International Journal of Computer Trends and Technology
ISSN: 2231–2803 / https://doi.org/10.14445/22312803/IJCTT-V73I8P101

Volume 73 Issue 8, 1-6, August 2025

Original Article

Accuracy of Random Forest-Based Model for Malaria Parasite Prediction

Abdurrahman Zangina Abdullahi[1], Ishola Dada Muraina[2]

[1,2]*Information and Communication Technology Department, Faculty of Computing, Northwest University, Kano, Nigeria.*

[2]*Corresponding Author : ishod2001@gmail.com*

Received: 05 June 2025 Revised: 09 July 2025 Accepted: 28 July 2025 Published: 13 August 2025

Abstract - *Infectious diseases like Malaria have reportedly been the most prominent amongst the communicable diseases and have led to the death of approximately 435,000 annually in the world, while the majority of these fatalities occur in sub-Saharan Africa. Despite the deployment of heavy investment and strategy to mitigate or eradicate the malaria parasite, specifically in Sub-Saharan Africa, a high and upward trend of malaria cases is still being recorded. This study aims to examine the viability of the Random Forest Algorithm to predict the presence of the malaria parasite in patients accurately. The study presents a model based on the Random Forest Algorithm using MobileNetV2 as feature extraction. Meanwhile, the study considers some metrics, such as Precision, Recall and F1-Score, to further determine the performance of the model towards predicting the accuracy of malaria parasite in patients. The results confirm the accuracy of the Confusion Matrix classifier in predicting malaria parasite cases, while the model shows high accuracy and performance. The study contributes to the Health Informatics-Based Machine Learning domain towards predicting the Malaria parasite in patients.*

Keywords - *Predictive Model, Random Forest Algorithm, MobileNetV2, Malaria Parasite, Infectious Disease, Health Informatics.*

1. Introduction

Infectious diseases have been classified as one of the diseases that quickly claim the lives of people as a result of their capacity to spread from human-to-human and animal-to-human [1]. This disease cuts across Malaria, Tuberculosis, Measles, Zika Virus, Pertussis, Influenza, SARS and many more that have the tendency of taking the lives of individuals if proper care is not taken at the earliest time. Studies have stressed that the causes of Infectious diseases have been traced to the existence of pathogenic microorganisms such as viruses, bacteria, parasites and fungi in the hosts, which can be transmitted from one person to another [2]. Therefore, diseases like Malaria have reportedly been the most prominent amongst the communicable diseases in Nigeria. However, a lot of measures have been put in place to mitigate or eradicate the spread of these communicable diseases. At the same time, the government and some agencies have created awareness about their attributed dangers.

Malaria, as one of the most prevalent infectious diseases with substantial health implications, has a long history that dates back to the 16th century. It is a severe illness caused by the Plasmodium parasite, primarily transmitted through the bites of infected females Anopheles mosquitoes [3,4]. A study has described the infection processes in patients, which begin with a migration of mature parasites into the liver and later to the bloodstream, infecting the red blood cells within a few days [4]. On the other hand, Malaria has the potential to be transmitted through different ways other than mosquito bites, such as organ transplants, blood transfusions and utilization of contaminated syringes and needles with infected blood [5]. This portrays that many people are prone to the risk of being infected, which may vary in different regions of the world as a result of disparities in the factors that induce malaria parasites in patients.

According to the World Health Organization (WHO), Malaria remains a leading cause of death globally, claiming approximately 435,000 lives annually, while the majority of these fatalities occur in sub-Saharan Africa. Studies have stressed that Malaria, as a disease, has been the major cause of child mortality in Africa, with 247 million infected people in the year 2022 as being recorded globally [6]. Meanwhile, a significant investment has been made to control Malaria by the WHO and Global Technical Strategy by voting for a sum of $6.4 billion yearly, hoping to reduce malaria incidence by 90% by the end of 2023 [7]. Despite the heavy investment and strategy being made to mitigate or eradicate the Malaria parasite by the WHO and other related health agencies, there is still a high and upward trend of Malaria cases in Sub-Saharan Africa [6]. Thus, there is a need to focus more on the early prevention of Malaria by working on the prediction of the accuracy of Malaria parasites in patients. Hence, this study

provides a model for predicting the Malaria parasite in patients through a Machine Learning Model, the Random Forest Algorithm.

