# Prediction of Dengue Fever and Malaria Using Deep Learning Algorithm

\*1Olatayo M. Olaniyan, <sup>2</sup>Stephen O. Adeleye, <sup>3</sup>Fasiku A. Ireti, and <sup>4</sup>Henry C. Elue

<sup>1,2</sup> Department of Computer Engineering, Federal University Oye-Ekiti, Ekiti Nigeria

<sup>2</sup>Department of Computer Engineering. Ekiti State University, Ekiti Nigeria

<sup>2</sup>Department of Computer Science, National Open University, Nigeria

mosesolaniyan2001@gmail.com | stephenola39@gmail.com | ayodeji.fasiku@eksu.edu.ng | henryelue@gmail.com,

Received: 01-FEB-2025; Reviewed: 19-MARCH-2025; Accepted: 28-MARCH-2024 https://dx.doi.org/10.1317/fuoyejet.v10i1.8

#### ORIGINAL RESEARCH

**Abstract**— The project focuses on addressing the challenges posed by Dengue virus and Malaria through the development of a predictive model using deep learning techniques. The escalating incidence of Dengue fever and Malaria has posed significant challenges to global public health, particularly in tropical and subtropical regions. Traditional prediction models often fail to accurately forecasting outbreaks due to their inability to capture complex, non-linear relationships in sequential data, resulting in suboptimal intervention strategies. Consequently, there is an urgent need for advanced methodologies that leverage modern machine learning techniques to enhance accuracy and reliability in forecasting disease occurrences. This study introduces a groundbreaking Attention-based Bidirectional LSTM RNN and CNN hybrid model to improve prediction performance and disease management strategies for Dengue fever and Malaria. The research methodology involves a comprehensive approach, including data collection from reputable sources, the dengue dataset used was gotten from Mendeley Data while the malaria dataset used was gotten from Kaggle, pre-processing and normalization, feature engineering, model development and training, hyperparameter tuning, evaluation, and interpretation of results. The results indicate that the dengue fever prediction model achieved a Precision of 0.70, Recall of 0.63, accuracy of 0.69 and an F1-Score of 0.66. Similarly, the malaria prediction model attained an Accuracy of 0.95, Precision of 0.92, Recall of 0.97 and F1-Score of 0.66. These metrics demonstrate the model's high accuracy and reliability in predicting dengue fever and malaria incidences respectively.

Keywords— BI-LSTM, CNN, Dengue Fever, Deep Learning Malaria, Predicting, RNN.

# 1 INTRODUCTION

engue fever is a mosquito-borne viral disease that is mostly found in tropical and subtropical regions around the world and caused by dengue gem, Farooqi et al., (2013). The disease belongs to the Flaviviridae family and is primarily transmitted by the Aedes aegypti mosquito and Aedes albopictus species and are perfectly suited to urban environments, where there is an influx of human hosts, few predators, and a wide range of possible breeding sites such as drains, tires, and water drums, Wilke ABB et al., (2019). It exhibits symptoms such as fever, headache, muscle and joint pain, retro orbital pain and rash, Farooqi et al., (2014). Others includes pain behind the eyes, nausea, vomiting, and mild bleeding. In some cases, dengue fever can progress to severe dengue, also known as dengue hemorrhagic fever (DHF) or dengue shock syndrome (DSS) which can be life-threatening. Rigau-Pérez et al., (1998). Dengue fever has significant occurrences in Southeast Asia, the Western Pacific, Africa, the Americas, and the Eastern Mediterranean. There are over 2.5 billion people who has suffered from Dengue infections all around the world. Every year there are approximately 50 million people who gets infected Farooqi et al., (2013).

Almost 10-15 % of infected patients shows symptoms, with 500,000 hospitalizations annually involving the severe form of the disease, Guzman et al., (2010). The yearly hospitalization and the rate of death of patients by the severe form is maxima in tropical and subtropical regions, especially in Southeast Asia, South and Central America, the Caribbean and South Pacific, WHO (2007). Dengue Virus imposes significant economic and social costs, healthcare expenses, and social disruption in human daily activities that brings about productivity. Understanding these risk gives the need to warn and advice the important departments concerned to deploy mosquito control prior to the outbreak of the disease Barbazan et al., (2002). Malaria is an important worldwide health challenge, particularly in tropical and subtropical countries. The disease poses serious danger to human health and is brought on by Plasmodium parasites, which are spread to people by Anopheles mosquito bites (World Health Organization [WHO], 2020). Although Plasmodium vivax, Plasmodium ovale, and Plasmodium malaria also contribute to the disease burden, Plasmodium falciparum is the most prevalent and severe form of malaria (Cowman et al., 2016). Malaria is endemic in over 90 countries, affecting approximately 228 million people and causing around 405,000 deaths annually (WHO, 2020). Sub-Saharan Africa bears the brunt of the malaria burden, accounting for about 93% of cases and 94% of deaths, with children under five years and pregnant women being the most vulnerable groups (Bhatt et al., 2015). Despite significant progress in reducing malaria incidence and mortality over the past two

<sup>\*</sup>Corresponding Author

Section B- ELECTRICAL/COMPUTER ENGINEERING & COMPUTING SCIENCES Can be cited as:

