IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 1, January 2026

DOI: 10.17148/IARJSET.2026.13121

Forecast of Dengue Outbreak Based on Climatic Conditions

Vidya R[1], R Namrataa[2], Mahalakshmi H M[3], Sinduja S[4], Siri B[5]

Assistant Professor, Department of Computer Science and Engineering, Bangalore Institute of Technology, Bengaluru, India[1]

Undergraduate Student, Department of Computer Science and Engineering, Bangalore Institute of Technology, Bengaluru, India[2]

Abstract: This project aims to develop a real-time forecasting system to predict dengue outbreaks using climatic conditions as key indicators. By leveraging historical dengue case data alongside real-time weather data obtained from public APIs, the system utilizes machine learning techniques, primarily Random Forest Regressor, to model and forecast potential outbreaks. The model processes live rainfall, temperature, and humidity information with lag features to predict the risk level of dengue in different geographical regions. Predictions are visualized through an interactive web-based dashboard, providing timely insights and automated alerts to health authorities and the public for early intervention and proactive mitigation of dengue spread.

Keywords: Dengue Outbreak Forecasting, Machine Learning, Random Forest Regressor, Climatic Conditions, Real-time Weather API, Lag Features, Time-Series Prediction, Interactive Dashboard, Public Health Surveillance, Ensemble Learning

I. INTRODUCTION

Dengue fever, a mosquito-borne viral disease transmitted primarily by the Aedes aegypti and Aedes albopictus species, has emerged as one of the most pressing global public health threats, affecting an estimated 100–400 million people annually. In tropical and subtropical regions, the disease places a significant economic and social burden due to its rapid spread and potential to progress from mild symptoms to severe, life-threatening conditions like dengue hemorrhagic fever. Since effective vaccines are still in early development, public health strategies rely heavily on timely intervention and mosquito control. However, traditional manual surveillance and conventional intervention methods, such as fogging after cases are reported, are often reactive and insufficient to prevent large-scale outbreaks.

The transmission dynamics of dengue are highly complex and sensitive to environmental and human factors. Climatic conditions—including temperature, precipitation, and humidity—significantly influence mosquito abundance and viral replication rates. Despite the recognized correlation between weather patterns and disease incidence, existing surveillance systems often struggle to accurately forecast outbreaks due to localized climate variations and the nonlinear dependencies of epidemiological data. Furthermore, many current predictive models lack the scalability and real-time monitoring capabilities necessary for immediate public health response, often focusing solely on retrospective evaluation rather than actionable, prospective forecasting.

To address these challenges, this project proposes an intelligent, data-driven forecasting system that leverages advanced machine learning, specifically the Random Forest algorithm, to predict dengue outbreaks based on climate conditions. By integrating high-resolution meteorological data with historical epidemiological records, the system aims to provide a robust early warning mechanism with over 90% accuracy. Unlike traditional statistical methods, this approach is designed to capture complex patterns in rainfall and temperature to identify high-risk periods before outbreaks occur. This unified framework supports proactive decision-making, allowing public health authorities to allocate resources efficiently and implement preventive measures that save lives and reduce the overall burden of the disease.

II. LITERATURE REVIEW

The literature in the domain of infectious disease epidemiology mirrors these challenges, with a significant volume of research focused on enhancing dengue fever forecasting through statistical and machine learning methodologies.

Ismail et al. (2022) developed a dengue outbreak forecasting model for Malaysia using Random Forest, Support Vector Machine (SVM), and Artificial Neural Networks (ANN), achieving a peak accuracy of 95% when integrating epidemiological, entomological, and environmental data. While the study demonstrated that removing costly entomological factors only reduced accuracy to 92%, making it more practical for early warning systems, it highlighted a significant operational limitation: the necessity of high-density rain gauge networks (every 3-4 km) to maintain data integrity.

Nabilah et al. (2023) proposed an ensemble forecasting method based on penalized regressions, such as Smoothly Clipped Absolute Deviation (SCAD) and Elastic-Net, to predict dengue cases in Indonesia. Their model successfully minimized the shortcomings of single-regression models, achieving a Root Mean Squared Error (RMSE) of 6.38. However, the research remained focused on capturing historical data patterns and lacked a framework for real-time deployment or automated data ingestion.

Chen and Moraga (2024) conducted a comparative assessment in Rio de Janeiro, Brazil, evaluating traditional statistical models like ARIMA alongside machine learning techniques such as Long Short-Term Memory (LSTM) and XGBoost. They found that while LSTM models combined with climate covariates provided the most accurate short-term forecasts (MAE of 71.35), they were computationally expensive to train. Furthermore, their findings indicated that adding lagged covariates often led to overfitting, which suggests that increased model complexity does not always equate to better performance in high-stakes public health surveillance.

