AUTOMATED IDENTIFICATION & CLASSIFICATION OF MALARIAL PARASITES IN THIN BLOOD SMEAR IMAGES

Soumya Das[1], Sony P[2], Jyothi R L[3]

[1]M-Tech student, Dept. of Computer Science & Engineering, College of Engineering, Chengannur, Kerala, India
[2,3] Asst. Professor, Dept. of Computer Science & Engineering, College of Engineering, Chengannur, Kerala, India

---***---

Abstract - Malaria is a dreadful disease in the hematological field causing millions of death. Fast diagnosis and acute treatment of malaria is important to reduce the death rate. Hence the rapid diagnosing and proper medication is the intense need of era. Malaria analysis is based on microscopic examination of blood films. In this work, image processing techniques are used for detection of malaria from microscopic images of Giemsa stained thin blood smear. Here, blood smear images are analyzed based on two methods. First, by extracting SIFT features from preprocessed images and lead to the classifier for recognition. A comparative analysis of SVM (Support Vector Machine) and ANN (Artificial Neural Network) is carried out for recognition of extracted SIFT features. Second method comprises of leading the image directly without any preprocessing to CNN based network. SIFT based feature extraction method produces recognition efficiency of 94% with SVM and 96% efficiency with ANN.CNN based feature extraction method produces recognition efficiency of 98%.

***Key Words*: RBC, Malaria, Thresholding, SIFT, ANN, CNN**

1. INTRODUCTION

Malaria is a dreadful infectious disease which affects humans and other animals. It is a mosquito borne and tropical Vector borne disease that cause sufferings and millions of death in different parts of the world. According to the World Health Organization (WHO) estimates, there were 212 million new cases of malaria worldwide in 2015. Malaria is the most dangerous one among the entire mosquito borne diseases.

Automated identification and classification of malarial parasites in thin blood smear images is a research project undertaken in the field of medical diagnosis and pathological analysis, to detect the presence of malarial parasite, classification into malaria infected or not and analyzing the efficient method of malarial diagnosis by implementing the work using ANN and CNN other than SVM. The aim of the project is to develop the algorithms to provide an efficient, diagnostic platform for a quantitative analysis of malaria infection.

The main objectives of this method are developing automated and accurate method to find presence of malarial parasite in blood and identifying the efficient method for malaria diagnosis. The existing methods for malaria diagnosis are inaccurate. This is the motivation of the project. Conventional microscopy is the existing method for malaria diagnosis. The disadvantages of this method are that method is time consuming, labour intensive and results depends on the skills of the microscopist. Here, lies the scope of this work.

This paper will progress through the following steps of image processing: image acquisition, pre-processing, segmentation, feature extraction and finally classification. Pre-processing includes removing the noise in the image, segmentation to separate foreground and background pixels, feature extraction to extract the features for classifying into malaria infected cells and non-infected cells and further classification into whether infected with malaria or normal sample.

2. LITERATURE SURVEY

A number of studies on the possibility of automating conventional microscopy have been done in the past. In this section a number of these studies are reviewed.

Kshipra C. Charpe et. al. [1], proposed automated malarial parasite and their stage detection, which proceeds in steps like image acquisition, segmentation, feature extraction and then classification. The features: color, shape, size, intensity, texture will be extracted and classification is done using SVM. In this work, the parasite infected RBC and their count is also found.

S.S. Savkare et. al. [2] developed a fully automatic system for counting and classification of Malaria parasite infected erythrocytes and detection of life stage of parasites. Here, Otsu threshold erythrocytes are segmented form pre-processed images, watershed algorithm is used to separate overlapped cells, statistical and colour features are extracted and given to the SVM binary classifier which classifies malaria infected erythrocytes and SVM multi classifier is used for detection of parasite life stages.

Hanung Adi Nugroho et. al. [3], developed based on the image processing technique to detect three stages of Plasmodium parasites while in human host, i.e. trophozoites, schizonts, and gametocyte plasmodium falciparum. Feature extraction based on histogram-based texture is used to extract feature parasite cell. Multilayer perceptron back propagation algorithm is used to classify all features. The results show that the proposed method achieves accuracy of 87.8%, sensitivity of 81.7%, and specificity of 90.8% for detecting infected red blood cells.

Aeggarut Pinkaew et. al. [4] developed an automated classification system operating on digitized images of thick blood film to classify between Plasmodium falciparum and Plasmodium vivax malaria parasite species. Here, five statistical features such as mean, standard deviation, kurtosis, skewness and entropy from four color channels (green, intensity, saturation, and value) are calculated. The features are then projected onto a subspace representing image characteristics from both species. The projected features are used by the support vector machine for classification.