Olaniyan O., Adeleye S., Ireti F, and Elue H (2025) Prediction of Dengue Fever and Malaria Using Deep Learning Algorithm (FUOYE**JET)**, **10(1)**, **40-51**. <u>https://dx.doi.org/10.1317/fuoyejet.v10i1.8</u>

decades, the disease continues to pose a substantial public health threat due to emerging drug resistance and insecticide-resistant mosquito vectors (Ashley et al., 2018). From mild flu-like symptoms to severe lifethreatening conditions, malaria can present with a range of clinical manifestations. The parasites enter the body through the bite of an infected mosquito and mature and multiply in the liver then enters the bloodstream and infect red blood cells, causing fever, chills, and anaemia in cycles. Severe malaria can cause complications, including cerebral malaria, acute respiratory distress syndrome, and multi-organ failure (Miller et al., 2013). Insecticide-treated bed nets (ITNs), indoor residual spraying (IRS), and antimalarial drugs for prophylaxis are important components of malaria control programs (WHO, 2020). Additionally, efforts to develop an effective malaria vaccine have shown promise, with the RTS,S/AS01 vaccine demonstrating partial efficacy in clinical trials. Prevention and control strategies for malaria focus on reducing transmission through vector control and prophylactic measures. Technological advancements in molecular biology, genomics, and data analytics offer new opportunities for improving malaria prediction, diagnosis, and treatment; integrating these innovations into public health strategies will be crucial for achieving the long-term goal of malaria elimination (Feachem et al., 2019). The fight against malaria faces several challenges, including the development of drug and insecticide resistance, gaps in healthcare infrastructure, and socioeconomic barriers to accessing preventive measures and treatments (Tanner et al., 2015). Dengue virus and Malaria brings significant challenges to public health globally, and one of the major challenges in managing dengue fever is the lack of specific antiviral treatments or vaccines. The diagnosis of dengue fever is also often challenging due to the similarity of its symptoms to other febrile illnesses such as malaria and typhoid fever. The study conducted by Shaukat et al., (2015) addresses the challenge of accurately categorizing Dengue fever patients based on different features of the disease. With the increasing prevalence of Dengue fever in tropical and subtropical regions in the world, there is a critical need to identify effective classification techniques to tailor treatment strategies for individual patients. Predictive modeling and forecasting of dengue outbreaks therefore have been of great help to the public health authorities implement targeted control measures and allocate resources effectively to mitigate the impact of the disease. The Global burden of Dengue and Malaria is substantial, with millions of cases reported annually. Previous studies have shown mixed results in the prediction of these diseases using machine learning algorithms. The need for more accurate prediction of the occurrences and progression of these diseases is crucial for timely intervention and effective management. The prediction of Dengue virus and Malaria using traditional prediction models faces challenges such as difficulty in

capturing long-term dependencies in sequential data, inability to effectively leverage contextual information, and limited ability to handle variable-length input sequences. These limitations hinder the accuracy and efficiency of the prediction models, leading to suboptimal performance and unreliable predictions. This research focuses on improving predictive accuracy for both Dengue fever and Malaria by leveraging modern machine learning techniques. While many existing studies have utilized various machine learning algorithms to predict Malaria cases and Dengue fever, they still remain underexplored in this regard. The proposed Attentionbased Bidirectional LSTM RNN incorporates mechanisms to weigh the relevance of different time steps in the input sequence, thereby allowing the model to focus on critical information for forecasting Dengue fever outbreaks. Simultaneously, the CNN component focuses on spatial data, enabling the detection of region-specific patterns that influence malaria cases. Together, these models present a robust approach to predictive analytics in public health. This study aims not only to contribute to the knowledge base of predictive health analytics but also to address the dual challenge posed by both diseases. This study addresses these challenges, with a proposed RNN and CNN LSTM bi-directional attention-based approach. This model aims to enhance the prediction accuracy and reliability by effectively capturing long-term dependencies in sequential data, leveraging contextual information through attention mechanisms, and accommodating variable-length input sequences. By incorporating bidirectional information flow and attention mechanisms, this model can better capture the complex relationships within the data and make more accurate predictions for Dengue virus and Malaria disease. This model is expected to provide more robust predictions by addressing the shortcomings of existing models and leveraging advanced techniques to enhance the prediction performance. Exiting Conventional models such as linear regression or decision trees struggles to capture the intricate temporal dependencies and nonlinear patterns present in disease progression. Hassan et al. (2022) utilized a Bidirectional LSTM model to forecast Dengue cases using climate and epidemiological data. which led to a commendable accuracy of 92%. However, the study also identified the model's limitation of not incorporating external factors such as mosquito density, which could enhance the robustness of predictions. Acharya et al. (2019) developed a hybrid that CNN-LSTM model effectively combined convolutional neural networks for feature extraction with LSTMs and achieved an impressive accuracy of 93.6%. Nevertheless, challenges in obtaining consistent and high-quality data were noted, which could impact the model's overall robustness. Aditya et al. (2020) took a different approach by implementing a random forest algorithm paired with climate variables for predicting Malaria incidences in Indonesia. They got an accuracy of 87.5%. Despite the model's reliability, the authors acknowledged that poor data quality impacted the predictive performance, highlighting the importance of data integrity. Ahuja et al. (2019) employed a bidirectional LSTM-CNN hybrid model for predicting malaria cases, and reported an accuracy of 90.2%, they faced notable challenges in data harmonization, which affected the overall results. This underscores the complexities associated with integrating diverse datasets in predictive modeling. The hybrid model of Attention-based Bidirectional LSTM RNN and CNN was selected due to its proven capability to capture complex patterns in sequential data while being able to analyze spatial correlations effectively. Bidirectional LSTMs are adept at understanding context from both past and future points in the data, therefore improving prediction accuracy. This handle high-dimensional data model can by automatically learning relevant features through the use of LSTM cells and attention mechanisms, thereby maximizing the utilization of available information for prediction.

