UNIABUJA Journal of Engineering and Technology
https://ujet.uniabuja.edu.ng/
ISSN: 2714-3236 (Online); 2714-3228 (Print)
Volume 2, Issue 1, 2025; 118-127

Development of a Malaria Progression Prediction System with Artificial Neural Networks (ANN) and Support Vector Machines (SVM)

Olatayo M. OLANIYAN[1], Kolade E. OLUWADARE[1], Henry C. ELUE[2] and Ayodeji I. FASIKU[3]

[1]*Department of Computer Engineering, Federal University, Oye-Ekiti, Nigeria*
[2]*Department of Computer Science, National Open University, Nigeria*
[3]*Department of Computer Engineering, Ekiti State University, Ado-Ekiti, Ekiti State, Nigeria*

Corresponding Author: olatayo.olaniyan@fuoye.edu.ng

Abstract
Malaria has been a major worldwide health problem, especially in areas with low availability of healthcare resources. Delaying intervention frequently results in higher fatality rates. This study aims to develop a Malaria Progressive Prediction System (MPPS) leveraging sophisticated machine learning approaches, specifically Support Vector Machine (SVM) and Artificial Neural Network (ANN), to address this important issue. A large Kaggle dataset containing clinical and laboratory data from malaria patients at different phases of the illness is used in this study. Multiple layers of linked artificial neurons are used to build the artificial neural network (ANN), which uses a backpropagation method for training. In contrast, the SVM approach discovers the optimal hyperplane for classification by transforming input data into a multidimensional space through supervised learning. The Malaria Progressive Prediction System (MPPS) continuously showed during the performance evaluation that the Artificial Neural Networks (ANN) model outperformed the Support Vector Machine (SVM). The ANN model had greater precision, accuracy, recall, and F1 score. This produces outcomes that ascertain that the ANN model is highly efficient in forecasting malaria progression. It outperforms SVM in accuracy and reliability, confirming its superiority in this important healthcare application.

Keywords: *Malaria, SVM, deep learning, ANN, machine learning, prediction system.*

1.0 Introduction

Malaria, caused by Plasmodium parasites transmitted through infected mosquitoes, remains a major global health challenge, particularly in regions with inadequate healthcare infrastructure (WHO, 2023). Globally, it causes a significant amount of illness and death, with approximately 229 million cases and 409,000 fatalities reported in 2022 (WHO, 2022). Timely identification and precise forecasting of malaria advancement can have a pivotal impact on enhancing patient results through facilitating prompt intervention and optimal allocation of resources.

In recent years, the medical field has experienced a rapid growth in adopting machine learning due to its ability to enhance disease diagnosis and prognosis. Comparing deep learning algorithm and the traditional machine learning algorithms to determine their effectiveness classifying malaria parasites images. Both algorithms have their pattern of learning from images and offer some edge in the feature extraction. Machine learning approaches such as SVM and ANN, were utilized for predictive modeling in various medical applications (Zhang, 2020). A computerized model that uses biological neural networks for both structural and operational design is called an Artificial Neural Network (ANN) (LeCun et al., 2015). Conversely, Cortes and Vapnik (1995) noted that SVM is a highly effective supervised learning method renowned for its ability to handle high-dimensional data and perform non-linear classification tasks with exceptional accuracy.

A thorough review of existing literature reveals a paucity research on the application ANN and SVM in predicting malaria progression. This study aims to contribute to the growing body of work in predictive analytics and malaria research. By evaluating the performance of ANN and SVM models, we aim to highlight their respective strengths and limitations in forecasting malaria progression, offering insights for subsequent clinical applications and research in the future.

SVM is generally known to achieve great performance with minimal data as it is focused on finding the optimal decision boundaries using support vectors, it is also less computationally intensive during training compared to ANN or other deep learning algorithms. While for ANN, it learns hierarchical feature representations directly from images oblivating the necessity for manual feature engineering, it scales well with large training data and improves performance as more data become available.

Received: 21-01-2025 / Accepted: 05-02-2025 /Published: 31-03-2025

This study intends to develop a Malaria Progression Prediction System (MPPS) that leverages computational models capable of identifying intricate patterns within data, particularly ANN and SVM. The research is based on a comprehensive dataset containing laboratory and clinical records of malaria patients at different stages of the disease. This dataset includes demographic details, symptoms, medical history, and diagnostic test results. The study covers data acquisition, preprocessing, model implementation, performance assessment, along with a comparison of the two machine learning techniques to determine their effectiveness in predicting malaria progression.

