is Levenberg marquardt and junction tree for bayesian network. Network generated with 7 neurons in input layer, 10 neurons in hidden and 1 neuron in output layer [6].

Comparison with these techniques with laboratory test suggests that ANN gives better results. The results are evaluated based on sensitivity and specificity. Microscopy test gives 100 % specificity. Sensitivity of ANN and BN is 45% and 15% greater than microscopy test. Sensitivity means the ability of test to identify the infected records and specificity means to separate the records from non infected case. These values are determined from FP, FN, TP and TN.

III. METHODOLOGY

The block diagram of system is as shown in Fig. 2. The technique based on image processing method and classification using neural networks. Each block shows the image obtained at the output of that particular block.

A. Input Image

A database contains 200 images which are acquired from public health image library. There are three classes of malaria and fourth class is normal RBC. These are stained images which highlight the infected RBC, WBC, parasite and other artifacts in image. Zoom level of all images must be approximately same. Each class has 50 images thus there are 150 infected images and 50 images of normal RBC.

These are jpeg standard images. The maximum size of image is 400 X 461 pixels. Size and shape are extracted from image and a statistical record is used to train the network. Dilute Giemsa stain (1:20 vol/vol) is used to prepare slide. It takes around 20 minutes. Image is read in the form of pixels and stored in the form of array.

B. Noise Filtering

Images are acquired from electronic media; noise in image should be removed. Median filter is used for this. The size of filter is 3X3. The jpeg colored image is converted to gray scale and then median filter approach is used. Red, Green and blue colors are separately considered from image and converted to gray scale. Average filter has a drawback that it blurs the object boundaries during filtering. So, median filter is used [7].

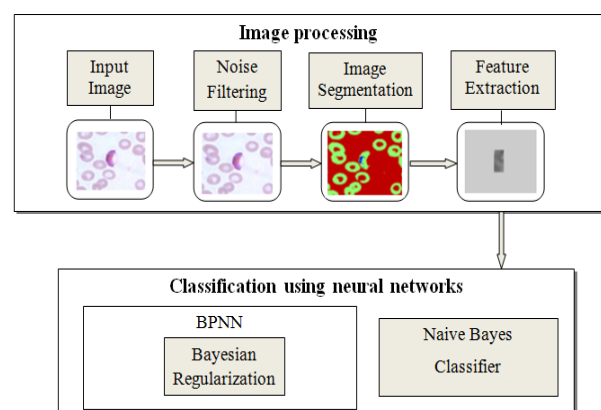


Figure 2: Block diagram of proposed system

C. Image Segmentation

Segmentation means to divide image into small segments depending on the similarity or discontinuity. Here intensity based segmentation is used. There are 4 clusters into which image is classified. A dark colored pixel means highlighted colored pixel are classified into first cluster means, parasite or WBC is classified into first two clusters. This process helps to segment the parasite from background image.

D. Feature Extraction

Features are extracted from segmented part are size and shape. Each parasite is differentiated by its shape and size. Such as P. Falciparum is elongated in shape, P. Malariae is round to oval and P. vivax is slightly bigger than P. Malariae round in shape. To determine the size, Number of dark pixels in the region is calculated and for shape canny edge detector algorithm is used. In this way entire data is collected for testing. Out of 200 images in database 40 images are used for training. Two features of each image means there are 80 records in matrix for training.

E. Classification

Classification is used using BPNN and naïve bayes classifier. To achieve local minima of error function when new data enters error may increase rapidly. For this network has to be generalized again. This process is called as a Bayesian Regularization. It calculates mean square error. Naïve bayes classifier is a probabilistic model which considers each event independently.

For this it generates a model for classifier. For this set of input matrix has size equal to 2×80 . Target output matrix has a size of 1×80 . For first class of malaria i.e. P. Falciparum, first 20 positions of target matrix i.e. from 1 to 20 columns is set to one. For next twenty images belong to second class i.e. P. Malariae, from 21 to 40 columns, first row of target matrix is set to two. Similarly, for third P. Vivax class 41 to 60 columns are set to three.