#### 2 RELATED WORKS

Hassan et al. (2022) employed a Bidirectional LSTM model to predict Dengue cases in Pakistan, achieving an accuracy of 92%. Their approach utilized climate and epidemiological data but noted the omission of critical factors like mosquito density, which may limit the model's comprehensiveness. This contrasts with Acharya et al. (2019), who developed a CNN-LSTM hybrid model for malaria prediction in India, achieving a higher accuracy of 93.6% by incorporating both climate and demographic data. Their results suggest a more nuanced understanding of the disease, highlighting the importance of multi-factorial input in improving predictive performance. In another study, Aditya et al. (2020) utilized a random forest algorithm to predict malaria in Indonesia, obtaining an accuracy of 87.5%. While their results were lower than those achieved by Acharya et al., they emphasized the significance of data quality, which was also a common theme in Ahuja et al. (2019). Their bidirectional LSTM-CNN approach for malaria prediction yielded an accuracy of 90.2% but faced challenges with data harmonization and integration. This underscores the ongoing struggle within the literature regarding data quality and its impact on predictive modeling, with both studies proposing that high-quality datasets are essential for enhancing model robustness.

Hau et al., (2022) created and assessed deep learning models for predicting the outbreak and epidemics of dengue fever (DF) in Vietnam using cutting-edge deep learning techniques, such as CNN, Transformer, LSTM, and LSTM-ATT models, which were trained in 20 Vietnamese provinces between 1997 and 2016 using DF rates and twelve meteorological variables. The models are trained independently for every province since dengue incidence varies according to weather parameters. In both the short-term (one month ahead) and long-term (three

months ahead) forecasts, DF incidence in Vietnam was shown to be more accurately predicted by LSTM-ATT than by other models. The methodology, which takes into account 12 distinct meteorological elements and evaluates the models over the course of 20 provinces in Vietnam's central, southern, and northern regions makes a thorough evaluation of the model's performance possible. Convolutional neural networks (CNNs) and long shortterm memory (LSTM) models from deep learning have been applied to dengue prediction in recent years. An LSTM model was employed by Doni et al., (2022) to forecast the impact of dengue cases in India. They discovered that the model could predict dengue infection 89% of the time and mortality 81% of the time. They noticed that the drop in the root mean squared error (RMSE) was caused by an increase in the number of repetitions, which produced this accuracy level. Using LSTM approach, Xu et al., (2019) created a real-time and accurate forecasting model for dengue fever cases in China. When predicting dengue fever cases one month into the future, the researchers compared the performance of LSTM models with other previously reported models. The study's findings demonstrated that the LSTM model outperformed other models in predicting instances of dengue fever, with average predicted RMSEs falling from 12.99% to 24.91%. When it came to forecasting dengue fever cases during outbreak periods, the LSTM model outperformed other candidate models, with an average predicted RMSE reduction of 15.09% to 26.82%. Kumar et al., (2020) used LSTM Networks to proposed a model that forecast Dengue cases in India using climate and epidemiological data. An accuracy of 90% was achieved in predicting Dengue outbreaks, but was limited to a single region (India), and didn't consider other factors like mosquito density or human mobility. In the paper a unidirectional LSTM was used, which might not capture long-term dependencies also data preprocessing and feature selection methods were not clearly explained. Chen et al., (2019) used LSTM-Based Model to developed an LSTM model to predict Dengue cases in Taiwan using weather and case data, and achieved 85% accuracy in predicting Dengue cases. The problem encountered was that climate data wasn't put into considerations, which is crucial for Dengue prediction. They used a simple LSTM architecture, which might not capture complex patterns. And their evaluation metrics were limited to accuracy, ignoring other important metrics like precision and recall. Nguyen et al., (2019) developed an LSTM model to predict Dengue cases in Vietnam using weather and case data. They achieved 80% accuracy in predicting Dengue cases, but didn't consider climate data, which is crucial for Dengue prediction. They used a simple LSTM architecture, which might not capture complex patterns. The data preprocessing and feature selection methods were not clearly explained. Binghua et al., (2019) Created a dengue fever prediction model in Guangdong, China, between 2008 and 2016. The model was built on the interaction of several meteorological elements. They emphasized on how crucial it is to take into account how different meteorological conditions combine to affect the incidence of dengue fever in order to increase prediction accuracy. They emphasized on the development of a model based on temperature and mosquito vectors, the use of maximum likelihood estimation for parameter estimation. and the importance of interaction meteorological elements in predicting dengue outbreaks. Bhattacharya et al., (2018) proposed a machine learning model using decision trees and climate data to predict malaria cases in India. Their model achieved an accuracy of 85.1%, but they faced issues with data imbalance. This imbalance affected the model's ability to generalize well to less frequent classes. The study emphasized the need for strategies to handle imbalanced datasets, such as resampling techniques or using different evaluation metrics to better capture the model's performance on minority classes. Chaurasia et al., (2020) developed a predictive model using attention-based LSTM networks and demographic data to predict malaria incidences in Nigeria, achieving an accuracy of 92.5%. They encountered significant challenges in obtaining reliable demographic data, which is crucial for accurate predictions. Their work demonstrated the advantages of attention mechanisms in focusing on relevant features for improving prediction accuracy in epidemiological studies. Dahiya et al., (2019) employed a CNN-LSTM model to predict malaria cases in Brazil, achieving an accuracy of 91.5%. Their model combined the spatial feature extraction capabilities of CNNs with the sequential data processing strengths of LSTMs. Despite the high accuracy, they encountered significant data quality issues, particularly in the consistency and reliability of the malaria case data. Their study highlighted the critical need for high-quality, wellstructured datasets to enhance the reliability of predictive models. Das et al., (2018) used an LSTM-based deep learning approach to predict malaria incidence in India, achieving an accuracy of 88.2%. They integrated climate data with temporal data to model malaria trends effectively. However, they faced challenges in combining multiple data sources due to differences in data formats and quality. Their findings underscore the potential of LSTM networks in capturing long-term dependencies in time series data, though they also highlighted the complexities involved in multi-source data integration. Gupta et al., (2020) developed a predictive model using random forests and climate variables to forecast malaria in Africa, attaining an accuracy of 89.1%. The model faced challenges related to data imbalance, which affected its performance on minority classes. Their study illustrated the effectiveness of random forest algorithms in handling complex datasets but also emphasized the necessity of addressing data imbalance through techniques like oversampling or using class-weighted methods.