Malaria has continued to pose a substantial global health challenge, especially in tropical and subtropical areas, impacting millions of individuals annually. Gaining an express understanding of the general features and occurrence patterns of malaria is essential for implementing efficient measures to prevent, control, and manage the disease.

Malaria has remained a significant general health issue in sub-Saharan areas of Africa and beyond. The Sub-Saharan Africa region bore the burden of approximately 234 million malaria cases and 593,000 fatalities. These statistics represent a staggering 95% of all reported malaria cases and 96% of malaria-related deaths worldwide for that year. Studies has observed that children below the age of 5 are disproportionately impacted, accounting for 80% Accounted for 95% of malaria deaths in the WHO African Region in 2023.

Malaria exerts a substantial influence on public health, resulting in both illness and death. It exacerbates socioeconomic inequalities, impeding economic progress and perpetuating the cycle of poverty in areas affected by the issue. The illness can result in significant healthcare expenses, encompassing diagnostic examinations, therapy, and hospital admittance.

Precise and prompt Accurate identification of malaria is essential for effective treatment and disease control. Multiple diagnostic techniques are accessible to identify malaria, each possessing distinct advantages and disadvantages. Examining blood smears under a microscope is still considered the most dependable method for diagnosing malaria (CDC, 2021).

Machine learning algorithms are computational models capable of learning and improving their performance through experience, without requiring explicit programming. These models can identify patterns, relationships, and trends within complex datasets. In healthcare, commonly used machine learning techniques include random forests, artificial neural networks (ANN), decision trees, support vector machines (SVM), and various deep learning approaches (Ching et al., 2018).

Machine learning combine with deep learning has extensive applications in the healthcare industry. A notable field is medical imaging analysis, in which algorithms may scrutinize images from diverse modalities (such as MRI, X-rays, or CT scans) that aids in disease detection and diagnosis. Deep learning algorithms have shown a remarkable ability to accurately detect anomalies in radiological pictures, as reported by Esteva et al. (2017).

Numerous studies have investigated the application of Artificial Neural Networks (ANN) and Support Vector Machines (SVM) in malaria research. Artificial neural network (ANN) models were used by Leite et al. (2019) to forecast occurrences of malaria by utilizing meteorological and environmental data. Their research emphasized the capability of Artificial Neural Networks (ANN) to accurately represent intricate connections between climate variables and the spread of malaria.

2.0 Materials and Methods

The Malaria Progression Prediction System utilizing ANN and SVM is the main emphasis of this study. The Artificial Neural Network (ANN) model is the deep learning strategy that will be applied in this study. The dataset used in this research was obtained from Kaggle and has 27,558 images and has different folders of the infected and uninfected. The dataset was splitted into 20% for testing and 80% for training of the datasets. This approach was used to provide better accuracy in the proposed prediction system. Both the support vector machine and artificial neural network were employed to train the data, and confusion matrix, accuracy, recall, precision, F1-measure, and precision were used to assess the model's performance. It was deployed to a web application following assessment. Figure 1 displays the architecture of the created model.