Wongsakorn Preedanan et al., [5] investigate automated detection of malaria parasites in images of Giemsa-stained thin blood films. Here, parasitemia is based on automatic segmentation, feature extraction and classification methods. Segmentation relies on adaptive thresholding and watershed methods. Statistical features are then computed for each cell and classified using SVM binary classifier. Accuracy of classification is validated based on the leave-one-out cross-validation technique. This method yields 92.71% sensitivity, 97.35% specificity and 97.17% accuracy.

Ahmedelmubarak Bashir et. al., [6] proposed an accurate, rapid and affordable model of malaria diagnosis using stained thin blood smear images. The method made use of the intensity features of Plasmodium parasites and erythrocytes. Images of infected and non-infected erythrocytes were acquired, pre-processed, relevant features extracted from them and eventually diagnosis was made based on the features extracted from the images. A set of features based on intensity have been proposed, and the performance of these features on the red blood cell samples from the created database have been evaluated using an artificial neural network (ANN) classifier.

Hassan Abdelrhman Mohammed et al., [7] proposed an image processing system to identify malaria parasites in thin blood smears and to classify them into one of the four different species of malaria. Many techniques were implemented in the preprocessing stage to enhance the images. In the first part of the system morphological processing is applied to extract the Red Blood Cells (RBC) from blood images. The developed algorithm picks the suspicious regions and detects the parasites in the images including the overlapped cells. Accordingly, the RBCs are classified into infected and non-infected cells and the number of RBCs in each image is calculated. The second part of the system uses the Normalized Cross-Correlation function to classify the parasite into one of the four species namely, Plasmodium falciparum, Plasmodium vivax, Plasmodium ovale. Compared to manual results, the system achieved 95 % accuracy for detection and counting of RBCs and 100% for detection and classifying the parasite into one of its four types.

3. PROPOSED METHOD

In the proposed methodology, image processing techniques are used for detection of malaria from microscopic images of Giemsa stained thin blood smear. Here, blood smear images are analyzed based on two methods. First, by extracting SIFT features from preprocessed images and lead to the classifier for recognition. A comparative analysis of SVM (Support Vector Machine) and ANN (Artificial Neural Network) is carried out for recognition of extracted SIFT features. Second method comprises of leading the image directly without any preprocessing to CNN based network.

3.1 SIFT BASED RECOGNITION SYSTEM

In this work, the image of blood smear goes through different stages. First, the stained image is acquired using microscope attached to camera then preprocessing and segmentation is done. Then the features are extracted and classified by a classifier. Finally the features are extracted using SIFT and then classified by a classifier.

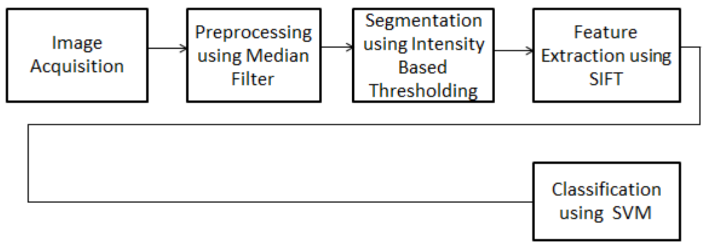

Fig -1: Block Diagram

3.1.1 IMAGE ACQUISITION

A database of stained blood smears images has to be taken from a laboratory by a microscope attached with camera. The images obtained are in jpeg format and the resolution will be according to the camera attached to the microscope. Then images can be captured and it can be viewed if we connect a laptop or mobile to it. Image acquired contain noise which is produced from instruments during image acquisition and it has to be removed which is done in image preprocessing.

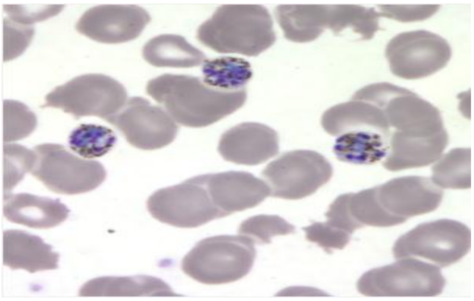

Fig -2: Image Acquisition

3.1.2 PREPROCESSING

Image pre-processing is useful technique for removing noise from images. The images acquired contain poisson noise. Poisson noise is produced by electronic equipment. Median filter, Weiner filter and Gaussian filter are used to remove poisson noise. Among this, median filter is found to be the best. So, median filter is used to remove noise.