2. Related Works

Malaria is globally known as an incidence threat among the group of infectious diseases caused by a class of parasite called Plasmodium. The result of being infected by the Plasmodium represents a significant illness in the world, especially in some endemic regions [3,8,4]. This implies that Malaria has been tagged as a significant health issue for decades among inhabitants in Tropical and Sub-Tropical countries of the world [9]. There has been a report that Malaria cases in the world by the end of the year 2022 surpassed pre-pandemic cases as recorded during the recent global Lockdown, with 249 million cases. Thus, many measures have been established to eradicate its existence through predictive techniques, including Mathematical Modeling, Statistical Modeling and Machine Learning Algorithms [10]. These approaches help in examining the data, handling its processing and forecasting outcomes based on previous experiences [7].

The study of [11] has argued that the birth of Artificial Intelligence is seen as a technological approach to address the issues in the healthcare industry. This has led to the promulgation of Machine Learning models like the Random Forest algorithm, which uses a dataset to identify relationship among some attributes and make predictions towards providing solutions or treatments to the issue under investigation, towards achieving some tasks [12]. Researchers have argued that the Random Forest algorithm is known for its popularity in achieving accuracy in its calculations in different areas of work, including the healthcare domain. Thus, it is capable of accepting large datasets, which can be numerical and categorical [6]; hence, it is suitable for predicting malaria infection in patients. Meanwhile, other studies have explored different Machine Learning models on the classification of diabetes and further checked their performance in dealing with the disease. Thus, their outcome places Random Forest algorithms above the other models in terms of accuracy of diagnosing diabetes in patients. Besides that, the Random Forest Algorithm can be termed as a predictive analytics-based machine learning as a result of its capacity to extract the current data together with the historical data to predict future occurrences [13]. This implies that the Random Forest algorithm possesses higher power to ascertain the likelihood of events.

The study of [6] explores the quantitative potential of the Random Forest algorithm towards accurately predicting Malaria parasite alongside other algorithms; MLR, ANN and ANFIS, while the obtained results show that the Random Forest algorithm possesses higher potential to predict Malaria parasite loads accurately. A related study by [14] introduces the MobForest package in R, thereby providing a framework for model-based recursive partitioning using the Random Forest algorithm, thus establishing its capacity to withstand complex scenarios. Besides that, there have been variations in the outcome of studies on Malaria predictions with different degrees of accuracy while using different models, as shown in Table 1.

Table 1. Malaria Detection Using Image Processing and Machine Learning [15]

Algorithm	Accuracy	Precision	Recall	F-Score
Cubic SVM	86.1	71.2	86.3	77.9
Linear SVM	79.2	51.2	84.3	63.87
Cosine KNN	74.4	70.2	64.7	67.33

The study of [15] depicts that current studies on the predictive accuracy of the Malaria parasite possess some level of inaccuracy in their outcomes due to the inadequacy of selecting the ideal metrics to assess the models. This has paved the way for the existing gap in understanding some required incidences for predicting the occurrence of the malaria parasite. It has been stressed that current techniques lack the Precision and efficiency needed for effective identification of Malaria parasite patterns [8], which may make it difficult to capture some subtle patterns of Malaria infection [15]. Therefore, this implies that there is a need to address the inaccuracy of predicting malaria parasites, which may be hidden in the bloodstream of patients. Hence, this study uses the Random Forest Algorithm with relevant evaluation metrics to achieve accurate Malaria parasite prediction.

3. Materials and Methods

This section explains the approaches used for model selection, model design, training of the model and the source of data used. The performance metrics are also discussed in detail.

3.1. Model Selection

This study bases its approaches on the study of [15] that uses Cubic SVM, Linear SVM and Cosine KNN as their classifiers, using Histogram-based feature extraction for analyzing Malaria parasite images. Thus, this study enhances [15] by replacing Histogram-based feature extraction with MobileNetV2, a transfer learning algorithm. The MobileNetV2 is known for its capacity to extract high-level, discriminative features from complex image datasets. Meanwhile, Machine Learning provides different models which suit different data types like images, sequences, numerical data and text. Thus, this study chose the Random Forest Algorithm as a result of its ability to handle classification tasks, which is useful in this Malaria parasite prediction study. Hence, a combination of MobileNetV2 with Random Forest is expected to produce sufficient accuracy and an effective predictive model for Malaria parasite prediction.