Received: 21-01-2025 / Accepted: 05-02-2025 /Published: 31-03-2025

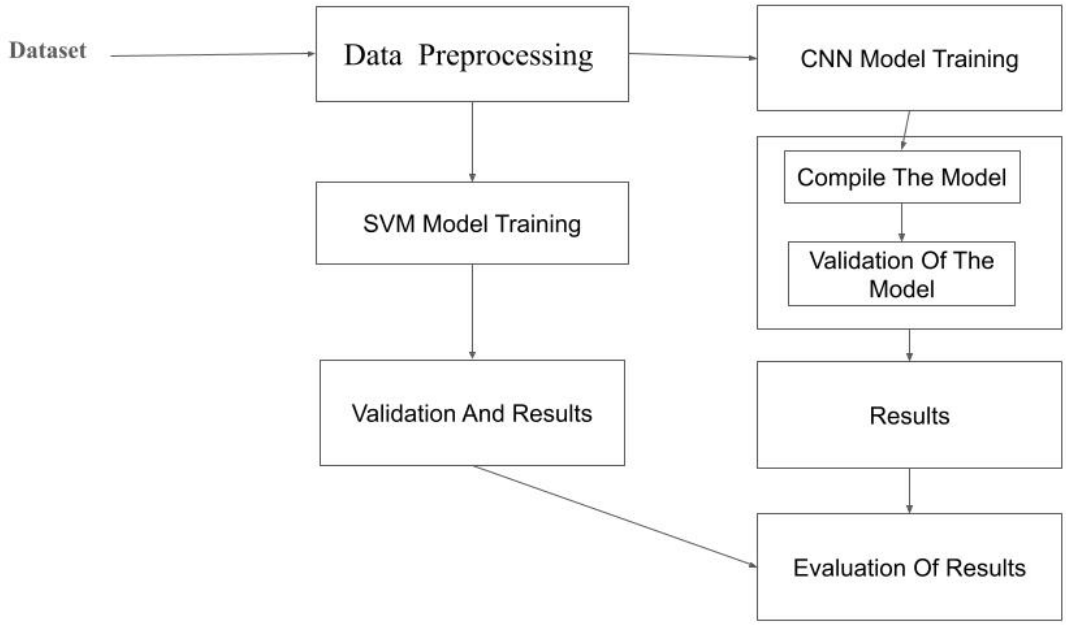

Figure 1: Block diagram of the Malaria Progression Prediction System

2.1 Design of the ANN-Based model

The CNN-based model's design and the performance analysis that follows are the two primary parts of the Artificial Neural Network (ANN)-based model architecture used to train the dataset.

2.2 Design of the CNN Model

In the development of the CNN model, there are major steps that would be involved in the model design and they include training the model and testing it before deploying to a web application

This would involve training the dataset, and in the dataset run the application for processing and it would classify the images into the ones that are infected or uninfected.

2.3 Implementation of the designed model

2.3.1 Classification and prediction algorithms (classifiers)

Classification is a method used to assign data into certain categories or groupings. The primary objective of a classification challenge is to determine the specific category or class to which a fresh set of data belongs. This research aims to enhance the accuracy of malaria prediction in patients through the application of deep learning methodologies.

2.3.1.1 Data collection

The dataset that was used in this project was obtained from Kaggle and it contains a log file of different parameters of test and has a size of 708MB. It contains images of the infected and the non-infected.

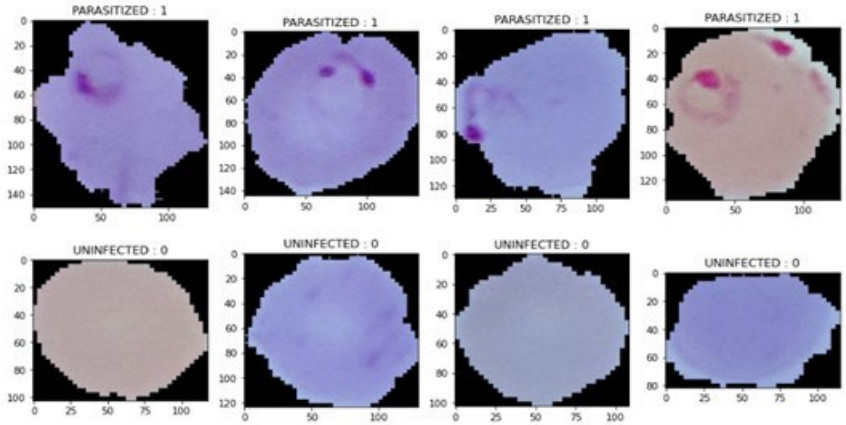

Figure 2: Infected and the Non-Infected Sample

2.3.1.2 Data pre-processing

This process is a phase in data mining involving altering or dropping of data before utilisation in order to ensure an effective performance. The dataset was cleaned and sampled, removal of some missing data and duplicate rows.

2.3.1.3 Data cleaning

In real-world scenarios, data often contain noise, incompleteness, and inconsistencies, necessitating data cleaning (or data cleansing) processes to address these issues. The steps involved in data cleaning typically include: Duplicate or irrelevant observations from the dataset are been removed, filtering unwanted outliers, fixing structural errors, Handling missing data and validation

2.3.3 Data splitting

Three distinct sets of data are utilized: training, testing, and validation. The training set contains the data necessary for constructing the model and the data used to qualify and assess the model's performance is found in the validation and testing sets, the X column contains the number of dataset and 19 labels while Y contains one label, the training set contains 591318 rows and 19 columns, and testing set contains 147830.