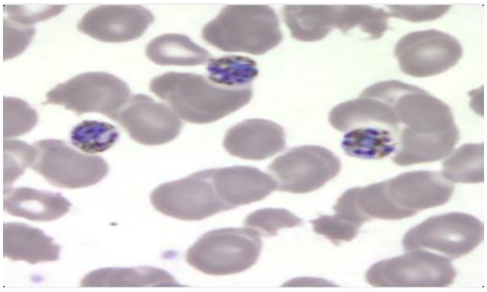

Fig -3: Preprocessed image

3.1.3 SEGMENTATION

Segmentation is the process of partitioning the image into different segments. Here, the image is partitioned into foreground and background. This is to isolate the RBCs. Simplest approach for segmentation is used here which is the selection of a suitable intensity threshold. All pixels with a value higher than a particular threshold value is classified as the region of interest and all pixels with a lower value are classified as background pixels. Such a distribution is called bimodal [12] because there are two mode values: one for the background and one for the feature.

Fig -4: Segmented image

3.1.4 FEATURE EXTRACTION

Feature extraction is applied to extract relevant features from images. In this work, SIFT is used for extracting features. The scale-invariant feature transform (SIFT) is an algorithm in computer vision to detect and describe local features in images [13]. SIFT key points of objects are first extracted from a set of reference images and stored in a database. An object is recognized in a new image by individually comparing each feature from the new image to this database and finding candidate matching features based on Euclidean distance of their feature vectors. From the full set of matches, subsets of key points that agree on the object and its location, scale, and orientation [14] in the new image are identified to filter out good matches. The determination of consistent clusters is performed rapidly by using an efficient hash table implementation of the generalized Hough transform. Each cluster of three or more features that agree on an object and its pose is then subject to further detailed model verification and subsequently outliers are discarded. Finally the probability that a particular set of features indicates the presence of an object is computed, given the accuracy of fit and number of probable false matches. Object matches that pass all these tests can be identified as correct with high confidence.

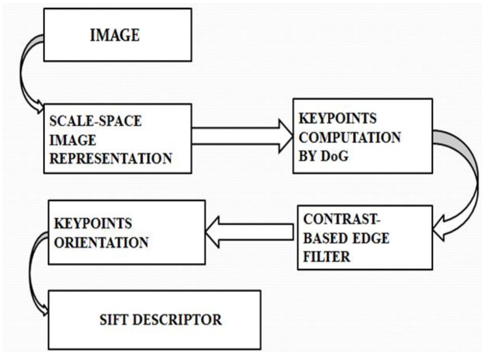

Fig -5: SIFT Algorithm Steps

3.1.5 CLASSIFICATION

This is the final stage where the features of input image is compared with that of the infected RBC image features and then classified. For this the classifier is trained by providing the features extracted in the previous stage. In the proposed methodology Support Vector machine (SVM) classifier is used first. Then, the performance of SVM is analyzed by implementing classification using ANN.

(A)Support Vector Machine (SVM)

Support vector machine constructs a hyper plane or set of hyper planes [13] in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks like outliers detection. Here, the classifier is trained to enable it to classify the input images into infected and normal RBC image. This is how the presence of the malaria parasite in the blood will be detected. The performance of

SVM classifier is analyzed by implementing the classification using Artificial Neural Network (ANN). Various performance measures such as specificity, sensitivity, precision and so on are analyzed.

(B) Artificial Neural Network

An Artificial Neural Network (ANN) is a computational model based on the structure and functions of biological neural networks. Here, two layer feed forward network is used. A feed forward neural network [13] is an artificial neural network wherein connections between the units do not form a cycle. It consists of a large number of simple neuron-like processing units, organized in layers. Every unit in a layer is connected with all the units in the previous layer. The weights on these connections encode the knowledge of a network. Often the units in a neural network are also called nodes. Data enters at the inputs and passes through the network, layer by layer, until it arrives at the outputs. During normal operation, that is when it acts as a classifier, there is no feedback between layers. SIFT based feature extraction method produces recognition efficiency of 96% efficiency with ANN.

3.2 CNN BASED RECOGNITION SYSTEM

Convolutional Neural Network (CNN) is the second method in the proposed work. The performance of the proposed work is compared using Convolutional Neural Network (CNN) to analyze which of these will give best results.

