IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

DETECTION OF MALARIAL PARASITES IN BLOOD SAMPLES USING IMAGE PROCESSING, MACHINE LEARNING AND DEEP LEARNING

[1]Mrs. A.Usha, [2]N.M.Vishal Kanna, [3]M.Swarna, [4]M.Shyam Ganesh
[1]Associate professor, [2]UG Student, [3]UG Student, [4]UG Student
Department of Electronics and Communication Engineering
Easwari Engineering College, Chennai, India

***Abstract*:** Malaria is a dreadful disease caused by single-celled microorganism instigating millions of death worldwide. Hence a rapid with accurate diagnosis and proper effectual medication is the need of the hour. The diagnosis of the infection starts with the microscopic examination of blood films. Diagnostic process automation with an intellectual framework which would diagnose malaria parasites may be successful in solving the above problem. Hence, several image processing techniques are used in this work to expose malaria from the microscopic images of Giemsa stained thin blood smear. For the identification of the derived SIFT (Scale Invariant Feature Transform) functions, a comparative study of SVM (Support Vector Machine), CNN (Convolutional Neural Network) is pursued. MATLAB (Matrix Laboratory) software is used for this image processing and classification processes. A recognition efficiency of 94% is obtained from SVM and 98% from CNN. Hence, CNN provides better efficiency when compared to SVM.

Keywords-Malarial parasites, Image processing, Feature extraction, SVM classifier, CNN classifier, Deep learning

I. INTRODUCTION:

Malaria is a mosquito-borne disease that affects humans and animals in different parts of the world causing misery and millions of deaths. A research undertaken in the field of medical diagnosis and pathological analysis is the automatic detection and categorization of malaria parasites in thin blood smear images. The current malaria diagnosis approach is traditional microscopy which is time consuming, labor intensive and results depend on the researcher's skills. Normal healthy blood cell (Fig-1) and malaria affected blood cell (Fig-2) can be differentiated very well. The research is aimed at developing algorithms to provide an efficient diagnostic tool for quantitative malaria infection analysis.

Malarial Parasites: Plasmodium vivax, Plasmodium ovale, Plasmodium falciparum, Plasmodium malariae are parasites instigating malaria. Among the four malarial parasites, Plasmodium falciparum is the deadliest. Plasmodium vivax is widely distributed and the infections caused by this parasite leads to severe disease due to splenomegaly. Plasmodium malariae causes "benign malaria", which is less dangerous when compared to other infections. Plasmodium ovale causes tertian malaria. All these parasites are transmitted through the female anepholes mosquitoes.

Different stages in the growth and development of these parasites (Fig-3) are Ring stage, Trophozoite stage, schizont stage, and Gametocyte stage. Trophozoites mature in Ring stage into schizonts that rupture releasing merozoites. Trophozoite stage is the activated, feeding stage of the protozoa's life cycle. The parasite replicates its DNA several times at the schizont level, and several mitotic divisions occur asynchronously. Gametocytes are the precursors of male and female gametes where, via the developmental transition from asexual replication in erythrocytes, malaria parasites emerge in the human host.

The rest of this paper is organized as follows. We provide review of literature, methodology followed by results and discussion, comparative analysis and performance analysis. We close with a conclusion.