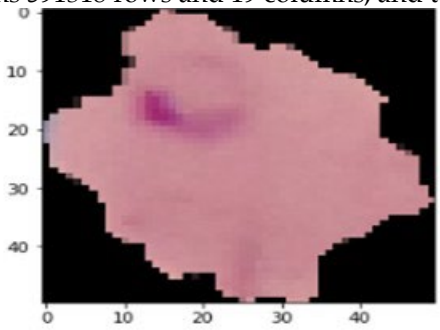

Figure 3: Infected Sample

2.3.4 Model training

Neural networks are employed to train data in the creation of malaria prediction models. The initial step in training the model is you set up your network layers and the hidden state. The hidden state's shape and dimension will be determined by the shape and dimension of the neural network. Then you loop through your inputs, feeding the ANN, the NN returns the output as well as a concealed state that has been updated.

2.3.5 Compiling the Artificial Neural Network model

The initial step in constructing the model is to define specific parameters, such as: loss function, optimizer and the metrics argument.

I. **Loss function**

The loss function will be used to determine the difference between the model's output and its real output.

II. **Optimizer**

The optimizer is used to update the weights and learning rate of neural networks. By minimizing the function, it addresses optimization problems.

III. **Metrics arguments**

The accuracy function will be used as a statistic to assess the recurrent neural network algorithm's performance. To validate the datasets, compare the accuracy and loss function for both the training and test datasets.

Using a different set of validation data, model assessment evaluates the trained model's actual performance. This process describes how effectively the model can generalize to new data that it hasn't encountered during training. Model performance is often assessed using a variety of measures, including accuracy, recall, and precision.

One of the most germane metrics in evaluating a model is accuracy, which is computed by comparing the sample labels with the predictions. Accuracy is the proportion of properly categorized samples among all samples. It provides an express assessment of the model's accuracy in assigning occurrences to the appropriate categories.

$$\text{Accuracy} = \text{True Negative} + \text{True Positive} \quad \text{(i)}$$

Received: 21-01-2025 / Accepted: 05-02-2025 /Published: 31-03-2025

True Negative + True positive + False Negative + False Positive

Out of all positive predictions the model makes, the precision is the percentage of real positive forecasts. Furthermore, it quantifies the precise number of expected clean samples that are clean, indicating the model's capacity to prevent misclassifying healthy people as diseased.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive + False Positive}} \quad \text{(ii)}$$

The percentage of accurate positive predictions among all real positives is known as recall. The malaria prediction system calculates the percentage of real, uninfected samples that the model properly identifies.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive + False Negative}} \quad \text{(iii)}$$

Deployment: Deploying of the trained model would be done on a web application using stream lit.

2.4 Methods of Evaluation

2.4.1 Accuracy

The percentage of properly identified samples relative to all samples is known as accuracy.

2.4.2 Precision and recall

When a dataset is imbalanced—that is, when there are more samples of one class than the other precision and recall become relevant measures.

2.4.3 F1-score

The F1-score is calculated as the harmonic mean of recall and precision. Model's total performance is represented by a single value that is obtained by combining the two measures.

2.4.4 Confusion matrix

A confusion matrix is a succinct tabular summary used to organize predictions into categories such as true positives, true negatives, false positives, and false negatives. It effectively summarizes the performance of a classification model.

3.0 Results and Discussion

Malaria is a potentially lethal parasite illness caused by infection with Plasmodium protozoa, which is spread by a female Anopheles mosquito carrying the infection. The following model tries to distinguish cells infected with malaria from the uninfected ones. This research was developed using python as a programming language which is an open source with Panda, Seaborn, Scikit-image, Scikit-learn and open CV libraries to enable it to make image manipulation, analysis, visualization and recognition. The system consists of four primary stages: picture acquisition, image pre-processing, image segmentation, and classification. The training of the developed model was conducted using 15 epochs at 165 iterations per epoch while total iteration was 2475 iterations. The training period was calculated and estimated to be on a continuous form. The model yielded offline accuracy of 60.04% at final iteration while the average segmentation accuracy was 65% and average computational time of 2 hours per each epoch. Total parameters are 57,585,818, trainable parameters are 57,584,218, non-trainable parameters are 1,600. Figure 4 and 5 shows the training progress, training process and recognition process respectively.