3.2. Model Design

The model begins with data loading and pre-processing images from the Malaria parasite dataset through the

ImageDataGenerator class, which applies a preprocessing function, preprocess_input, from the MobileNetV2 architecture. This is done to achieve data normalization, thus resizing to a target size of 224 by 224 pixels. Thus, the data is then organized into batches for training and testing purposes and represented in the procedural task as shown in Algorithm 1.

Algorithm 1: RF without optimization
1.0 Data Loading and Feature Extraction:
1.1 train_generator = load_data("train_data_directory")
1.2 test_generator = load_data("test_data_directory")
1.3 feature_extractor = build_feature_extractor()
1.4 X_train, y_train = extract_features(train_generator, feature_extractor)
1.5 X_test, y_test = extract_features(test_generator, feature_extractor)
2.0. Random Forest Training and Evaluation:
2.1 Create a Random Forest classifier (with default or chosen hyperparameters)
2.2 Train classifier on X_train, y_train
2.3 Predict on X_test
2.4 print classification report
2.5 End of program

Besides, the base model of MobileNetV2 excludes the top layer to allow the extraction of high-level features. The output from MobileNetV2 is then passed through a GlobalAveragePooling2D layer, which reduces the spatial dimensions of the feature maps to a single vector per image, summarizing the learned features efficiently in the model as shown in Figure 1 and the procedure presented in Algorithm 2.

Algorithm 2: RF with optimization
1.1 Define Snake Optimization Algorithm
1.2 defSnake_Optimization_Algorithm (classifier, hyperparameters, data, labels):
1.3 best_hyperparameters = None
1.4 best_performance = 0
1.5 snake population = initialize_snake_population (hyperparameters)
1.6 for each generation in range (max_generations):
1.7 for each snake in snake_population:
1.8 current_hyperparameters = snake.hyperparameters
1.9classifier.set hyperparameters (current_hyperparameters)
1.10 performance = classifier.evaluate (data, labels)
1.11 if performance & gt; best_performance:
1.12 best_hyperparameters = current_hyperparameters
1.13 best_performance = performance
1.14 snake.update_position(performance)
1.15snake_population = update_snake_population(snake_population)
1.16 return best_hyperparameters, best_performance
3.0 Load and preprocess data
3.1 data, labels = load_malaria_data()
3.2 data = preprocess_input (data, target_size = (224, 224))
4.0 Extract features using MobileNetV2
4.1MobileNetV2_base = load_MobileNetV2 (include_top = False)
4.2features = GlobalAveragePooling2D (MobileNetV2_base(data))
5.0 Initialize classifiers
5.1 Random_Forest_Classifier = RandomForest()
6.0 Optimize hyperparameters using the Snake Optimization Algorithm for Random Forest
6.1optimized_RF_hyperparameters,best_RF_ performance = Snake_Optimization_Algorithm (Random_Forest_Classifier,RF_hyperparameters, features,labels)
7.0 Set optimized hyperparameters
7.1Random_Forest_Classifier.set_hyperparameters (optimized_RF_hyperparameters)
8.0 Evaluate optimized classifiers
8.1performance_RF = Random_Forest _Classifier. evaluate (features, labels)
9.0 Output results
9.1print ("Optimized Random Forest Performance: ", performance_RF)
10.0 Save the optimized models
10.1save_model(Random_Forest_Classifier, "optimized_random_forest_model")
10.2 End of program