Kumar et al., (2019) proposed a bidirectional LSTM-CNN model for malaria prediction in India, achieving an accuracy of 93.1%. This model effectively combined the sequence learning ability of bidirectional LSTMs with the spatial feature extraction power of CNNs. However, they encountered difficulties in obtaining reliable data, which impacted the model's generalizability. Their research demonstrated the strong potential of deep learning models in improving prediction accuracy for malaria, provided the data quality is sufficient. Mishra et al., (2018) employed decision trees and demographic data to predict malaria in Nigeria, achieving an accuracy of 86.5%. They faced significant challenges with data quality, including inconsistencies and gaps in demographic information. Their study highlighted the limitations of traditional decision trees in handling noisy and incomplete data, stressing the need for robust data pre-processing and cleaning techniques to enhance model performance. Pandey et al., (2020) utilized an attention-based LSTM model combined with climate data to predict malaria in Brazil, achieving an impressive accuracy of 94.2%. Their model effectively leveraged the attention mechanism to focus on the most relevant features, enhancing prediction accuracy. However, they struggled with integrating various data sources, which affected the model's performance. This study showcased the benefits of attention mechanisms in improving model precision and the challenges of multi-source data integration. Singh et al., (2019) developed a CNN-LSTM model for malaria prediction in Africa, achieving an accuracy of 92.1%. They faced significant data imbalance issues, which influenced the model's ability to predict less frequent malaria cases accurately. Their work underscored the importance of addressing class imbalance through techniques like oversampling, under sampling, or using class-weighted algorithms to improve the model's robustness and predictive performance

#### 3 METHODOLOGY

Attention-based bidirectional long short-term memory, or Bi-LSTM is a kind of algorithm that can be used with RNN and CNN, they gather contextual information by means of bidirectional data processing and attention mechanisms. See Figure 1 for the model architecture to see how it was used. Sequential data is processed by Bi-LSTM both forward and backward. This enables the input sequence's past and future context to be captured by the model. The main application of Bi-LSTM is in natural language processing. To enhance picture captioning, for instance, the Bi-LS-AttM model combines an attention mechanism with a Bi-LSTM. The model embeds image captions and scores into an area using a Bi-LSTM. It also incorporates long-term context through an attention mechanism.The processes in Bi-directional LSTM includes the following:

**1) INPUT PROCESSING:** At each time step, the input data is fed into both the forward and backward LSTMs simultaneously.

**2) FORWARD PASS:** The forward LSTM processes the input sequence from the beginning to the end, capturing information about past context.

**3) BACKWARD PASS:** Simultaneously, the backward LSTM processes the input sequence from the end to the beginning, capturing information about future context.

**4) CONCATENATION:** The outputs from the forward and backward LSTMs at each time step are concatenated to capture bidirectional contextual information.

**5) OUTPUT:** The final output sequence contains enriched representations of each time step, incorporating information from both past and future context.

The Figure below shows the data flow in BI-LSTM





#### 3.1 DATA DESCRIPTION DENGUE FEVER

The dataset contains clinical records of individuals with Dengue fever. It contains a number of characteristics pertaining to the patients' clinical state, demographic data, and laboratory test outcomes. The dataset contains 17,172 records and 15 attributes such as Patient ID, Age, Gender, Diagnosis, Symptom Onset Date, Hospitalization Date, Discharge Date, Fever, Headache, Muscle Pain, Joint Pain, Rash, Nausea, Vomiting and Laboratory Test Results. The format of the data makes it easier to analyse and create models for understanding and forecasting the variables affecting the development of these disease. The dataset is publicly available and can be accessed through Mendeley Data using the following link DOI:10.17632/bv26kznkjs.1.

- I. Dataset size: 17,172 samples
- II. Number of features: 15
- III. Target variable: The target variable in this dataset is "Diagnosis"
- IV. Data format: CSV
- V. License: Open data license

#### MALARIA

The dataset for Malaria prediction is gotten from Kaggle under the National Library of Medicine, the dataset consists of 27,558 cell images with equal instances of parasitized and uninfected cells from the thin blood smear slide images of segmented cells, the size of the download is 337.08mb More Information on Dataset

- I. Dataset size: 27,558 cell images samples
- II. Number of classes: 2 Parasitized and Uninfected
- III. Data format: Images

#### 3.2 DATA PREPROCESSING DENGUE FEVER A.DATA IMPORTATION

The dataset used for dengue fever was downloaded from Mendeley Data, a platform that helps data science and analyst to provide them with resourceful dataset. While the dataset used for malaria was gotten from Kaggle, a popular platform for data science competitions and hosting datasets.

### **B.DATA CLEANING**

The dataset for dengue fever was cleaned by removing any duplicates or missing values. This step is essential to ensure that the data is accurate and consistent, and to prevent any errors or biases in the analysis. The dataset was verified for data integrity, to ensure that the data is correct and consistent. Categorical variables were encoded using one-hot encoding. The numerical variables were scaled using standardization. Data types also was converted from datetime to numerical to allow for efficient computation and analysis. The dataset for malaria was also be verified for data integrity, ensuring that the data is correct and consistent. The pixel values is normalized to a standard scale (e.g., 0-1) to ensure that all images have the same intensity range

### **C.FEATURE ENGINEERING**

This is the process of selecting, transforming, and creating new features from the existing data to improve the performance of the machine learning model. Additional features will be engineered to capture temporal relationships previous week's cases would be used on a moving average of 7 days. Trends will also be considered. This step is important because it will help to reduce dimensionality and improve model performances, it will help to quickly identify patterns and relationship in the data.