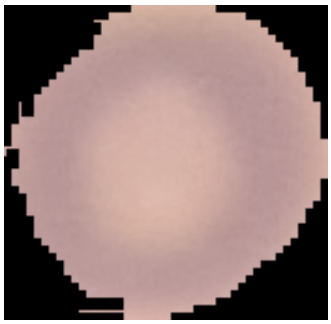

Fig-1: Healthy blood cell

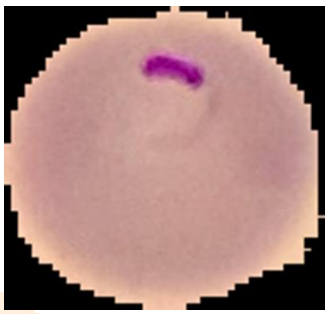

Fig-2: Malaria infected blood cell

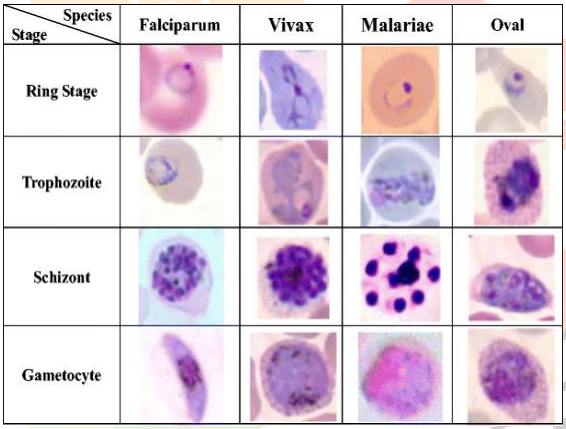

Fig-3: Stages in the growth and development of a malarial parasite in a blood cell [5]

I. REVIEW OF LITERATURE:

1. Min Liu et al., (2017) '**ABO Blood Group Detection Based on Image Processing Technology'** proposed a quick, accurate and reliable blood group judgment system based on the ABO blood group image features. To obtain the best approximation of the original image, the median filter is used to remove the noise. The characteristic parameters of the blood group ABO are determined according to the image's gray level distribution.
2. Abubakar Yamin et al., (2017) '**Image Processing Based Detection & Classification of Blood Group Using Color Images'** proposed a program that would use image recognition to identify blood groups. Preprocessing methods, HSV Luminance and morphological operations are measures used to identify the type of blood group using image processing techniques.
3. Alireza Karimian et al., (2016) '**Design a new algorithm to count white blood cells for classification Leukemic Blood Image using machine vision system'** proposed that color space transfer models can be used to classify the white blood cells. The community of leukocytes is divided by division of watershed conversion and cleanup of the picture is completed.
4. Corentin Dallet et al., (2014) '**Real Time Blood Image Processing Application for Malaria Diagnosis Using Mobile Phones'** proposed a simple and stable Android cell phone application platform to analyze blood sample images. The application is based on MATLAB 's novel Annular Ring Ratio process, checked and validated. The system detects blood components including Red Blood Cells (RBCs), White Blood Cells (WBCs), parasitemia of the infected RBC parasites.
5. I.Kale et al., (2011) '**A Novel Method to Count the Red Blood Cells in Thin Blood Films',** hereby**,** a novel concept to classify the total number of red blood cells (RBCs) and their position in a Giemsa stained thin blood film picture was proposed. The method uses basic knowledge on component cell structure and brightness to detect the RBCs in the picture. It removes the segmentation procedures used to segment the cells into the microscopic image and prevents pre-processing of images to deal with non-uniform illumination before cell detection.