Similarly last 61 to 80 columns belong to normal RBC is set to four. Thus, input and target output matrix is created for training. A network with bayesian regularization as training algorithm and 0 neurons in hidden layer with 500 iterations are used. Tansig and purelin are activation functions. A feed forward neural network has one input layer, one output layer. The process executes separately for shape and size [8].

Network is trained with different training samples such as for 10, 20, 30, 35 samples. The performance is plotted with this is as shown if Fig. 3. It is a plot of percentage accuracy VS training samples. It shows that optimum performance is obtained at 20 samples. Thus network is trained with 20 Samples. Above 20 samples network becomes saturated and cannot give much more improvement in the performance. For 20 samples 92.50% accuracy is achieved for Bayesian Regularization and 97.50% for naïve bayes classifier. Similarly, to select number of neurons in hidden layer a process is carried for different neurons but minimum error is obtained for 50 neurons as shown in Fig.4. As training epochs increased the network is genetalized at 100 epochs for less number of neurons network requires more iterations to

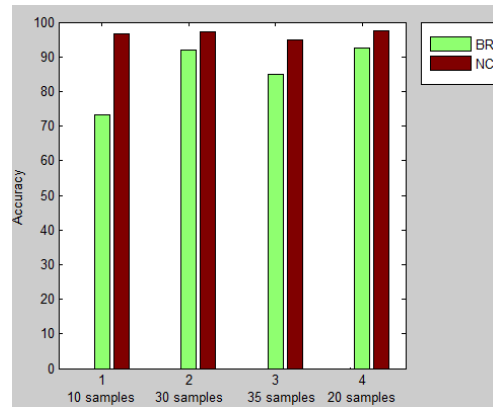
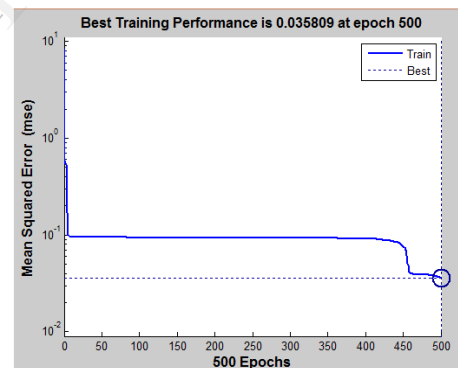


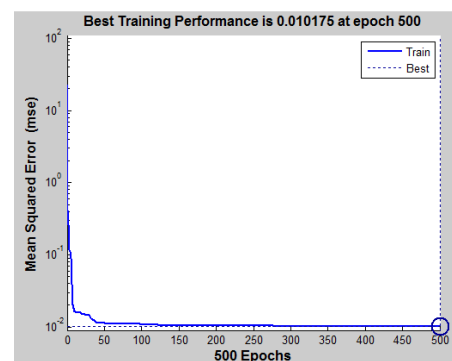
Figure 3: Accuracy plot for different training samples

generalized the network. After 50 neurons network gets saturated and does not shows any further improvement in network [9].

For naïve bayes classifier a mean of each class is calculated. There are 4 classes with shape and size of each class. Means there are 8 records. Each feature will have separate mean say from m1 to m8. First class say P. Falciparum have shape with mean m1 and size with mean m2. Similarly second class P. Malariae will have means m3 and m4. When image is to be tested its size and shape is subtracted from mean of a particular class.



(a) Plot for 10 neurons in hidden layer



(b) Plot for 20 neurons in hidden layer

Figure 4: Error plot with different number of neurons

Image is of shape 'x' and size 'y'. Now 'x' is subtracted from m1, m3, m5 and m7 and 'y' is subtracted from m2, m4,

m6, m8. Then minimum absolute value obtained from shape is stored in variable 's1' and from size is stored in variable 's2'. If 's1' is equal to 's2' then correct class is determined otherwise class is not determined. It uses normal means Gaussian distribution. Prior probability is calculated from the relative frequency of input class data. For testing there is no need to calculate shape and size separately. Input and output matrix is entered to the network and image to be tested is selected. From the statistical data, probability of each class of parasite is calculated and correct classification is possible. As there are 200 images in a database and 40 of which are of each class.