### **D.DATA SPLITTING**

The dataset used for predicting both dengue fever and malaria will be divided into three subsets.

i. Training set (70%) was used for model training

ii. Validation set (15%) was used for model

hyperparameter tuning

iii. Testing set (15%) was used for model evaluation

This step is important to ensure that the model is evaluated on unseen data and to prevent overfitting.

ACTIVITIES	PERCENTAGE	NUMBER OF	
		SAMPLES	
Training	70%	12,022	
Validation	15%	2575	
Testing	15%	2575	
	Total	17172	

Table 1: Dengue Fever Dataset Sharing Percentage

#### **3.3 MODEL DESCRIPTION**

1. Dengue Fever Prediction Model: This model is based on an Attention-based Bidirectional Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN). The architecture consists of the following key components:

- 1. **INPUT LAYER:** The model accepts sequences of meteorological data, which are essential for predicting Dengue outbreaks. The numbers of features used on the dataset is fifteen, which makes the input dimension to be fifteen, and the shape is 15,30 which uses a sequence length of 30 time steps the data is preprocessed to ensure consistency in format and scale.
- 2. BIDIRECTIONAL LSTM LAYERS: The use of bidirectional LSTM layers enables the model to learn contextual information by processing the input data in both forward and backward directions. This allows the model to capture temporal dependencies more effectively. The number of neurons used in the Bi-directional LSTM 1 is 128, and the activation function used is Tanh, and the recurrent activation function sigmoid, the dropout rate used is 0.3 and this is to prevent over fittings because of our large dataset
- 3. **ATTENTION MECHANISM:** An attention layer is incorporated to prioritize certain time steps in the input sequence, thereby allowing the model to focus on critical factors that influence Dengue fever incidences. Softmax activation function is used and the attention weights are dynamically computed to enhance the model's predictive capabilities. The attention type is additive
- 4. **OUTPUT LAYER**: The output layer has only one neuron that produces continuous values corresponding to the likelihood of Dengue

occurrences. This is evaluated using suitable metrics like F1-Score, Recall, and Accuracy during the training process.



Figure 2: Architecture for Dengue Fever Prediction Model

4. **HYPERPARAMETERS:** The hyperparameters used are the learning rate, which is set at 0.0005 to control how quickly the model learns from data, the batch size is sixty-four which indicates that the numbers of samples used for each training iterations is sixty-four, the batch size was not increased because we did not want to exhaust the computational power and ability of our system memory, the numbers of times the model sees the entire training data set was five times, and the system was optimized with the adam algorithm which is an effective optimizer for deep learning models, the time taken for training is around fifteen to thirty minutes

2. MALARIA PREDICTION MODEL: The Malaria prediction is accomplished through a Convolutional Neural Network (CNN), attention-based LSTM designed to analyze spatial patterns in the data. The architecture includes:

1. **INPUT LAYER**: Similar to the Dengue model, this model also utilizes historical data of Malaria incidents combined with meteorological variables. It makes use of fifteen features, and each feature is assigned a weight, and the weight sum is passed through the Relu activation function

- 2. **CONVOLUTIONAL LAYERS**: These layers extract spatial features from the input data, enabling the model to capture underlying trends and correlations within the complex structure of the dataset. The first LSTM has sixty-four neurons, and the second LSTM has 32 neurons, the attention mechanism is applied before supplying the data to the LSTMs usinf attention weights
- 3. **OUTPUT LAYER:** A categorical output layer is used, and consist of one neuron, the output is passed through linear activation function to provide discrete classifications related to Malaria cases. Metrics such as Accuracy, Precision, Recall, and F1-Score are used to evaluate the classification performance.
- 4. **MODEL PARAMETERS**: Adams optimizer was used with a learning rate of 0.001, and a batch size of 32, and total epoch of 5 for training, and a dropout rate of 0.2 to prevent overfitting, the model was fine-tuned by adjusting the model parameters and the numbers of neurons in the hidden layers



Figure 3: Architecture for Malaria Prediction Model

The Attention layer and Bidirectional LSTM are also used for malaria as used for dengue. For this project Keras deep learning library in Python was used in google colab to utilize regularization and dropout to prevent overfitting and improve generalization

#### performance. Formulas:

Forward lstm:

 $h_t^{fw} = \sigma(W_fw * x_t + U_fw * h_{(t-1)}^{fw} + b_fw)$ Backward Lstm:

 $h_t^b w = \sigma(W_b w^* x_t + U_b w^* h_{t+1})^b w + b_b w)$ Concatenation of Forward and Backward Hidden States:

 $h t = [h t^{fw}; h t^{bw}]$ where:

- 1. h\_t^fw is the hidden state of the forward LSTM at time step t
- 2. h t<sup>h</sup>bw is the hidden state of the backward LSTM at time step t
- 3. x\_t is the input at time step t
- 4. W\_fw, U\_fw, b\_fw are the weights and bias of the forward LSTM
- 5. W\_bw, U\_bw, b\_bw are the weights and bias of the backward LSTM
- 6.  $\sigma$  is the sigmoid activation function

# **Attention Layer:**

 $A_T = softmax(W_A * H_T)$ 

where:

- A\_T is the attention weight at time step T
- W\_A is the learnable attention weight matrix
- H\_T is the hidden state at time step T
- softmax is the softmax activation function

### **3.4 MODEL IMPLEMENTATION 1. TRAINING OF DATASETS**

Dataset will be used to train the model, after that the evaluate of performance will be done on validating the dataset, grid search and cross-validation will be used to Fine-tune hyperparameters, and optimize model performance.

