

# Malaria Parasite Classification Using Machine Learning Techniques

S. Regina Lourdhu Sugnathi<sup>1</sup>, B. Janani<sup>2</sup>

Dean of PG Science, Department of Computer Science, Mount Carmel College, Autonomous, Bangalore, India<sup>1</sup>

Student, Department of Computer Science, Mount Carmel College, Autonomous, Bangalore, India<sup>2</sup>

**Abstract:** Malaria is a disease caused by protozoan parasites of the genus Plasmodium that are transmitted through the bites of infected female mosquitoes. It mainly infects the red blood cells in the human body. Initially malaria may cause no symptoms but gradually it may throw up symptoms such as vomiting, headaches and eventually may result in coma. Some current malaria detection techniques include manual microscopic examination and Rapid Diagnostic Test. These approaches are vulnerable to human errors. Early detection of malaria can help in reducing the fatality rates and examination of features in blood cells may aid in vaccine development. This research work focuses on identifying the type of plasmodium using suitable machine learning model that would classify the type of plasmodium with greater accuracy. Binary classification using machine learning algorithms namely Convolutional Neural Network (CNN), Support Vector Machines (SVM), Random Forest (RF), Decision Tree (DT) and ensemble technique are implemented to predict the model with higher accuracy. A comparative analysis is performed to find the model with greater accuracy. The comparative analysis reveals that the CNN model has outperformed the other models with highest accuracy for each of the species classification.

**Keywords:** Histogram Oriented Gradients, Gabor, Kaze, Support Vector Machines, Decision Tree, Random Forest, Ensemble model, Convolutional Neural Network.

## I. INTRODUCTION

Malaria is one of the most dangerous infections like Tuberculosis and Coronavirus 2019 (Covid-19). There are five Plasmodium species that cause malaria in human. The types of species are: Plasmodium falciparum, Plasmodium vivax, Plasmodium malariae, Plasmodium ovale. Each of the parasite species goes through various stages during their development cycle. In chronological order, the stages are ring stage, trophozoite stage, schizont stage, and gametocyte stage. Accurate counts of parasites are not only important for malaria diagnosis in the human body but they are also necessary for administering the proper medication, effectiveness measurement of the drug and classification of the severity of the disease. Traditionally, this disease and the type of species is manually detected by pathologists with the help of a microscope by observing the parasite genus Plasmodium. Several real life problems involve the classification of data into the type of species.

Besides medical research and analysis, contemporary computer technology plays an important role in creating more tools that attempt to fight the disease. To enhance diagnosis, image analysis software's and machine learning strategies have been used to quantify parasitemia in microscopic blood slides. Machine learning and big data approaches are attractive research areas. Due to its dynamic solution style and learning capability, these approaches are applied in many real-world applications. Classification algorithms such as Support Vector Machines, Random Forest, Decision Tree can be used for image classification. The success in the accuracy rate of these classification algorithms help to implement classification models for many real world applications.

Deep learning is a class of machine learning algorithms and the models are based on Convolutional neural network. CNN is an appropriate model for image analysis as it consists of multiple layers for processing input data. CNN model basically consists of two computational processes such as feed-forward and back-propagation. The CNN use relatively minimum pre-processing compared to other image classification algorithms. The neural network learns the data automatically without human interference and hence reduces the error significantly.

The paper is organized as follows: Section 2 presents literature review of various techniques. Section 3 explains the methodology used. Section 4 presents the experimental results and Section 5 discusses the scope of this research work.

## **II. LITERATURE REVIEW**

Accuracy of classification model is based on the choice of features selected from the data set. Decision tree classifiers and Neural Network based classifiers provide better classification accuracy for certain problem domains. It has been proven by the researchers[1] that the combination of more than one classifier generally provide better performance in classification problems.