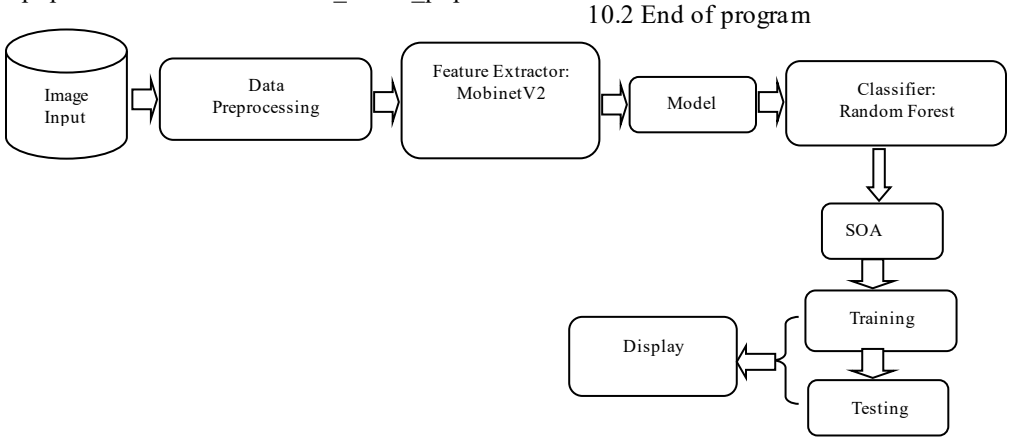

Fig. 1 Procedural model

Moreover, the pooling layer captures the global context of the image, which is used to differentiate between infected and uninfected cells in the Malaria dataset. Then, the extracted features are used as input for a Random Forest classifier to predict the presence of Malaria parasites, which is done by constructing multiple decision trees and aggregating their predictions to enhance accuracy and reduce overfitting. Hence, the classification results are presented while the Confusion Matrix is plotted to visualize the performance across different classes after training a classifier.

3.3. Model Training and Data Source

The model is trained so that the classifier can extract features to predict the status of cell image vis-à-vis Malaria infection. The model's training process uses 80% of the available data to build and fine-tune the classifiers pre-trained on the ImageNet dataset. The classifier is trained by constructing multiple decision trees, where each tree is built using a random subset of the training data. This method is known as an ensemble and helps reduce the prediction variance and improve generalization for unseen data. On the other hand, the remaining 20% of the data was reserved for testing the model's performance. Thus, the model's predictions are compared against the actual labels, and various performance metrics, such as Accuracy, Precision, Recall and F1 score.

The dataset used in this study is sourced from Kaggle. This well-known open-source platform provides a wide array of datasets for model development, testing and validation, thus facilitating the implementation of various Machine Learning Algorithms. The obtained dataset is divided into two main categories: training and testing. The first folder, labeled "Parasite," contains 220 images of blood smears with Malaria infections. The second folder, labeled "Uninfected," contains 196 images of blood smears free of Malaria. Indeed, the data testing is grouped into two: the parasite folder within the test-set containing 91 images, while the uninfected folder contains 43 images. This represents the division of data into training and testing sets, ensuring the model is adequately trained and obtains a fair assessment of its performance.

3.4. Performance Metrics

This study considers some metrics to achieve the objective of the study, which are accuracy, Precision, recall and F1 score for determining the performance of the algorithm on the subject under investigation [16].

Accuracy: is the percentage of accurate predictions, which is the ratio of the number of correctly classified instances to the total number of instances and mathematically represented below;

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}}$$

Precision: is the ratio of positively predicted instances among the retrieved instances, represented below;

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Recall is the metric that measures the ability of a model towards identify relevant instances, where a high recall value denotes a good model with regard to the correctness of data.

$$\text{Recall} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Negative (FN)}}$$

The True Positive (TP) represents the number of positive instances the model correctly identifies, and the False Negative (FN) represents the number of positive instances the model wrongly identifies as negative.

F1 Score: This metric is used to evaluate the machine learning model's performance by combining the precision and recall metrics.

$$\text{F1 Score} = \frac{2 * (\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})}$$

4. Result and Discussion

This study aims to examine the accuracy of the Random Forest algorithm in curbing the prevalence of Malaria in societies. The results are presented in two different ways: classification results with and without optimizations, whereby the values of the considered four metrics are compared with the study of Motwani et al. (2020), alongside their visualization using a Confusion Matrix. As shown in Table 2, the result of the experiment on the performance of the Random Forest algorithm without optimization shows an accuracy of 95%. At the same time, the Precision of the parasite class is set at 0.98. This implies a high proportion of correctly identified instances out of all instances predicted as positive.