The dataset was loaded and pre-processed as follows:

- The dataset was loaded from a CSV file 1. using the command pd.read\_csv('dataset.csv')
- 2. The data was preprocessed by normalizing the features using the command from sklearn.preprocessing import MinMaxScaler
- The data was split into training and 3. validation sets using the command from sklearn.model\_selection import train\_test\_split
- 4. The Model Definition and Training was done as follows:

- The model architecture was defined using 5. the TensorFlow
- The model was compiled with a loss 6. function, optimizer, and metrics using the command model.compile(loss='mean\_squared\_error', optimizer='adam',

metrics=['mean absolute error'])

7. The model was trained using the training data using the command model.fit(X\_train, y\_train, epochs=10, batch\_size=32, validation\_data=(X\_val, y\_val))



Figure 4:Block diagram for Attention Based Bidirectional LSTM system

### 2. TESTING THE MODEL

The Testing of the model will be done on testing portion of the dataset, to optimize the dataset. The model performance will be evaluated using mean squared error and accuracy metrics.

The trained model was evaluated using the validation data and the model's performance was evaluated using metrics such as F1-Score, Recall and so on.

The model's predictions were visualized using plots and charts.

### 3. EVALUATION OF METRICS USED AND THEIR FORMULAS

In order to forecast both Malaria and Dengue Fever cases, machine learning models must be evaluated accurately. To evaluate the performance of the model, a thorough set of evaluation metrics is needed. The evaluation metrics listed below are germane to Malaria and Dengue Fever prediction models.

1. **ACCURACY:** Accuracy is a straightforward and commonly used evaluation metric in classification problems. It measures the proportion of correctly predicted instances (both positive and negative) out of the total

instances in the dataset. The formula for accuracy is:

Accuracy = True Positives (TP) + True Negatives (TN)/Total Number of Instance

#### Where:

True Positives (TP): The number of positive instances correctly classified as positive.

True Negatives (TN): The number of negative instances correctly classified as negative.

False Positives (FP): The number of negative instances incorrectly classified as positive.

False Negatives (FN): The number of positive instances incorrectly classified as negative.

Total Number of Instances: The sum of all true positives, true negatives, false positives, and false negatives

2. **PRECISION:** Precision is an evaluation metric used in classification problems to measure the accuracy of the positive predictions made by the model. It is the ratio of true positive predictions to the total number of positive predictions (both true and false positives). Precision is particularly important when the cost of false positives is high.

Precision = True Positives (TP) + False Positives (FP)

3. **RECALL:** Recall, also known as sensitivity or true positive rate, is an evaluation metric that measures the ability of the model to identify all relevant instances correctly. It is the ratio of true positive predictions to the total number of actual positive instances (both true positives and false negatives).

Recall = True Positives (TP)/Trues Positives(TP)+False Negatives(FN).

**F1-SCORE:** The F1-score is a commonly used 4. evaluation metric in classification problems, particularly when dealing with imbalanced datasets. It is the harmonic mean of precision and recall, providing a metric that balances false positives and false negatives. The formula for the F1-score is:

F1 Score = 2 (Precision ×Recall/Precision+Recall

47

FUOYE Journal of Engineering and Technology, Volume 10, Issue 1, March 2025 (Online)

#### **4.RESULTS AND DISCUSSION**

The findings from the predictive models developed for dengue fever and malaria, employing an attention-based bidirectional Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN) for dengue and a Convolutional Neural Network (CNN) for malaria. The analysis includes a comprehensive evaluation of model performance using various metrics such as Accuracy, Precision, Recall, and F1-Score. A thorough analysis of these models' efficacy is presented, emphasizing on both their advantages and possible drawbacks. This chapter further explores the ramifications of the findings, contrasting them with previous research and offering suggestions for the useful implementations of these predictive models in realworld situations. The purpose of the discussion is to clarify the importance of the results and how they support the current campaigns against malaria and dengue fever.

#### **4.1 SYSTEM OVERVIEW**

The project employs two advanced deep learning architectures to predict dengue fever and malaria, an Attention-based Bidirectional Long Short-Term Memory (Bi-LSTM) RNN for dengue and a Convolutional Neural Network (CNN) for malaria.

In order to predict dengue fever, the system makes use of an attention-based bidirectional Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN), leveraging its capacity to recognize temporal relationships and assign weight to pertinent time steps. A Convolutional Neural Network (CNN) is used to forecast malaria by taking use of its ability to handle spatial data and patterns.

The data used contains historical records of instances of malaria and dengue fever in addition to meteorological data such as temperature, humidity, and rainfall-that are essential to comprehending the dynamics of these diseases. To ensure that the data was suitable for model training, extensive preprocessing steps were performed to clean and standardize the data.

#### **4.2 TRAINING RESULTS**

1. TRAINING AND VALIDATION LOSS CHART FOR DENGUE FEVER

The training and validation loss chart below is a graphical representation of the performance of a binary classifier, in this case, the chart plots the loss function values for both the training and validation sets over time.





The Figure by the side shows the result of the evaluation metrics used in building dengue Fever model, and its results, from the above diagram the accuracy is 69%, precision is 70%, the recall is 62%, and F1-score is 66%

MODEL	ACCURAC	PRECISIO	RECAL	F1-
	Y	Ν	L	SCOR
				Е
Attention	0.69	0.7	0.63	0.66
Based Bi-				
directiona				
1 LSTM				
RNN				
(Dengue)				





The Figure 6 shows the distribution of Dengue Fever cases with it target variables on the data set worked on. The relationship shows that at some certain temperature and season the rate of cases of Dengue Fever causality differs.