Deep Learning based classification of malaria using images was performed by the researchers [2]. Here the images are first converted to grayscale. To improve the quality of the acquired image, isolated noise pixels are eliminated by applying image morphological operation. Image Hough Circle transform was applied to detect disk like overlapped RBCs and the images are segmented for better classification. Deep algorithms such as LeNet-5, AlexNet and GoogleNet were used to learn inherent features of malaria cells. An SVM was trained on pre-selected features extracted from the same dataset. Simulation results showed that very high recognition accuracy could be achieved by these deep learning techniques. They not only achieved higher, but also have the advantage of being able to automatically extract multiple layers of features from the input data.

The authors[3] discuss the importance of deep learning technique using Convolutional neural networks for classification. Images are fed to the system and the system selects the most efficient features for training. These features are then passed through the artificial neural network layers. The model is built with three convolutional layers and each layer followed by a max pooling layer. Relu function was used as an activation function. The binary cross-entropy algorithm has been used. Adaptive Moment Estimation(ADAM) was selected as optimization algorithm. The system provided very accurate results and with an accuracy of 0.95%.

Another Deep learning approach used to detect parasite infected red blood cells has been discussed by Zhaohui Liang et.al.,[4] Convolutional Neural Network model has been applied. The study compares the accuracy between a CNN model and transfer learning model. Deep learning has been implemented by both training a newly configured CNN model and transfer learning in order to evaluate its use for malaria blood smear classification. A pre-trained AlexNet has been used as the feature extractor for transfer learning. It is linked to a conventional SVM classifier to implement transfer learning as a comparison to the CNN model. The proposed CNN model worked well in sensitivity, specificity, and precision compared to transfer learning model.

Automatic method of detecting malaria parasite using image processing and comparison of various machine learning techniques to identify the suitable model was presented by the Researchers[7][8]. The images are converted to mask image. It has been observed that there were still some contiguous problems and to overcome that the watershed segmentation has been used. Features are extracted from each image using Speeded Up Robust Features method. The extracted features are trained using different machine learning algorithms. Later machine learning techniques namely Linear SVM, Fine Gaussian SVM, Cosine KNN, Boosted tree, Subspace KNN are compared. It is notable that Machine learning algorithms are most powerful tool for the classification of plasmodium parasite for the detection of malaria. Among the classifiers implemented Fine Gaussian SVM performs better in the classification tasks and Subspace KNN has the best overall performance.

In 2019 a study of various models of Deep Learning has been done and monitored to understand the models that would provide a better accuracy and faster resolution than previously used deep learning models[9]. It presents an outline of published papers to sketch pathway to improve deep learning algorithm for malaria parasite detection. The models used for comparison are Densenet121, VGG16, Alexnet, Resnet50, FastAI, and Resnet101. Resnet-50 gives very excessive matthews correlation coefficient, with a good overall performance than the other algorithms used under consideration. The results show that Resnet 50 model gave the highest accuracy.

## **III. FRAMEWORK**

The process flow diagram of the classification of the malarial species combined with model evaluation component is depicted in Figure 1. The image dataset for the work has been taken from Kaggle data repository. The images are pre-processed for appropriate feature extraction. In feature extraction step the most dominant features are extracted from the images. The extracted features are then sent as the input to the classification model. The classification model uses these features to perform the binary classification for each of the species. Classifiers namely Decision Tree, Random Forest and Support Vector Machine are used for classification. An ensemble of these models using Bootstrap Aggregation and CNN model are also used for classification of the species. The images of four different Plasmodium are shown in Figure 2.

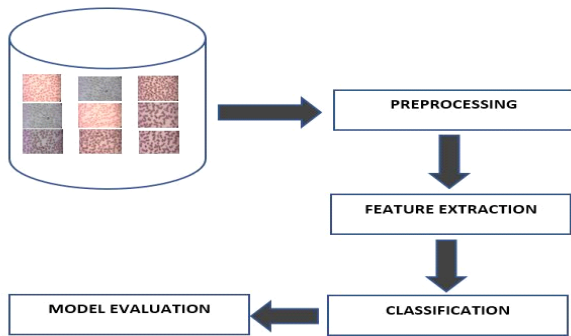


Figure 1: Workflow Diagram

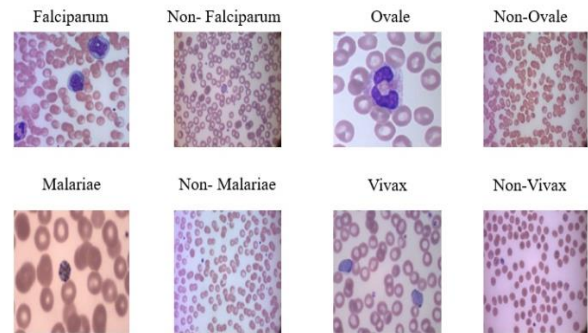


Figure 2: Plasmodium Images

The following sections describe in detail various steps undertaken at each stage as per the components in the process flow diagram.