Received: 21-01-2025 / Accepted: 05-02-2025 /Published: 31-03-2025

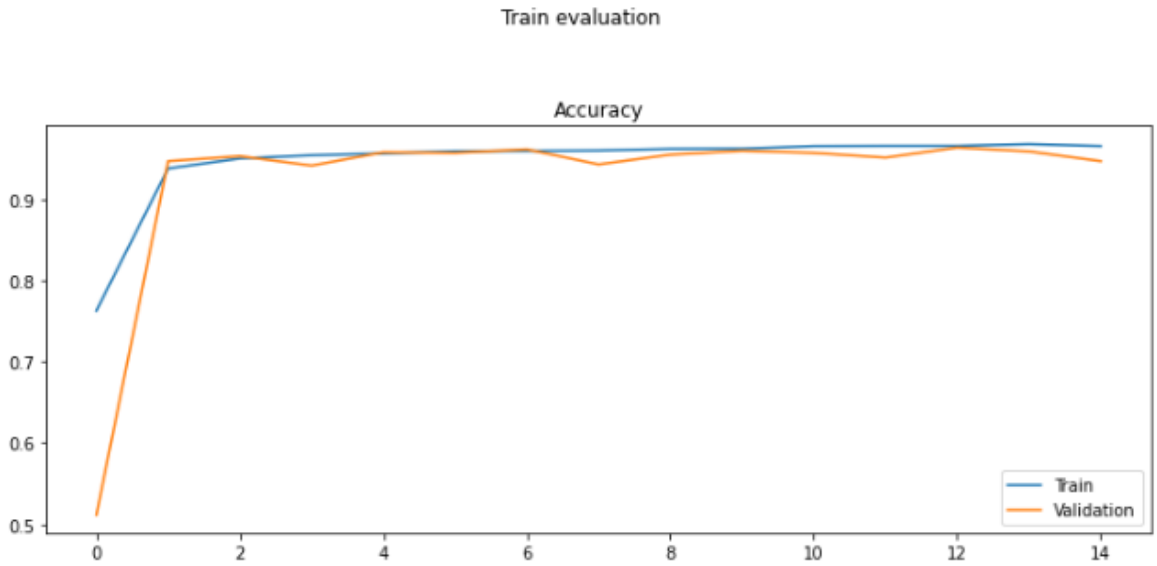

Figure 4: Comparing Validation and Training Accuracy

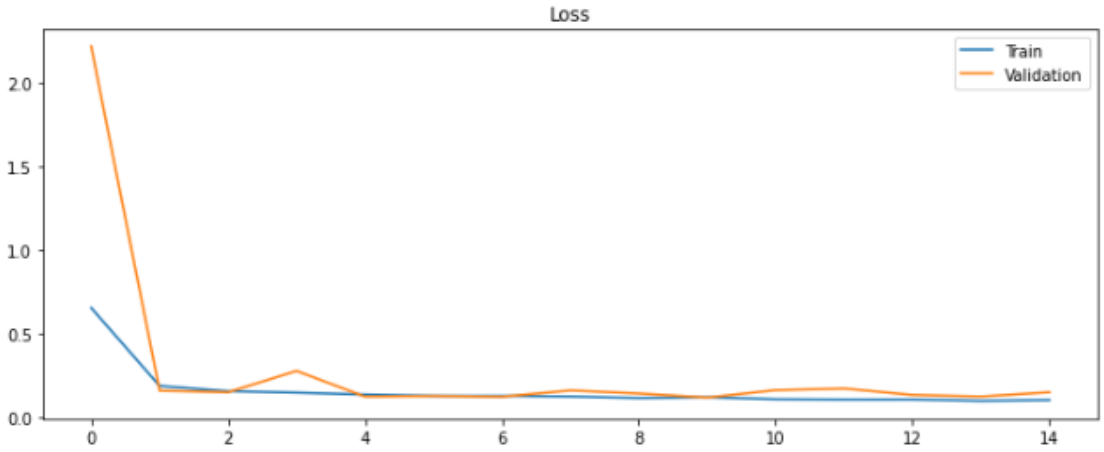

Figure 5: Comparing Training Loss and Validation Loss

This image is showing the training of each data in progress, its showing each loss and accuracy at every stage of training the data. The dataset has the parasitized cell images first followed by the images of the uninfected cells. In order to ensure better performance the dataset is shuffled randomly. Now, the data has been cleaned up and sorted into test, validation, and training sets. 80%, 10%, and 10% divided, accordingly. The CNN model approach is used to tackle this problem. The model uses the Adam optimizer with accuracy being the metric.