The prior probability of any class is e.g. P. Falciparum is $\frac{40}{200}$. Let image 'x' to be tested have a values of shape and size in between class P. Malariae and P. Vivax. These values of 'x' are near to the 'y' values of a class P. Malariae and 'y₁' values of a class P. Vivax. Similarly, posterior probability of x being P. Vivax is calculated. 'x' achieves largest probability between the two classes. The Naive bayes classifier network 'nb' is trained and it generates model as follows.

nb = Naive Bayes classifier with 4 classes for 2 dimensions.

Feature Distribution(s): normal

Classes: Falciparum, Malariae, Non infected, Vivax

IV. RESULT

Image is tested on different techniques. For naïve bayes classifier maximum probability is given to final class. Let the image to be tested is of class vivax means third class out of four classes.

Class = 'Vivax'

Post = 0.0000 0.0001 0.9998 0.0002

Images are tested on bayesian regularization and naïve bayes classifier separately. To plot the accuracy all the images are identified, memory required for bayesian regularization is more than naïve classifier. The class is identified as P. Vivax because it is with maximum probability. The result is evaluated as shown in Table 1.

Table 1: Comparison of neural network techniques

Feature	Bayesian regularization	Naive bayes classifier
Correct identification	92%	97.5%
Class detection time	Around 10% more than Naive bayes	Less than BR
Memory requirement	787 bytes	313 bytes
Training time	Around 58.17 seconds (for 500 epochs)	Around 0.77915 seconds

V. CONCLUSION

Malaria is harmful disease and hence fast, reliable and accurate detection techniques are required. Malaria Images are classified from their morphological characteristics. All images are of same zoom level so that parasites can be highlighted and features extracted easily. Noise is removed using median filter and shape is determines from canny edge detector.

Naive Bayes classifier is a probabilistic model and it is more accurate than bayesian regularization. Hidden neurons, Number of epochs can be selected in BPNN. Bayesian Regularization requires more memory to build the network than naive bayes classifier also it requires 10% more time for its execution. So, Naive Bayes classifier is proved to be better technique.

VI. ACKNOWLEDGMENT

I express my sincere gratitude towards the faculty members who make this project. I would like to express my thanks to my guide for his whole hearted co-operation and valuable suggestions, technical guidance throughout the Project. Finally, I would like to thank to all our staff members of Electronics & Telecom-communication Department who helped me directly or indirectly to complete this work successfully.

VII. REFERENCES

1. World health organization. What is malaria? <http://www.who.int/topics/malaria/en/>
2. Public Health Image Library, Centers of Disease control and Prevention, <http://phil.cdc.gov/phil/sessionexpired.asp>
3. Neetu Ahirwar, Sapnojit. Pattnaik, and Bibhudendra Acharya" Advanced image analysis based system for automatic detection and classification of malarial parasite in blood images", International Journal of Information Technology and Knowledge Management January-June 2012, Volume 5, No. 1 pp. 59-64
4. S. Kareem, I. Kale, R.C.S Morling,' Automated P. falciparum Detection System for Post-treatment Malaria Diagnosis Using Modified Annular Ring Ratio Method', 2012 IEEE, pp.432-436.
5. S. Kareem, I. Kale, R.C.S Morling,' A Novel Method to Count the Red Blood Cells in Thin Blood Films', 2011 IEEE, pp. 1021-1024
6. Austeclino Magalhaes Barros Junior, Angelo Amancio Duarte, Angelo Amancio Duarte" Artificial neural networks and bayesian networks as supporting tools for diagnosis of asymptomatic malaria", 2010 IEEE.
7. Rafael C. Gonzalez, Richard E. Woods, Steven L. Eddins 'Digital Image Processing Using MATLAB', Second edition, chapter 4, pp 163-231.
8. Support and compatible compilers for R2011a,<http://www.mathworks.in/support/compilers/R2011a/win64.html>
9. Simon Haykin 'Neural Networks and Learning Machines', Third Edition, Section 4.16 pp-197.