### A. Data Pre-processing

The malaria species dataset has been used for binary classification between each of the species as falciparum cell vs non-falciparum cell, ovale cell vs non-ovale cell, malariae cell vs non-malaria cell, and finally vivax cell vs non-vivax cell. In data pre-processing step, the cell images are first converted to gray-scale images in order to reduce variations and noise in the images, and to standardize each of the cell images.

### B. FEATURE EXTRACTION

In this stage, the features from the images are extracted and converted into feature vectors. Different feature extraction methods such as Gabor, HOG and Kaze have been used.

❖ **Gabor** : It is a linear filter used for feature extraction and texture analysis from the images, which basically means that it examines whether there is any particular frequency content in the image in specific directions in a localized region around the area or region of analysis. Figure 3 shows the extracted Gabor features in the cell images.

❖ **Histogram of Oriented Gradients(HoG)** : This is implemented by extracting gradient and orientation of the edges of an image. These orientations are computed in small portions. HOG generates a histogram for each of these regions respectively. The histograms are generated using the gradients and orientations. Figure 4 shows the extracted HoG features of the cell images.

❖ **Kaze** : The images are analysed and features are described in a nonlinear scale space by keeping significant image details after removing noise from the images. Figure 5 shows the extracted KAZE features of the cell images.

The extracted features from the cell images using the above three filters are stored in a pickle file which is then used as the input to the classifiers.

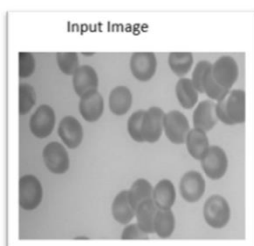


Figure 3: Gabor Features

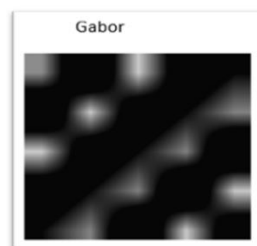


Figure 4: HoG Features

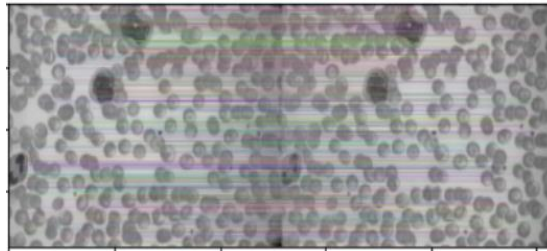


Figure 5. KAZE Features

### C. CLASSIFICATION

The binary classification between each of the species is performed using the mentioned algorithms. In the first phase the dataset of each species is organized and trained using CNN algorithm. This model learns the features automatically through various layers. Images are organized as training and test data of which 80% is used as training data and 20% is used as validation data. The image dataset are then loaded to the classification model and trained using CNN. The learning model is fit using functions imported from Keras library with sub sampling techniques, convolution filters, optimizers and activation functions. Later the model evaluation is performed.

In the next phase the dataset is trained using algorithms namely Decision Tree, Random Forest, Support Vector Machines. Firstly the features are extracted using different feature extraction techniques namely Gabor, HOG and Kaze. The extracted features are then stored in pickle file and taken as the input for the classification model. Once the models are trained an ensemble technique is performed to predict the species with the highest accuracy.