The layers used in this CNN model are convolutional layer, batch _normalization, dropout, flatten, dense at every stage of training the model.

3.1 Convolutional Layer

This particular neural network layer transforms a stream of data. A one-dimensional array is used to represent any type of data, including time series, text, and other types of data. Batch size can be any number, number of time steps is the length of the input sequence, number of output channels is the amount of filters used in the layer, number of features is the total number of weights and biases in the layer.

3.2 Batch Normalization

To accelerate and stabilize the training of artificial neural networks, batch normalization is employed. It involves normalizing the inputs of each layer through re-centering and re-scaling them.

3.3 Dropout

The Dropout layer is a technique that is used during training to prevent overfitting by forcing the network to learn to rely on more than just a few of its neurons. In this training the dropout rate is 0.

3.4 Flatten

The Flatten layer is generally used after a convolutional layer to reshape the output into a one-dimensional array layer into a format that can be fed into a fully connected layer.

3.5 Dense

The Dense layer is completely connected, this implies that each neuron in the layer is connected to every neuron in the preceding layer. The number of neurons in the dense layer can be specified using the unit option.

Table 1: CNN Model: "sequential"

Layer	Output Shape	Param #
conv2d	(None, 48, 48, 32)	896
max_pooling2d	(None, 24, 24, 32)	0
Batch normalization	(None, 24, 24, 32)	108
dropout	(None, 24, 24, 32)	0
conv2d_1	(None, 24, 24, 118)	36992
max_pooling2d_1	(None, 10, 10, 108)	0
batch_normalization_1	(None, 10, 10, 108)	506
dropout_1	(None, 10, 10, 108)	0
conv2d_2	(None, 10, 10, 108)	147584
max_pooling2d_	(None, 6, 6, 108)	0
batch_normalization_2	(None, 6, 6, 108)	502
conv2d_3	(None, 6, 6, 502)	590336
dropout_3	(None, 6, 6, 502)	0
conv2d_4	(None, 6, 6, 510)	0
Flatten	(None, 4608)	0
dense (Dense)	(None, 3999)	18436000
dense_1 (Dense)	(None, 3999)	16004000
dense_3 (Dense	(None, 999)	4001000
dense_4 (Dense)	(None, 2)	2002

Received: 21-01-2025 / Accepted: 05-02-2025 /Published: 31-03-2025

3.2 Machine Learning (Support Vector Machine) Model

Support Vector Machine (SVM) is a well-known supervised learning algorithm commonly utilized for classification problems in machine learning. It has proven effective in various applications, including image recognition and medical diagnostics

The goal of SVM in this project is creating a decision boundary that can differentiate and separate the non-infected from the infected. It was used to learn the relationship between the clinical symptoms, laboratory tests and the progression of the illness (mild, moderate or severe). To achieve this, we collected the data, extracted features from the dataset (blood cell counts and parasitic level), after training the SVM model on the dataset, the trained model was then employed to generate predictions for the progression of malaria for new patients.

3.3 Evaluation Results from Developed Malaria Progression Prediction System

In evaluating the developed system, two key metrics were used: segmentation accuracy and computational time. The system achieved an average segmentation accuracy, with variations observed due to the nature of the datasets for different patients. The overall accuracy achieved was 94.30%. The accuracy by the model was slightly higher at 96.53%, while the accuracy determined by validation was 94.67%. These metrics provide insights into the performance and effectiveness of the developed system.

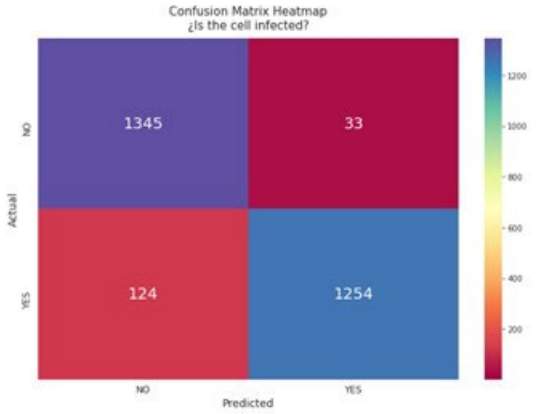

Figure 6: Confusion Matrix on ANN model

Table 2: Classification report on performance evaluation

	0	1	Accuracy	macro avg	weighted avg
Precision	0.69	0.65		0.67	0.67
Recall	0.71	0.64		0.67	0.67
f1-score	0.70	0.65	0.67	0.67	0.67
Support	368	321	689	689	689

Table 2 presents some metrics used in the analysis of malaria prediction which showcase the accuracy, recall, precision, F1-Score, TP, FP, FN, TN, as seen in this table, the accuracy is 0.67, recall is 0.706, precision is 0.64 and F1-Score is 0.65.