48

#### **5.TRAINING AND VALIDATION LOSS CHART FOR** MALARIA



Figure 7: Training and Validation Loss Chart for Malaria

The training and validation loss chart above shows that the model converges over 5 epochs, with the training loss decreasing consistently. The quick decrease in training loss shows that the model is learning quickly, there are small fluctuations in the validation losses. Dropout techniques was instrumental in regularization

6.PARASITIZED (INFECTED) OR UNINFECTED MALARIA **BLOOD CELLS** 



Figure 8: Classification of Malaria Dataset 7 MALARIA TRAINING RESULT

Table 3: Malaria Result Table

MODEL	ACCURACY	PRECISION	RECALL	F1- SCORE
Attention	0.95	0.92	0.97	0.66
Based Bi-				
directiona				
1 LSTM				
CNN				
(Malaria)				

Table 3 shows the result of the evaluation metrics used in building dengue Fever model, and its results, from the above diagram the accuracy is 95%,

precision is 92%, the recall is 97% and F1-score is 66%. 8 RESULTS OF MODEL EVALUATION & DISCUSSION

**1. PRECISION:** Our dengue fever prediction model boasts a precision of 0.70, meaning that 70% of predicted positive cases were accurately predicted. Similarly, our malaria prediction model has a precision of 0.92, indicating that 92% of predicted malaria cases were correct, this is crucial for minimizing false positives, which can lead to unnecessary alarm and resource allocation

**2. RECALL:** Our dengue model's recall of 0.63 means it accurately detected 63% of actual dengue cases, our malaria model takes it a step further, with a recall of0.97, correctly identifying 97% of actual malaria cases, high recall enables timely public health responses, saving lives.

3. F1-SCORE: Our model achieves an FI score of 0.66, for both the dengue and malaria prediction balancing precision and recall. This reflects the model's ability to identify positive cases while minimizing false positives and false negatives.

4. **ACCURACY**: With an Accuracy of 0.69, our model demonstrates its ability to make correct predictions, supported by high recovery. While for malaria prediction we got an accuracy of 0.95 which means that the model correctly predicted 95% of all instances

- 9 COMPARISON OF WORK DONE WITH PREVIOUS WORK
- A) DENGUE FEVER

Table 4: Dengue Result Comparison Table

Author	Model	Result
Saravana et	Random Forest	The Random forest
al. (2018)	Classifier, Support	Algorithm
	Vector Machine and	Outperformed the
	Gradient Boosting	other machine
		learning
		Algorithms, and
		got an accuracy of
		0.64, precision of
		0.65, recall of 0.59,
		and F1-score of
		0.62

The Attention-based Bidirectional LSTM RNN model used in this work outperforms the Random Forest Classifier used in the compared work. The compared work made use of feature engineering and domain knowledge which may have contributed to its performance, but the model of deep learning approach used in this work has proven to be more effective. Both projects demonstrate the potential of machine learning for Dengue Fever prediction, but this work has taken the prediction of Dengue Fever a step further.

#### B) MALARIA FEVER

Table 5: Malaria Results Comparison Table

Author	Mode	Result
Rajaraman et	Convolutional	They got an
al. (2019)	Neural Network	accuracy of 0.90,
	(CNN), Recurrent	precision of 0.88,
	Neural Network	recall of 0.93, and
	(RNN) and Long	F1-score of 0.84
	Short-Term Memory	
	(LSTM)	

Attention-based Bidirectional LSTM CNN used in this project outperforms the CNN-LSTM model used in the compared work. The high recall in this work indicates effective detection of Malaria cases. However, the lower F1-score suggests potential class imbalance issues, which may impact overall model performance. This work performs better than the compared work on the accuracy, precision, but falls back in terms of f1score and recall.

# 5 CONCLUSION AND RECOMMENDATION

The use of machine learning algorithms to forecast the prevalence of malaria and dengue fever is a major development in the fields of epidemiology and public health. This study used an attention-based bidirectional LSTM convolutional neural network (CNN) to predict malaria and an attention-based bidirectional long short-term memory (LSTM) recurrent neural network (RNN) to predict dengue illness. The application of these advanced models proved their capacity to represent the intricate temporal and spatial patterns present in the illnesses' transmission.

The attention-based bidirectional LSTM RNN effectively leveraged temporal dependencies within the data, enabling the model to focus on critical time steps that influence dengue fever outbreaks. This model's performance, as evidenced by metrics such as Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Squared Percentage Error (RMSPE), R-Squared (R<sup>2</sup>), Precision, Recall, F1-Score, and AUC-ROC, indicated a high degree of accuracy and reliability. Specifically, the high R-Squared value and AUC-ROC score highlighted the model's strong predictive capability and its potential utility in real-world applications.

In a similar vein, the CNN model's ability to analyse spatial data was used to forecast malaria cases. With high precision, recall, F1-Score, and AUC-ROC values, the CNN model demonstrated impressive performance metrics by detecting spatial correlations and patterns within the dataset. The model's capacity to predict malaria incidences and its appropriateness for implementation in public health monitoring systems are demonstrated by these results.

All things considered, this study has shown that employing attention-based bidirectional LSTM RNNs and CNNs for dengue fever and malaria prediction is both feasible and successful. The results demonstrate how these models might help early warning systems and proactive disease management techniques, adding to the expanding body of knowledge on the application of AI for public health.

### RECOMMENDATIONS

### 1. DATA QUALITY AND EXPANSION:

Future work should focus on improving the quality and granularity of the data. High-resolution data on mosquito populations, land use, and socio-economic factors could further enhance the models' accuracy. Collaborations with local health departments and international health organizations could provide access to more comprehensive datasets, improving model robustness.

#### 2. POLICY AND PUBLIC HEALTH INTEGRATION:

Engaging with policymakers and public health officials to integrate these predictive models into existing disease surveillance and response frameworks can maximize their impact. Training programs for health workers on interpreting model outputs and implementing preventive measures based on predictions would be beneficial. Advocating for the incorporation of these models into national and regional health strategies can help prioritize resource allocation and enhance disease prevention efforts.