I. METHODOLOGY:

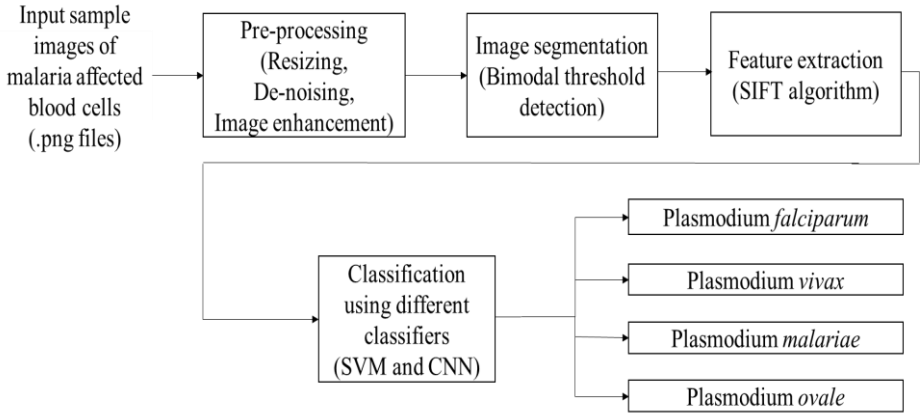

Fig-4: Block diagram of malarial parasite detection

System architecture (Fig-4) used for malaria parasite detection involves the following main steps: image acquisition (input image), preprocessing, segmentation, feature extraction, classification and final result (identification of the malaria causing parasite). Pre-processing involves eliminating image noise, segmenting to distinguish foreground and background pixels, extracting features such as contrast, similarity, homogeneity, capacity, entropy for the classification of malaria-infected cells and non-infected cells and further classification of the malarial parasite using SVM and CNN. From the outcome of the classification system using SVM and CNN, the final result i.e., the species of the malarial parasite in the given input image (thin blood smear) is identified and the results produced by both the classifiers SVM and CNN are compared for accuracy. MATLAB is implemented in all stages of the method.

- ***IMAGE ACQUISITION*** *(input image)*: A total of 400 photographs of thin blood smears stained by giesma was collected as a dataset [4]. Such photos have various magnifications and scale characteristics.

- ***PREPROCESSING***: The aim of pre-processing is to remove unwanted objects and noise from the image to facilitate image segmentation into meaningful regions. Preprocessing steps include: transforming colored images into grayscale presentation, background estimation using morphological opening technique, removing the background image from the original image, improving image contrast and transformation of grayscale images into binary images with thresholding techniques, turning images into negative presentation.

- ***SEGMENTATION:*** Segmentation is the mechanism whereby the image is partitioned into different parts. The image is divided into foreground and background here. It is for isolating the RBCs. Simplest segmentation algorithm used here is the selection of an acceptable threshold of intensity. All pixels with a value greater than a given threshold value are marked as the interest zone and the lower-value pixels are labeled as background pixels. Such a distribution is considered bimodal since there are two mode values: one for background and another one for feature. In certain cases, after the foreground has been stripped, there are still clumped RBCs and the iterative separation of clumps is done as such. The individual cells are retained while the clumps extracted are separated. The boundary is extracted for each clump and their curvature is determined. For each object, the boundary curvature vector is analyzed for regional maxima, indicating concavity points from which potential split lines will be drawn. An iterative process is used for the splitting of RBC clumps. The original individual RBCs and resulting individual RBCs from the process of clump splitting are merged into a binary RBC image. Unnecessarily separated cells are morphologically rebuilt and inserted back into the RBC binary image.

- ***FEATURE EXTRACTION:*** Extraction of the feature is used to extract specific characteristics from images. Here SIFT is used to extract characteristics. SIFT is a computer vision algorithm for identifying and defining local features in images. SIFT key points of objects are first taken from a collection of reference images and saved in the database. An object is recognized in a new image by comparing each feature from the new image to this database individually, and identifying matching features for candidates based on Euclidean distance from their feature vectors. In the new picture, subsets of key points agreeing on the object and its position, size and orientation are selected from the full set of matches to filter out great matches.

- ***CLASSIFICATION:*** This is the final step in which the input picture features such as contrast, correlation, homogeneity, energy and entropy extracted using SIFT is compared and then listed with that of the infected RBC image features. The classifier is trained for this by providing the features extracted in the preceding level. Classification is carried out using SVM and CNN classifiers in the proposed methodology and then their respective output is measured and analysed.

(A)Support Vector Machine (SVM):

A hyper plane or set of hyper planes are constructed by SVM in a high- or infinite-dimensional space (Fig-5), which may be used for classification , regression, or other tasks such as detection of outliers. The classifier is trained here to enable classification of the input images into infected and normal RBC images. This is how you detect the presence of malaria parasite in the blood.