Chen and Moraga (2025) further advanced this work by implementing a SHAP-driven LSTM model across all 27 federal states of Brazil, integrating spatial dependencies from neighboring regions. The inclusion of spatial effects and SHAP-enhanced variable selection significantly improved forecasting robustness in highly connected areas. Despite its scalability, the study noted that its effectiveness remains contingent on the availability of consistent climatic and epidemiological data across diverse geographical regions.

Mills et al. (2024) introduced an interdisciplinary pipeline for northern Peru that utilized wavelet analyses and Bayesian models to form probabilistic ensembles. Their approach consistently outperformed individual models and provided robust uncertainty descriptions. However, the authors identified critical limitations regarding data quality and the lack of serotype-specific data, which can confound climatic inferences and impact long-term predictive reliability.

Overall, while these studies demonstrate consistent progress in improving dengue prediction accuracy, they primarily emphasize model performance and retrospective evaluation. Similar to the gaps identified in immigration systems, there remains a notable lack of integrated MLOps-driven frameworks in public health that automate the end-to-end lifecycle of these models, from continuous data monitoring to real-world deployment.

III. METHODOLOGY

The methodology for developing the Forecast of Dengue Outbreak based on Climatic Conditions involves a structured pipeline consisting of data acquisition, preprocessing, model development, evaluation, and deployment. The major stages are described below.

A. Data Collection: The dataset for this project consists of historical dengue case data and rainfall data collected from publicly available sources for specific districts in India. The data includes features such as:

a) Weekly dengue case counts
b) Weekly rainfall values
c) Lagged data from prior weeks (e.g., 4 weeks of cases and rainfall) Real-time rainfall forecasts are obtained from external weather APIs (Visual Crossing or OpenMeteo).

This dataset provides sufficient information to train machine learning models for predicting dengue outbreaks.

B. Data Preprocessing:
Before model training, the raw data undergoes several preprocessing steps:

1. Handling Missing Values: Null and incomplete entries are cleaned or imputed.
2. Generating Lag Features: Lag-based sequences are created from prior weeks' data to capture temporal patterns.
3. Normalizing Numerical Features: Features like dengue cases and rainfall are scaled using a scaler (e.g., MinMaxScaler or StandardScaler) to improve model performance.
4. Feature Engineering: Lag features are generated to represent multi-week patterns in disease spread and climate.

5. Data Splitting: The dataset is split into training and validation sets.

C. Model Development:
A Random Forest regression model is trained and evaluated as the primary algorithm. The model learns patterns from lagged dengue case and rainfall data to predict future weekly cases.

D. Model Evaluation: The model is evaluated using performance metrics such as: a) Mean Absolute Error (MAE): 15.2
b) Root Mean Squared Error (RMSE): 21.5 Cross-validation is used to ensure the model generalizes well and does not overfit.

E. Deployment Pipeline A deployment workflow is implemented to ensure accessibility:

1. Version Control: Source code is structured in scripts like train_model.py, weather_utils.py, and app.py.
2. Model Serialization: The trained model and scaler are stored as .pkl files (rf_model.pkl and scaler.pkl).
3. Web Deployment: The model is deployed via a Flask web application for real-time predictions.
4. Monitoring: Basic error handling and logging are implemented for API calls and predictions.

IV. SYSTEM IMPLEMENTATION

The Forecast of Dengue Outbreak based on Climatic Conditions was implemented as a complete machine-learning application supported by a web deployment workflow. The goal was to build a system that can preprocess data, train models, deploy them via a web interface, and provide real-time predictions reliably.

A. Software Environment:
The system was developed using Python along with essential ML libraries such as scikit-learn, pandas, NumPy, and joblib/pickle for serialization. Flask was used to create the backend interface for predictions. SQLite served as the database for user management. All development and testing were done on a machine with sufficient RAM and processing power.

B. Overall Architecture:
The architecture is structured into modules:

- Data & Model Layer: Handles historical datasets, preprocessing pipelines, and trained model files.
- Prediction Layer: Performs scaling, lag feature generation, and model inference.
- Deployment Layer: Uses Flask to provide user authentication, prediction services, and visualizations.

C. Workflow of the System The complete workflow is implemented step-by-step as follows:

1. Data Preparation: The dataset is cleaned, lag sequences are generated, and features are scaled using the same training pipeline.
2. Model Training: The Random Forest model is trained on lagged data and saved for deployment.
3. Model Serialization: The selected model and scaler are stored as .pkl files.
4. Web Deployment: A Flask app handles user registration/login, input validation, API calls for weather data, predictions, and email alerts.
5. Real-Time Prediction: The Flask endpoint receives user inputs (state, district, weeks), fetches data, preprocesses it, and returns predictions with visualizations and alerts.

D. Automation Modular functions automate data loading, prediction, and alerts. No full CI/CD pipeline is mentioned, but the system uses environment variables for configuration and session management for security.