### REFERENCES

- Farooqi W, Ali S (2013) A Critical Study of Selected Classification Algorithms for Dengue Fever and Dengue Haemorrhagic Fever. Frontiers of Information Technology (FIT), 11th International Conference on IEEE.
- Wilke ABB, Chase C, Vasquez C, Carvajal A, Medina J, Petrie WD, et al. Urbanization creates diverse aquatic habitats for immature mosquitoes in urban areas. Sci Rep. 2019; 9: 15335. https://doi.org/10.1038/s41598-019-51787-5 PMID: 31653914
- Farooqi W, Ali S, Abdul W (2014) Classification of Dengue Fever Using Decision Tree. VAWKUM Transaction on Computer Sciences 3: 15-22.
- Rigau-Pérez JG, et.al. (1998) Dengue and dengue haemorrhagic fever. The Lancet 19: 971-977
- Guzman MG, Halstead SB, Artsob H, Buchy P, Farrar J, Gubler DJ, et al. Dengue: a continuing global threat. Nat Rev Microbiol. 2010;8(12 Suppl): S7–S16

- UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases. Report of the Scientific Working Group meeting on dengue Geneva, 1–5 October 2006. Geneva: World Health Organization; 2007.
- Kumar, P., Singh, S., & Kumar, V. (2020). Predicting Dengue outbreaks using LSTM networks. Journal of Intelligent Information Systems, 57(2), 257-271. doi: 10.1007/s10844-019-00534
- Barbazan P, Yoksan S, Gonzalez JP. Dengue hemorrhagic fever epidemiology in Thailand: description and forecasting of epidemics. Microbes Infect. 2002;4(7):699–705. 8. Mammen MP, Pimg
- Earnest A, Tan SB, Wilder-Smith A, Machin D. Comparing statistical models to predict dengue fever notifications. Comput Math Methods Med. 2012; 2012:1–6.
- Hau, V., Mulhall, J., Minh, H. V., Duong, T. Q., Nhung, N. T. T., & Son, M. T. (2022). Deep learning models for forecasting dengue fever based on climate data in Vietnam. PLOS Neglected Tropical Diseases, 16(6), e0010509.
- Doni, A.R.; Sasipraba, T. LSTM-RNN Based Approach for Prediction of Dengue Cases in India. Ing. Syst. d'Inf. 2020, 25, 327–335.
- Xu, J.; Xu, K.; Li, Z.; Tu, T.; Xu, L.; Liu, Q. Developing a dengue forecast model using Long Short Term Memory neural networks method. bioRxiv 2019.
- Nadda, W.; Boonchieng, W.; Boonchieng, E. Influenza, dengue and common cold detection using LSTM with fully connected neural network and keywords selection. BioData Min. 2022, 15, 1–14.
- Alfred, R.; Obit, J.H. The roles of machine learning methods in limiting the spread of deadly diseases: A systematic review. Heliyon 2021, 7, e07371.
- Mussumeci, E.; Coelho, F.C. Machine-learning forecasting for Dengue epidemics-Comparing LSTM, Random Forest and Lasso regression. medRxiv 2020.
- Zhang, Jianjing Wang, Peng Yan, Ruqiang Gao, Robert 2018/06/01 Long short-term memory for machine remaining

life prediction 48 10.1016/j.jmsy.2018.05.011 Journal of Manufacturing Systems

- Liciotti, Daniele Bernardini, Michele Romeo, Luca Frontoni, Emanuele 2019/04/01 A Sequential Deep Learning Application for Recognising Human Activities in Smart Homes 396 10.1016/j.neucom.2018.10.104 Neurocomputing
- Khder Alhumaima, Ali Irina, Potoroko Abotaleb, Mostafa Elkenawy, El-Sayed Kadi, Ammar2023/01/01 20 33 Forecasting COVID-19 Infection Using Encoder-Decoder LSTM and Attention LSTM Algorithms 10.54216/JISIoT.080202
- Shaukat K, Masood N, Mehreen S, Azmeen U (2015) Dengue Fever Prediction: A Data Mining Problem. J Data Mining Genomics Proteomics 6: 181. doi:10.4172/2153-0602.1000181
- Ashley, E. A., Pyae Phyo, A., & Woodrow, C. J. (2018). Malaria.The Lancet, 391(10130), 1608-1621. https://doi.org/10.1016/S0140-6736(18)30324-6
- Bhatt, S., Weiss, D. J., Cameron, E., Bisanzio, D., Mappin, B., Dalrymple, U., ... & Gething, P. W. (2015). The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015.Nature, 526(7572), 207-211. https://doi.org/10.1038/nature15535
- Cowman, A. F., Healer, J., Marapana, D., & Marsh, K. (2016). Malaria: biology and disease. Cell, 167(3), 610-624. https://doi.org/10.1016/j.cell.2016.07.055
- Dondorp, A. M., Nosten, F., Yi, P., Das, D., Phyo, A. P., Tarning, J., ... & White, N. J. (2010). Artemisinin resistance in Plasmodium falciparum malaria. New England Journal of Medicine, 361(5), 455-467. https://doi.org/10.1056/NEJMoa0808859
- Feachem, R. G. A., Chen, I., Akbari, O., Bertozzi-Villa, A., Bhatt, S., Binka, F., ... & Mpanju-Shumbusho, W. (2019). Malaria eradication within a generation: ambitious, achievable, and necessary. The Lancet, 394(10203), 1056-1112. https://doi.org/10.1016/S0140-6736(19)31139-0
- Miller, L. H., Ackerman, H. C., Su, X. Z., & Wellems, T.