Table 2. Random Forest Classification Report without Optimization

	Accuracy	Precision	Recall	F1-Score	Actual Data
Parasite	95%	0.98	0.95	0.96	91
Uninfected	100%	0.89	0.95	0.92	43

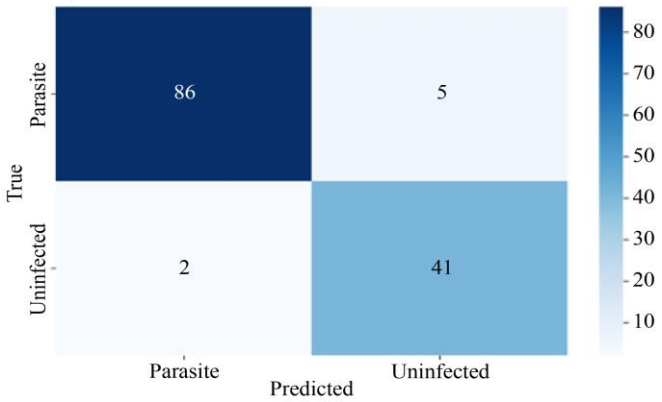

Fig. 2 Confusion Matrix for Random Forest without Optimization

Besides, the recall for the parasite class is found at 0.95, which indicates a good ability to identify actual positive cases correctly. Meanwhile, the F1-score, a balanced measure of Precision and Recall, agrees at 0.96 for the parasite class. On the other hand, in the aspect of the uninfected class, the Precision was 0.89, with a recall of 0.95, resulting in an F1-score of 0.92.

Moreover, the Confusion Matrix as the classifier for Random Forest, shown in Figure 2, portrays a high level of accuracy while identifying Malaria parasites in uninfected samples. The total of 86 out of 91 parasite samples are correctly classified as parasitic, while 5 are incorrectly labeled as uninfected. On the other hand, among 43 uninfected samples, only 2 are mistakenly classified as parasitic, with 41 correctly identified as uninfected. This distribution reflects that the classifier performs well while distinguishing between the two classes, with a low rate of false positives and false negatives.

In addition, the result of the Random Forest classifier with optimization, as shown in Table 3, also performs better compared to the existing classifiers that examine the Malaria parasites in patients. The accuracy of Random Forest in this regard is 96% with a Precision of parasite class at 0.99, Recall at 0.96 and F1-score at 0.97. Meanwhile, the uninfected class in Table 3 shows a Precision of 0.91, with a Recall of 0.98, while the F1-score is at 0.94.

Table 3. Random Forest Classification Report with Optimization

	Accuracy	Precision	Recall	F1-Score	Actual Data
Parasite	96%	0.99	0.96	0.97	91
Uninfected	100%	0.91	0.98	0.94	43

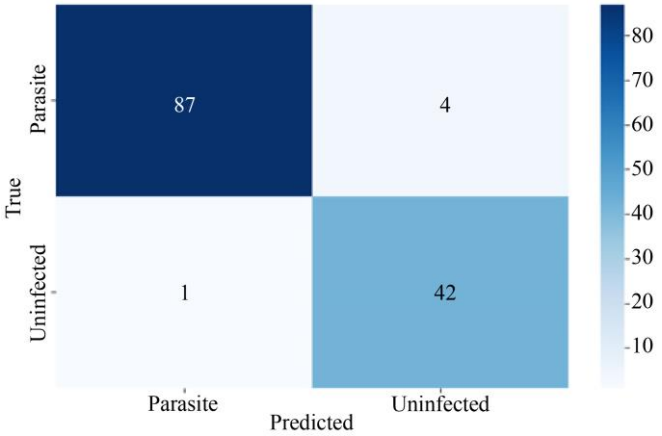

Fig. 3 Confusion Matrix for Random Forest with Optimization

The confusion matrix in Figure 3 shows that of all 91 parasitic samples, 87 are correctly classified as parasitic, while only 4 are incorrectly labeled as uninfected. Meanwhile, only 1 out of 43 unaffected samples is mistakenly classified as parasitic, and 42 are correctly identified as uninfected. This is a good reflection of the distribution of classifying analysis, which shows a low rate of false positives and false negatives.

5. Conclusion

This study aims to examine the viability of the Random Forest algorithm to predict the accuracy of the Malaria parasite. The study uses Precision, Recall and F1-score to measure the performance of a designed Random Forest-Based model towards achieving the accuracy of Malaria prediction. The MobileNetV2 is used for feature extraction, while the Confusion Matrix is used to confirm the accuracy of the classifier in predicting the Malaria parasite in patients.