Received: 21-01-2025 / Accepted: 05-02-2025 /Published: 31-03-2025

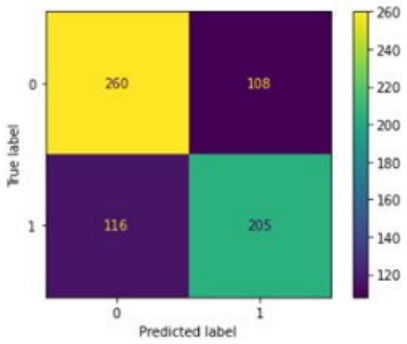

Figure 7: Confusion Matrix on ANN model

3.4 Comparing Support Vector Machines and Artificial Neural Networks

The dataset contained various clinical parameters of individuals diagnosed with malaria and the intention was to use these criteria to forecast how the disease would proceed.

The findings indicates that ANN and SVM can effectively develop a prediction model for malaria model but ANN with good accuracy surpassed SVM. Because ANNs can recognise patterns and non-linear correlations in datasets, they are good for medical diagnosis applications, where the interactions between various features can be complex using traditional methods.

SVM had a good performance with its classification ability to handle data efficiently but it cannot effectively capture the correlations between features with complex medical datasets.

Performance of both ANN and SVN could be improved with availability of larger and diverse datasets. SVM had an accuracy of 67.5% while ANN had an accuracy of 96%.

4.0 Conclusion

Malaria continues to be a major global health issue, making accurate prediction of disease progression essential for timely interventions and improved patient care. This study addresses this pressing problem by developing a Malaria Progression Prediction System (MPPS) using Artificial Neural Networks (ANN) and Support Vector Machines (SVM). By utilizing a diverse dataset and applying machine learning techniques, the proposed MPPS exhibits strong predictive capabilities. The comparative analysis of ANN and SVM highlights their strengths and limitations in forecasting malaria progression. This research offers valuable insights for healthcare professionals, aiding in informed decision-making, resource optimization, and improved disease management. Future research could focus on expanding the dataset, incorporating additional features, and investigating more machine learning models to enhance the accuracy and effectiveness of the MPPS.

Acknowledgements

I sincerely express my gratitude to Federal University, Oye-Ekiti, Nigeria, for providing me with an excellent platform to pursue my academic and research goals, I will also like to express our gratitude to the Department of Computer Engineering for its unwavering support, guidance, and resources throughout the development of this research project. The department's dedication to nurturing creative thinking and practical problem-solving has been invaluable in shaping my knowledge and skills. The encouragement and mentorship provided by the faculty members, staff, and my peers have contributed to the realization of this project.

References

Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., & Abbeel, O. J. (2018). Opportunities and obstacles for deep learning in biology and medicine. Journal of The Royal Society Interface, 15(141), 20170387.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115-118.

Centers for Disease Control and Prevention (CDC). (2021). Malaria diagnosis (United States). Retrieved from https://www.cdc.gov/malaria/diagnosis_treatment/diagnostic_tools.html

World Health Organization (WHO). (2022). Guidelines for the treatment of malaria (3rd ed.). Geneva: World Health Organization.

Centers for Disease Control and Prevention (CDC). (2023). Malaria biology. Retrieved from https://www.cdc.gov/malaria/about/biology/

World Health Organization (WHO). (2023). World malaria report 2023. Geneva: World Health Organization.
Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273-297.
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.
Nair, V. (2018). A review on support vector machine. International Journal of Innovative Research in Computer and Communication Engineering, 6(1), 802-805.
World Health Organization (WHO). (2022). World malaria report 2020. Geneva: World Health Organization.
World Health Organization (WHO). (2023). Malaria. Retrieved from https://www.who.int/news-room/fact-sheets/detail/malaria

Received: 21-01-2025 / Accepted: 05-02-2025 /Published: 31-03-2025