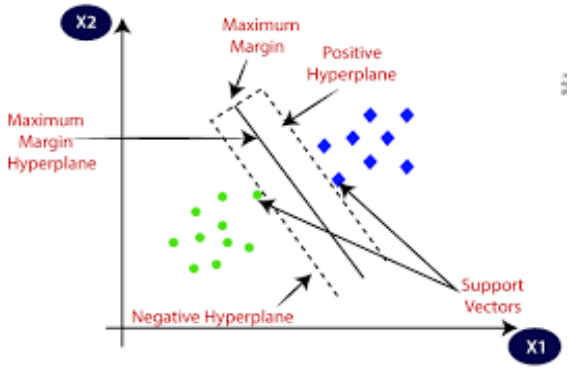

Fig-5: Hyperplane SVM algorithm [6]

(B)Convolutional Neural Network (CNN):

The CNN approach is second. Comparison of the output of the proposed work is made using CNN to determine which of these would yield the best results. CNN is a class of profound, feed-forward artificial neural networks that use a multilayer perceptron (Fig-6) variation designed to allow minimal preprocessing. These are also known as shift invariant or space invariant artificial neural networks (SIANN), based on their characteristics of shared-weight architecture and invariance in translation. Compared with other image classification algorithms, CNNs use very little pre-processing. This ensures that the network learns from the filters that were hand-engineered in conventional algorithms.

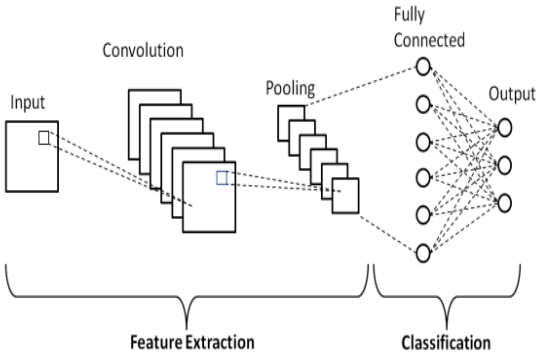

Fig-6: Neural network layer structure [7]

- ***IDENTIFICATION OF PARASITE***: From the outcome of the classification system using SVM and CNN, the final result i.e., the species of the malarial parasite in the given input image (thin blood smear) is identified and the results produced by both the classifiers SVM and CNN are compared for accuracy.

V. RESULTS AND DISCUSSION:

A dataset containing blood sample images is acquired from [15] in .png format and then converted to .jpeg format. 100 images per species type are used for SVM algorithm and 1000 images are used for CNN algorithm. Out of the total images in the dataset, 75% of them were used for training and 25% of them were used for testing.

(i) Output of SVM classifier:

(A)Species type1-Plasmodium falciparum

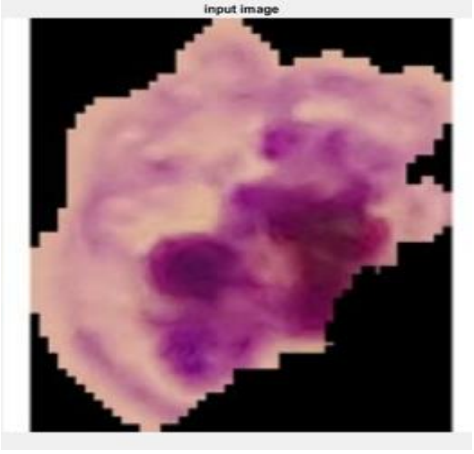

Fig-7: Input image of blood sample of species type-1 (SVM)

The blood sample image (Fig-7) is fed as input to the SVM algorithm. JPEG format conversion and Gray-scale conversion takes place in order to support pre-processing techniques.

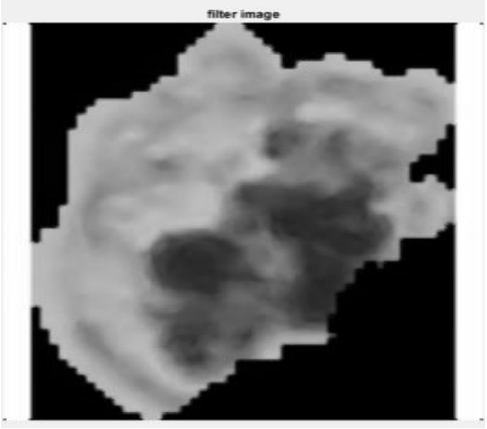

Fig-8: Filtered input image using median filter of species type-1 (SVM)

Resizing of images, de-noising of images (Fig-8) with the courtesy of median filter, contrast enhancement and morphological operations (dilation and erosion) are applied to the subjected image.