Thus, the designed model achieved high accuracy with effective performance metrics, making it a promising tool for Malaria diagnosis in patients. Hence, the study contributes to the domain of Health Informatics-Based Machine Learning in the area of infectious disease prediction and eradication.

References

[1] Qiu Li et al., "A New Prediction Model of Infectious Diseases with Vaccination Strategies Based on Evolutionary Game Theory," *Chaos, Solitons & Fractals*, vol. 104, pp. 51-60, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[2] WHO,. Infectious Diseases, 2019. [Online]. Available: www.who.int/topics/infectious_diseases/en/

[3] Azam Mehmood Qadri et al., "A Novel Transfer Learning-Based Model for Diagnosing Malaria from Parasitized and Uninfected Red Blood Cell Images," *Decision Analytics Journal*, vol. 9, pp. 1-11, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] Kyle Manning, Xiaojun Zhai, and Wangyang Yu, "Image Analysis and Machine Learning-Based Malaria Assessment System," *Digital Communications and Networks,* vol. 8, no. 2, pp. 132-142. 2022. [CrossRef] [Google Scholar] [Publisher Link]

[5] Mosabbir Bhuiyan, and Md. Saiful Islam, "A New Ensemble Learning Approach to Detect Malaria from Microscopic Red Blood Cell Images," *Sensors International*, vol. 4, pp. 1-11, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[6] Dilber Uzun Ozsahin et al., "Quantitative Forecasting of Malaria Parasite using Machine Learning Models: MLR, ANN, ANFIS and Random Forest," *Diagnostics*, vol. 14, no. 4, pp. 1-13, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[7] Amit Kumar, Pankaj Verma, and Poonam Jindal, "Machine Learning Approach to Surface Plasmon Resonance Sensor Based on MXene Coated PCF for Malaria Disease Detection in RBCs," *Optik*, vol. 274, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] Charles Ikerionwu et al., "Application of Machine and Deep Learning Algorithms in Optical Microscopic Detection of Plasmodium: A Malaria Diagnostic Tool for the Future," *Photodiagnosis and Photodynamic Therapy*, vol. 40, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[9] Odu Nkiruka, Rajesh Prasad, and Onime Clement, "Prediction of Malaria Incidence using Climate Variability and Machine Learning," *Informatics in Medicine Unlocked*, vol. 22, pp. 1-12, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[10] Kate Zinszer et al., "Predicting Malaria in A Highly Endemic Country Using Environmental and Clinical Data Sources," *Online Journal of Public Health Informatics*, vol. 6, no. 1, 2013. [Google Scholar] [Publisher Link]

[11] S.C.A. Devadoss, "The AI Revolution in Healthcare Product Management," *International Journal of Computer Trends and Technology*, vol. 72, no. 2, pp. 1-8, 2024. [CrossRef] [Publisher Link]

[12] Alvin Rajkomar, Jeffrey Dean, and Isaac Kohane, "Machine Learning in Medicine," *New England Journal of Medicine*, vol. 380, no. 14, pp. 1347-1358, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[13] Manohar Sai Jasti, "Predictive Analytics in Data Engineering," *International Journal of Computer Trends and Technology*, vol. 72, no. 8, pp. 19-25, 2024. [CrossRef] [Publisher Link]

[14] Nikhil R Garge, Georgiy Bobashev, and Barry Eggleston, "Random Forest Methodology for Model-Based Recursive Partitioning: The Mobforest Package for R. *BMC*," *Bioinformatics*, vol. 14, pp. 1-8, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[15] K. Motwani, A. Kanojiya, C. Gomes, and A. Yadav, "Malaria Detection using Image Processing and Machine Learning," *International Journal of Engineering Research and Technology*, vol. 9, no. 3, pp. 39-44, 2020. [Google Scholar] [Publisher Link]

[16] A.K. Santra, and C.C. Josephine, "Genetic Algorithm and Confusion Matrix for Document Clustering," *International Journal of Computer Science*, vol. 9, no. 1, no. 2, pp. 322-328, 2012. [Google Scholar]