Infectious Risk Prediction and Analysis System

Project Report #3

CS-351: Artificial Intelligence

Fall 2025

Instructor: Ahmed Nawaz

Submitted By:

Zain Jamshaid	2023772
Hamid Quyyum	2023418
Hiba Zulfiqar	2023246

Faculty of Computer Science and Engineering

GIK Institute of Engineering Sciences and Technology

Date: 23 November 2025

1 Objective

The primary objective of this phase was to transition from traditional machine learning algorithms to advanced concepts to further enhance the predictive accuracy of the Infectious Risk Prediction System. While the previous phase established a baseline model using Random Forest, this phase focuses on implementing an Artificial Neural Network (ANN) to better capture complex, non-linear relationships between climatic variables and disease outbreaks.

2 Recap of Previous Phase (Random Forest)

In the previous report (Project Report #2), we implemented a Random Forest Regressor. With feature engineering (lag features, rolling windows, and cyclical temporal encoding), we achieved significant improvements over the raw baseline.

Previous Best Results (Random Forest with Feature Engineering):

- Malaria Cases: R^2 Score of 0.94
- Dengue Cases: R^2 Score of 0.83

While the model performed exceptionally well for Malaria, there was still room for improvement in predicting Dengue cases.

3 Methodology: Shift to Neural Networks

In this report, we shifted our approach to advanced Machine Learning by implementing a Feedforward Neural Network (ANN). ANNs can approximate complex functions and interactions that ensemble tree methods might miss.

3.1 Data Preparation Changes

An additional preprocessing step was introduced:

- Feature Scaling: We applied StandardScaler to both the input features X and the target variables Y to ensure stable training.

3.2 Model Architecture

We designed a deep Artificial Neural Network using the TensorFlow/Keras framework with the following architecture:

- Input Layer: Matches the number of features.
- Hidden Layers: Four dense layers with decreasing neuron counts 256, 128, 64 and 32 respectively.
- Regularization: Applied Batch Normalization for stability and Dropout layers (rates 0.2 - 0.3) to prevent overfitting.
- Activation Function: ReLU for hidden layers.
- Output Layer: 2 for multi output (Malaria & Dengue).

3.3 Training Strategy

- Optimizer: Adam with an initial learning rate of 0.001.
- Loss Function: Mean Squared Error (MSE).
- Callbacks: Implemented EarlyStopping (to prevent overfitting), ReduceLROnPlateau, and ModelCheckpoint.

4 Experimental Evaluation (Epoch Variations)

To determine the optimal training duration, we trained model with three different Epoch configurations: 20, 50, and 100 Epochs.

4.1 Comparative Analysis of Epochs

The table below summarizes the performance of the ANN under different training durations on the Test Set (2022-2023 data).

Experiment Configuration	Actual Epochs Trained	Train Loss (MSE)	Val Loss (MSE)	R2 Malaria	R2 Dengue
20 Epochs	20	0.0981	0.7460	0.9417	0.9504
50 Epochs	26 (Stopped Early)	0.0881	0.9662	0.9427	0.9290
100 Epochs	29 (Stopped Early)	0.0865	0.9337	0.9417	0.9360

Observation: The model converged relatively quickly. Both the 50 and 100 epoch runs triggered EarlyStopping around 26-29 epochs. However, the 20 Epoch model provided the most generalized performance, achieving the highest score for Dengue without overfitting to the training data.

5 Final Results & Comparison

The shift to Neural Networks yielded a substantial improvement particularly for Dengue cases, which was the weaker metric in the previous phase.

Target Variable	Previous Best (R2) (Random Forest)	New Best (R2) (ANN - 20 Epochs)	Improvement
Malaria Cases	0.94	0.94	Stable (Maintained High Accuracy)
Dengue Cases	0.83	0.95	Significant Increase (+0.12)

6 Visual Analysis

The scatter plots below visually confirm the improvement achieved by the Artificial Neural Network over the Random Forest model. Each plot displays Malaria predictions on the left and Dengue predictions on the right.

- Malaria (Left): The points tightly cluster around the red "Perfect Prediction" line, consistent with the previous model's high performance.
- Dengue (Right): The ANN significantly tightened the prediction cloud compared to the previous Random Forest model. The variance has been reduced, pushing the R^2 score from 0.83 to 0.95.

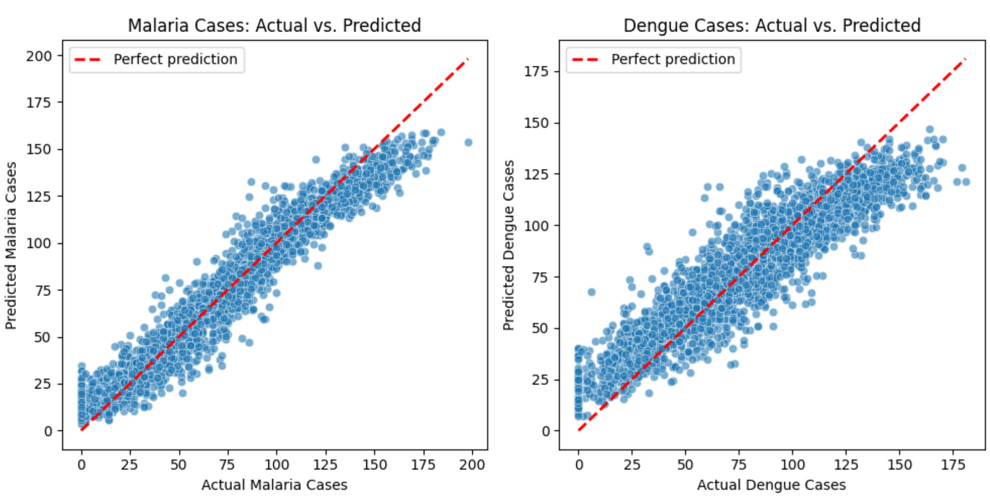

Figure 1: Random Forest model performance

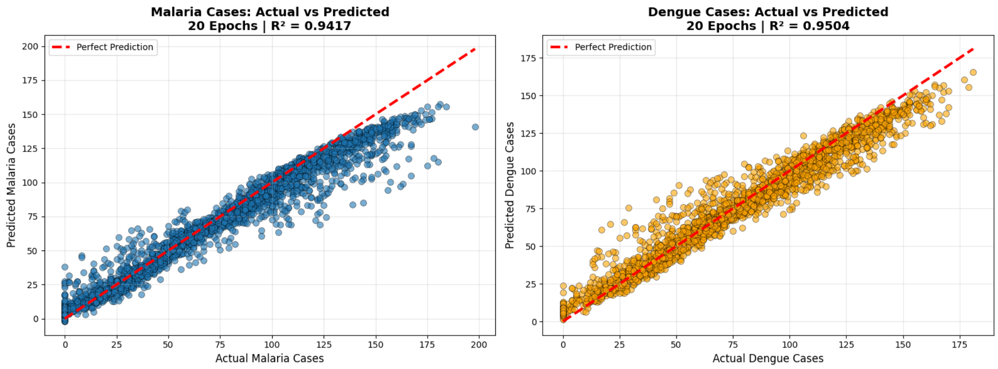

Figure 2: Improved neural network results