Fig-9: Binary converted image of the corresponding filtered image of species type-1 (SVM)

Binary conversion takes place through bi-modal threshold detection (Fig-9). Iterative method of clump splitting is conducted to recognize the cells of interest.

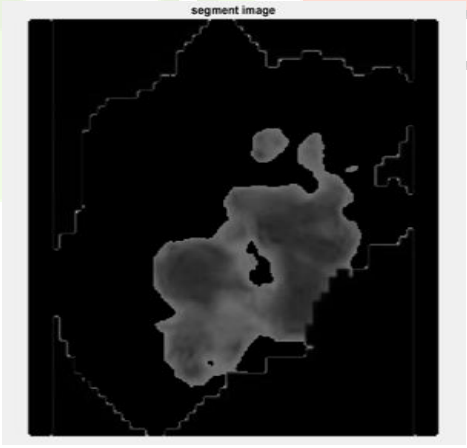

Fig-10: Segmented image of species type-1 (SVM)

Image segmentation process takes place (Fig-10). Feature extraction occurs with the help of SIFT algorithm. Outlier detection removes futile details in the background.

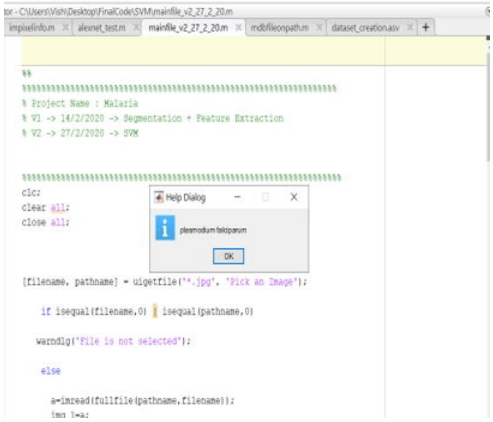

Fig-11: Species type-1 malarial parasite identified (SVM)

Final stage is where classification process takes place. Identification process is ensued by a dialog box indicating the type of malarial species (Fig-11).
In SVM algorithm, every blood sample image from the acquired dataset undergoes the same procedural flow similar to the steps explained in the previous segment for species type-1: Plasmodium *falciparum*. Species types 2, 3 and 4 are subjected to the same conditions as briefly explained earlier.

(B)Analysis of statistical features:

Statistical features such as Contrast, Correlation, Energy, Homogeneity, and Entropy of the four types of species are evaluated. The values are obtained using the GLCM matrix and other in-built functions exclusive for MATLAB environment using the Scale Invariant Feature Transform (SIFT).

Table-1: Statistical feature analysis of the four type of species

Species type / Statistical features	Plasmodium *falciparum*	Plasmodium *vivax*	Plasmodium *malariae*	Plasmodium *ovale*
Contrast	204.8366	234.2128	146.0947	210.5449
Correlation	0.5011	0.5030	0.5649	0.9107
Energy	0.6830	0.7486	0.7872	0.4943
Homogeneity	0.8968	0.9200	0.9316	0.8372
Entropy	1.8376	1.4551	1.3216	3.5631

- ***Contrast (CTR):*** Obtained by the change in brightness of image from one pixel to another, pointed by the data cursor (in-built feature).

$$CTR = \sum_{n=0}^{G-1} n^2 \left\{ \sum_{i=1}^{G} \sum_{j=1}^{G} P(i, j) \right\}, |i - j| = n \qquad (1)$$

Where, P- intensity value of the pixel
n- order
G- total number of pixels in the image
(i, j)- co-ordinates of the pixel

- ***Correlation (COR):*** Correlation is a measure of gray level linear dependence between the pixels at the specified positions relative to

each other.