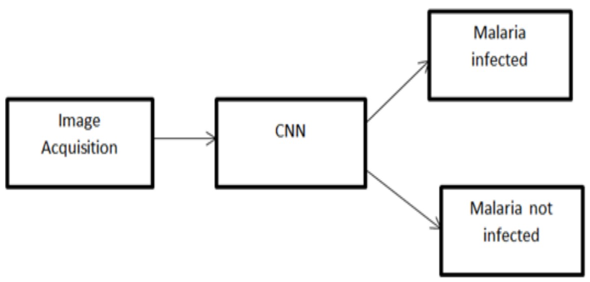

Fig -6: CNN Based Recognition System Diagram

3.2.1 CONVOLUTIONAL NEURAL NETWORK

A convolutional neural network (CNN) is a class of deep, feed-forward artificial neural networks [13] that use a variation of multilayer perceptron designed to require minimal preprocessing. They are also known as shift invariant or space invariant artificial neural networks (SIANN), based on their shared-weights architecture and translation invariance characteristics. CNNs use relatively little pre-processing compared to other image classification algorithms. This means that the network learns the filters that in traditional algorithms were hand-engineered. CNN based feature extraction method produces recognition efficiency of 98%.

Table -1: Comparison between SVM, ANN and CNN

Comparison between SVM, ANN and CNN			
PARAMETERS	SVM	ANN	CNN
Accuracy	0.9434	0.9620	0.9820
Sensitivity	09574	0.9790	0.9811
Specificity	0.8333	0.8531	0.8730
Precision	0.9783	0.9811	0.9941
Prevalence	0.8868	0.8570	0.8531

4. CONCLUSIONS

In this work, a system for diagnosing malarial parasites using microscopic images of stained blood samples was developed. A novel technique of segmenting Plasmodium parasites from infected RBC is developed. Here, malaria detection is carried out based on two methods. First, based on SIFT features and second based on deep CNN network. In SIFT based recognition method, SIFT features are extracted from preprocessed images and lead to the classifier for recognition but in CNN, the image is directly led without any preprocessing to CNN based network.

The detection of the parasites in blood samples requires more accuracy and for that the number of features considered plays a vital role in it. Features extracted using SIFT gives more accuracy in the proposed method. Thus, the proposed method is able to identify and classify the blood smear images and produced a recognition efficiency of 94% with SVM and 96% with ANN as explained before. CNN based feature method produced recognition efficiency of 98%. Therefore, CNN is found to produce best results for malaria detection.

This procedure will be a stepping stone in malaria diagnosis, as it avoids the usual ways of converting the image to different regions and thereby segmenting it. The method has been implemented in the Mat lab environment, for the flexibility and speed of prototyping the image processing options. The proposed method not only classifies malaria infected and non- infected, but also studies the classification using SVM and ANN and finds that ANN is found to be better classifier than SVM for the proposed work. It also analyzes the implementation of work using CNN and concludes that CNN produces the best results for malaria detection.

ACKNOWLEDGEMENT

I would like to thank my Principal of the institute Dr. Jacob Thomas V., HOD Dr. Smitha Dharan, CSE staff, parents and well-wishers for their immense technical and nontechnical support.

REFERENCES

[1] Miss. Kshipra C. Charpe, Dr. V. K. Bairagi "Automated Malaria Parasite and there Stage Detection in Microscopic Blood Images," 2015.

[2] S.S. Savkare, S.P. Narote, "Automatic System for Classification of Erythrocytes Infected with Malaria and Identification of Parasite's Life stage," 2nd International Conference on Communication, Computing & Security (ICCCS), 2012.

[3] Hanung Adi Nugroho, Son Ali Akbar, E. Elsa Herdiana Murhandarwati "Feature Extraction and Classification for Detection Malaria Parasites in Thin Blood Smear," 2nd Int. Conference on Information Technology, Computer and Electrical Engineering (ICITACEE), Indonesia, 2015.

[4] Aeggarut Pinkaew et. al., "Automated classification of malaria parasite species on thick blood film using support vector machine," Biomedical Engineering International Conference (BMEiCON), 2015.

[5] Wongsakorn Preedanan et. al., "Automated Detection of Plasmodium Falciparum from Giemsa-Stained Thin Blood Films," IEEE, 2016.

[6] Ahmedelmubarak Bashir et. al., "Detection of Malaria Parasites Using Digital Image Processing," International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE), 2017.

[7] Hassan Abdelrhman Mohammed et. al, "Detection and Classification of Malaria in Thin Blood Slide Images," International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE),2017.

[8] World Health Organization, Basic malaria microscopy, 2nd edition, WHO press, 2010.

[9] Rafael C Gonzalez and Richard E. Woods, "Digital Image Processing", 3rd edition.

[10] Parker, Jim R, Algorithms for image processing and computer vision, John Wiley & Sons, 2010

[11] Nixon, Mark, Feature extraction & Image processing, Academic Press, 2008

[12] http://what-when-how.com/biomedical-image-analysis/intensity-based-segmentation-thresholding-biomedical-image-analysis/

[13] www.wikipedia.org

[14] Jian Wu, Zhiming Cui et al., (2013) A Comparative Study of SIFT and its Variants, Measurement Science Review, Volume 13, No. 3