### IV. EXPERIMENTAL ANALYSIS

The dataset is categorized as falciparum and non-falciparum images, ovale and non-ovale images, malariae and non-malaria images, and vivax and non-vivax images. This dataset has been divided in the ratio 80:20 for Training and Validation purpose. CNN technique has been implemented. Activation function Rectified Linear(ReLU) is used. Loss function binary cross-entropy have been applied. The accuracy of the model trained with Plasmodium falciparum images is shown in Figure 6:

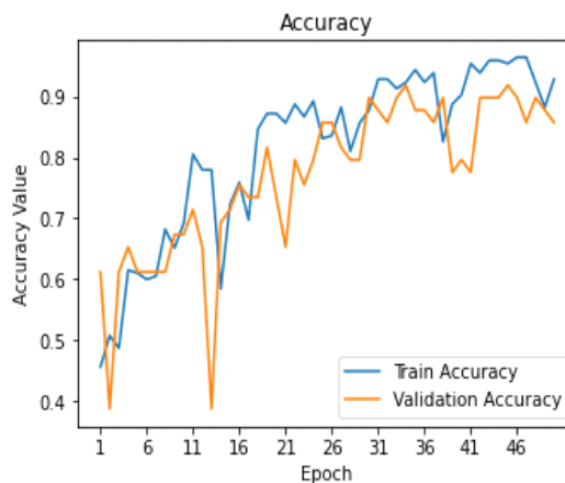


Figure 6. Plasmodium falciparum model accuracy

The accuracy of the model trained with *Plasmodium ovale* images is shown below in Figure 7:

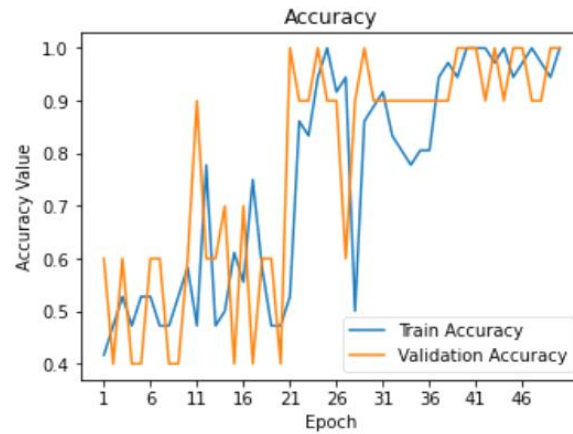


Figure 7. *Plasmodium ovale* model accuracy

The accuracy of the model trained with *Plasmodium malariae* images is shown below in Figure 8:

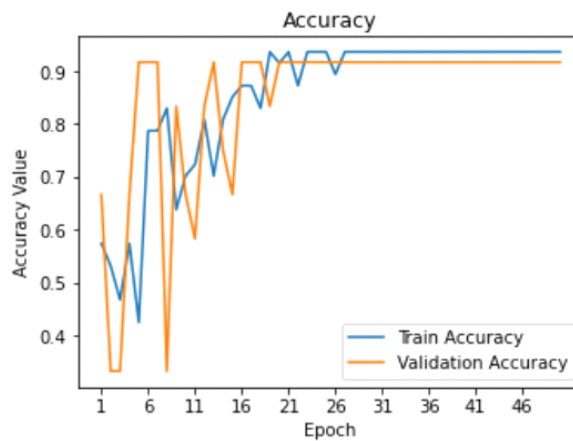


Figure 8. *Plasmodium malariae* model accuracy

The accuracy of the model trained with *Plasmodium vivax* images is shown below in Figure 9:

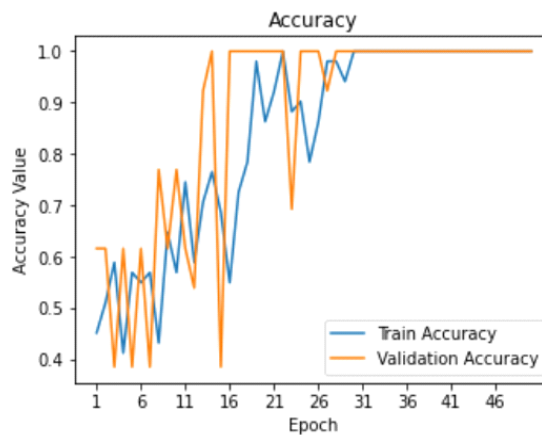


Figure 9. *Plasmodium vivax* model accuracy



The experimental results of the various Machine Learning classifiers used for performing binary classification between each of the species is listed below. In Table 1:

Sl. no	Species name	Decision Tree	Random Forest	SVM	Ensemble	CNN
1	P.falciparum	0.60	0.68	0.65	0.61	0.87
2	P.ovale	0.68	0.78	0.78	0.84	1.00
3	P.malariae	0.70	0.70	0.62	0.75	0.97
4	P.vivax	0.76	0.92	0.76	0.84	1.00

Table 1: Comparison of Model Accuracy

## V. CONCLUSION AND SCOPE

From the results, it could be concluded that Machine learning algorithms are a very powerful tool for the classification of the type of Plasmodium species. Among the four classifiers implemented, the Convolutional Neural Network model has the overall best performance compared to the other models. It has been observed that automatic learning of features by Convolutional Neural Network has outperformed in comparison to extracting features from the images and administering it as the input to the classifiers. The accuracy of the model could be improved by increasing the size of the image data set for training and validation. The hidden layers in the model can be increased to include complex features. The learning rate could also be improved for better performance. This work could be extended using multi class classifier to simultaneously address all the malarial species in place of binary classifier. Understanding the necessary and important features along with the application of appropriate feature extraction techniques will undoubtedly enhance the results.

## REFERENCES

- [1] S Regina Lourdu Suganthi, U K Sneha, Shwetha S, "Diabetic Retinopathy Classification Using Machine Learning Techniques", IJETT, Volume 68 , 2020
- [2] Yuhang Dong<sup>1</sup>, Zhuocheng Jiang<sup>1</sup>, Hongda Shen<sup>1</sup>, W. David Pan<sup>1</sup> Lance A. Williams<sup>2</sup>, Vishnu V. B. Reddy<sup>2</sup>, William H. Benjamin, Jr.<sup>2</sup>, Allen W. Bryan, Jr. "Evaluations of Deep Convolutional Neural Networks for Automatic Identification of Malaria Infected Cells", 2017
- [3] Soner Can Kalkan , Ozgur Koray Sahingoz "Deep Learning Based Classification of Malaria from Slide Images", 2019
- [4] Zhaohui Liang, 1 Andrew Powell,<sup>2</sup> Ilker Ersoy,<sup>3</sup> Mahdih Poostchi,<sup>4</sup> Kamolrat Silamut,<sup>5</sup> Kannappan Palaniappan,<sup>4</sup> Peng Guo,<sup>6</sup> Md Amir Hossain,<sup>7</sup> Antani Sameer,<sup>8</sup> Richard James Maude,<sup>5</sup> Jimmy Xiangji Huang, 1 Stefan Jaeger,<sup>8\*</sup> George Thoma<sup>8</sup> "CNN-Based Image Analysis for Malaria Diagnosis", IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2016
- [5] Alif Bin Abdul Qayyum\*, Tanveerul Islam and Md. Aynal Haque "Malaria Diagnosis with Dilated Convolutional Neural Network Based Image Analysis", IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON), 2019
- [6] Gautham Shekar, S. Revathy, Ediga Karthick Goud "Malaria Detection using Deep Learning", International Conference on Trends in Electronics and Informatics (ICOEI 2020), 2020
- [7] Divyansh Shah, Khushbu Kawale, Masumi Shah, Santosh Randive, Rahul Mapari "Malaria Parasite Detection Using Deep Learning (Beneficial to humankind)", International Conference on Intelligent Computing and Control Systems (ICICCS 2020) ,2020
- [8] Adedeji Olugboja, Zenghui Wang "MALARIA PARASITE DETECTION USING DIFFERENT MACHINE LEARNING CLASSIFIER,International Conference on Machine Learning and Cybernetics,2017
- [9] Satabdhi Nayak, Sanidhya Kumar, Mahesh Jangid "Malaria Detection Using Multiple Deep Learning Approaches", International Conference on Intelligent Communication and Computational Techniques (ICCT) , 2019