$$COR = \sum_{i-0}^{G-1}\sum_{j-0}^{G-1}\frac{(i-\mu_i)(j-\mu_j)P(i,j)}{\sigma_i\sigma_j} \quad (2)$$

$$\mu_j = \sum_{i-0}^{G-1}\sum_{j-0}^{G-1} jP(i,j) \quad (3)$$

$$\mu_i = \sum_{i-0}^{G-1}\sum_{j-0}^{G-1} iP(i,j) \quad (4)$$

$$\sigma_j = \sum_{i-0}^{G-1}\sum_{j-0}^{G-1} (j-\mu_j)^2 P(i,j) \quad (5)$$

$$\sigma_i = \sum_{i-0}^{G-1}\sum_{j-0}^{G-1} (j-\mu_i)^2 P(i,j) \quad (6)$$

Where, P- intensity value of the pixel
(i,j)- co-ordinates of the pixel
G- total number of pixels in the image
μ- mean
σ – variance

- ***Homogeneity (IDM):*** It is the uniformity in composition. Also known as Inverse Difference Moment (IDM).

$$IDM = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1}\frac{1}{1+(i-j)^2}P(i,j) \quad (7)$$

Where, P- intensity value of the pixel
(i,j)- co-ordinates of the pixel
G- total number of pixels in the image

- ***Energy (E):*** Determines the intensity of pixels. Calculated as follows:

$$E = \sum_{i=0}^{G-1}[P(i)]^2 \quad (8)$$

Where, P- intensity value of the pixel
(i,j)- co-ordinates of the pixel
G- total number of pixels in the image

- ***Entropy (H):*** Statistical measure of randomness used to characterize texture of image. Calculated as follows:

$$H = -\sum_{i=0}^{G-1} P(i)\log_2 P(i) \quad (9)$$

Where, P- intensity value of the pixel
(i,j)- co-ordinates of the pixel
G- total number of pixels in the image

(ii)Output of CNN classifier:

(A)Species type1-Plasmodium falciparum

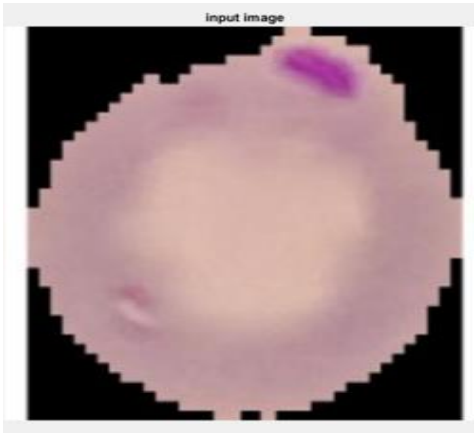

Fig-12: Input image of blood sample of species type-1 (CNN)

Image acquisition (Fig-12) process takes place in the CNN algorithm, which is an efficient and time-saving algorithm.

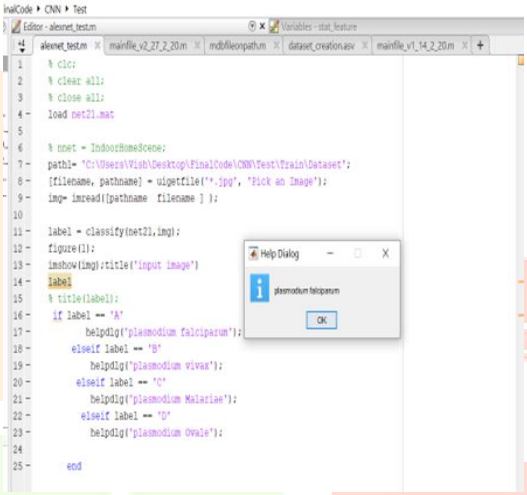

Fig-13: Species type-1 malarial parasite identified (CNN)

The type of malaria parasite present in the input blood sample image is identified (Fig-13) using CNN algorithm.
In CNN algorithm, every blood sample image from the acquired dataset undergoes the same procedural flow similar to the steps explained in the previous segment for species type-1: Plasmodium *falciparum*. Species types 2, 3 and 4 are subjected to the same conditions as briefly explained earlier.

V. COMPARATIVE ANALYSIS:

Table-2: Number of case types for existing and proposed systems

CASE TYPE	EXISTING SYSTEM	PROPOSED SYSTEM
TRUE POSITIVE	50	75
FALSE POSITIVE	8	3
FALSE NEGATIVE	10	5

Table-3: Accuracy calculation

TYPE	SVM (%)	CNN (%)
EXISTING SYSTEM	83.33	86.21
PROPOSED SYSTEM	93.75	96.15

Based on the values of the features extracted using GLCM matrix and SIFT algorithm, two tables are drawn for better interpretation. With a pre-set 75% of the total acquired images used for training and the remaining 25% used for testing, the number of case types of existing and proposed systems are estimated and presented in the table. The three case types are:

- True Positive- A test result that correctly indicates that the condition being tested for is present.
- False Positive- A test result which wrongly indicates that a particular condition is present.
- False Negative- A test result which wrongly indicates that a particular condition is absent.

The accuracy of existing system and proposed system, when subjected to SVM and CNN algorithms is calculated using formula:

$$\text{Accuracy} = \left[\frac{True\ Positive}{True\ Positive + False\ Negative}\right] * 100 \quad (10)$$

The existing system data is acquired from [1] [2] [3].

I. PERFORMANCE ANALYSIS:

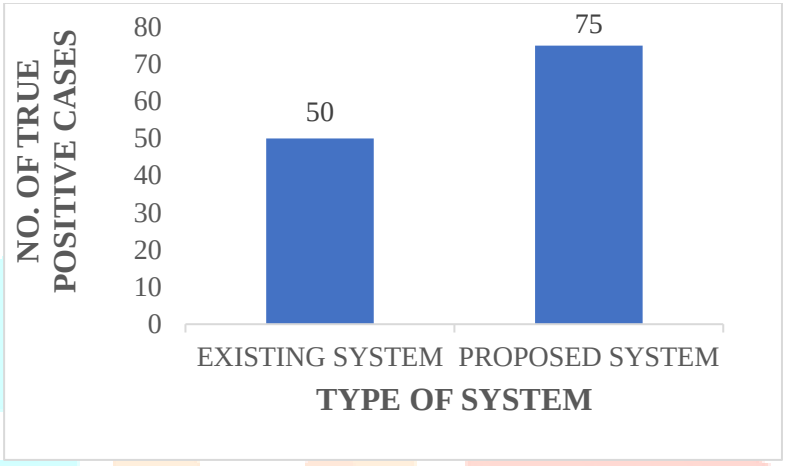

Fig-14: Number of true positive cases Vs. Type of system

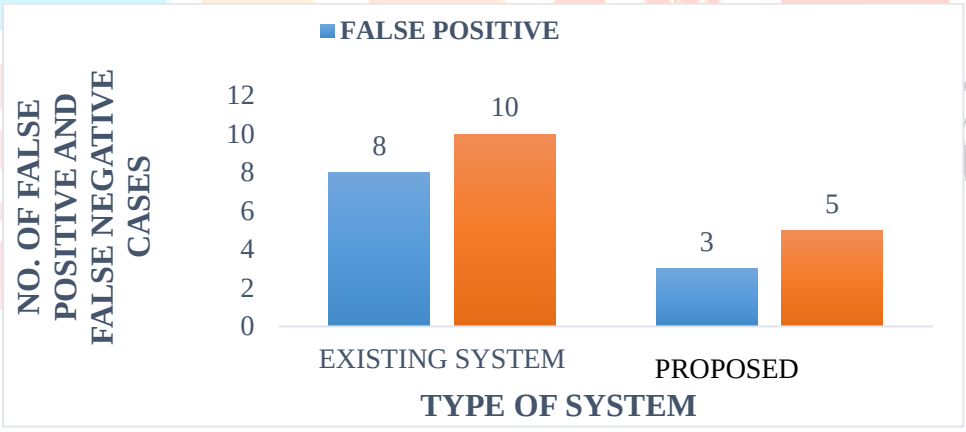

Fig-15: Number of false positive and false negative cases Vs. Type of system

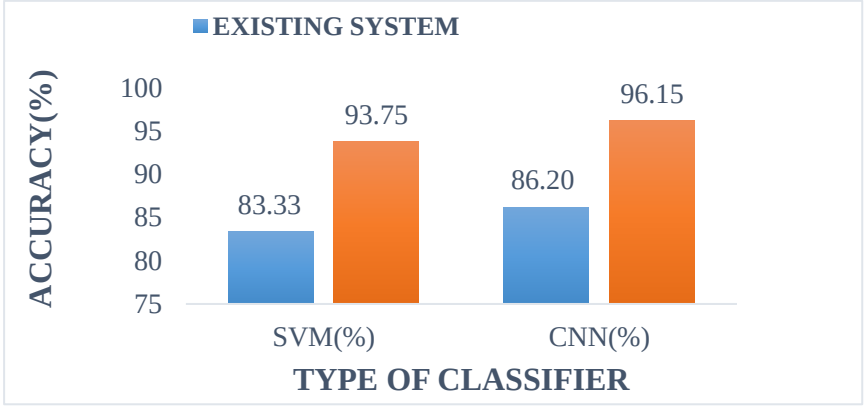

Fig-16: Accuracy classification between the two classifiers

Performance analysis is done between the existing and proposed systems and the results are represented using bar charts. The number of True Positive cases for the existing and proposed systems are assessed (Fig-14). A significant increase in the number of True Positive cases for the proposed system is represented. The number of false positive and false negative cases is represented against the type of

systems considered (Fig-15). The proposed system is factually better than its counterpart. The accuracy (in %) value of the two classifiers are compared (Fig-16).

I. CONCLUSION:

This paper proposes an automated diagnosis of malarial parasites present in the blood cell images. Hence, the parallax errors and time consumption which are the major drawbacks of the conventional microscopy method can be reduced to a greater extent. The classification process to identify the species of the malarial parasite in the blood cell image is done by two classifiers-SVM and CNN. From the outcomes obtained, CNN was found to be more effective in the classification with higher accuracy than SVM.

This proposal can further be extended by acquiring large number of sample images i.e., larger training and test datasets than which are available now and considering more features of the input image. This can improve the accuracy and reduce the errors in detection further.

REFERENCES:

[1] Hassan Abdelrahman Mohammed and Iman Abuel Maaly Abdelrahman (2017) 'Detection and Classification of Malaria in Thin Blood Slide Images', International Conference on Communication, Control, Computing and Electronics Engineering.

[2] Jyothi, R. and Sony, P.L. (2018) 'Automated Identification and Classification of Malarial Parasites in Thin Blood Smear Images', International Research Journal of Engineering and Technology.

[3] Kristofer E. delas Penas, Pilarita T. Rivera and Prospero C. Naval Jr., (2017) 'Malaria Parasite Detection and Species Identification on Thin Blood Smears using a Convolutional Neural Network', IEEE/ACM International Conference on Connected Health.

[4] Dataset of giesma stained thin blood cell images (input images). URL:
https://www.kaggle.com/iarunava/cell-images-for-detecting-malaria
https://lhncbc.nlm.nih.gov/publication/pub9932

[5]Stages in the growth and development of a malarial parasite in a blood cell. URL:

https://www.biologydiscussion.com/parasitology/parasiticprotozoa/histologyof-malarial-parasite-sporozoa/62006

[6] Hyperplane SVM algorithm. URL:

https://www.javatpoint.com/machine-learning-supportvector-machine-algorithm

[7]Neural network layer structure. URL:
https://missinglink.ai/guides/convolutional-neuralnetworks/convolutional-neural-network-architecture-forgingpathways-future/

IJCRT