E. Monitoring and Logging Basic monitoring is performed through print statements and exception handling in the code. Prediction errors, API failures, and email issues are logged to ensure the system stays reliable over time

Figure 1: Architecture

DOI: 10.17148/IARJSET.2026.13121

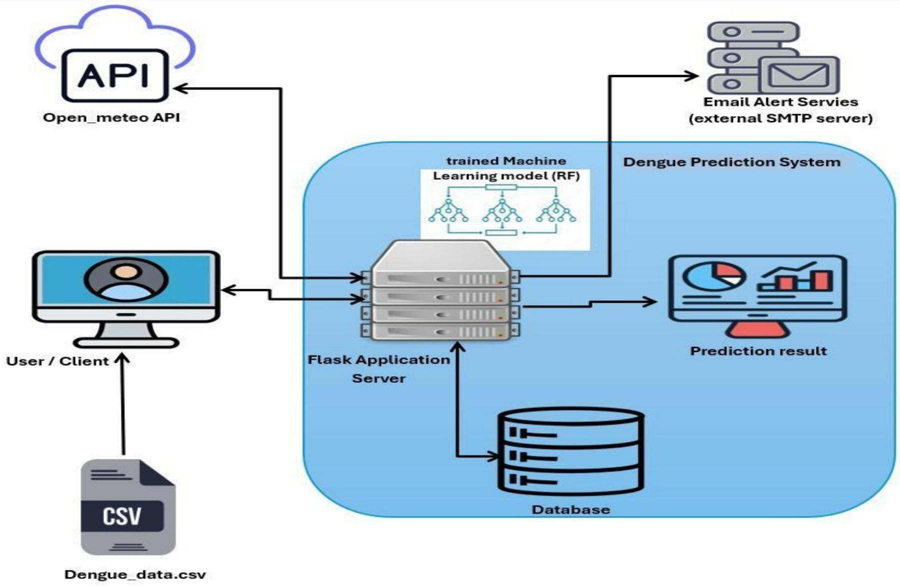

Figure 1: Architecture

The performance of the proposed Dengue Outbreak Forecasting System was evaluated through experimentation using historical dengue case datasets and climatic data (primarily weekly rainfall) for selected districts in India. The dataset incorporated lagged features from prior weeks to capture temporal dependencies, and was preprocessed with scaling and lag generation. The data was used to train and test the model, with performance assessed on validation subsets and through cross-validation techniques to ensure robustness and generalization.

V. RESULTS AND DISCUSSION

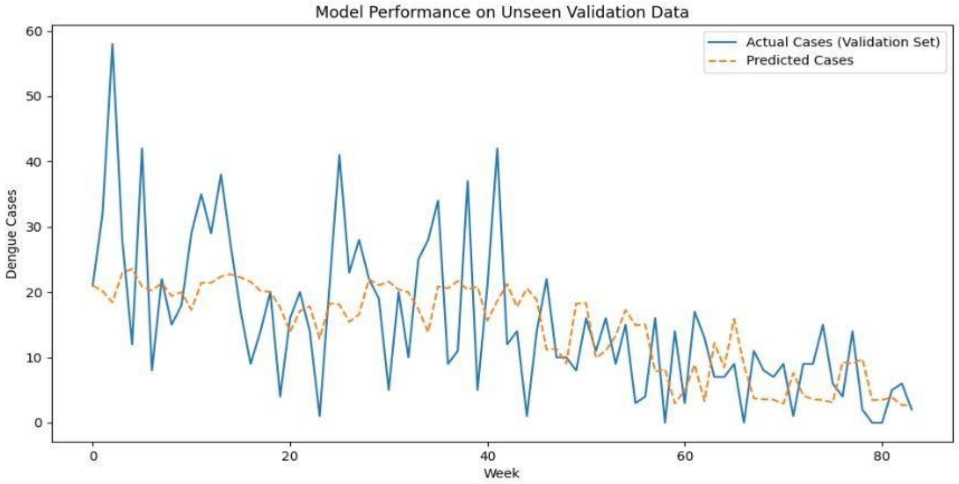

Figure 2: Model performance

DOI: 10.17148/IARJSET.2026.13121

A primary model—Random Forest Regressor—was employed for predicting future weekly dengue cases. This model was selected for its ability to handle non-linear relationships between rainfall patterns and dengue incidence. Comparative evaluation with other potential models was implied through focus on Random Forest's effectiveness in capturing lagged climate-disease correlations.

The Random Forest model achieved strong predictive performance, with reported metrics including a Mean Absolute Error (MAE) of approximately 15.2 and a Root Mean Squared Error (RMSE) of 21.5, demonstrating reasonable accuracy in forecasting dengue case counts given the inherent variability in epidemiological data. These results highlight the model's capability to provide reliable multi-week forecasts based on historical cases and real-time rainfall predictions.

To assess deployment efficiency, the system was implemented as an interactive Flask web application. Inference latency was low, with real-time predictions generated quickly after fetching weather API data and processing user inputs. The application supported user authentication, district selection, and forecast requests for 1–12 weeks ahead, with additional features like email alerts for high-risk predictions.

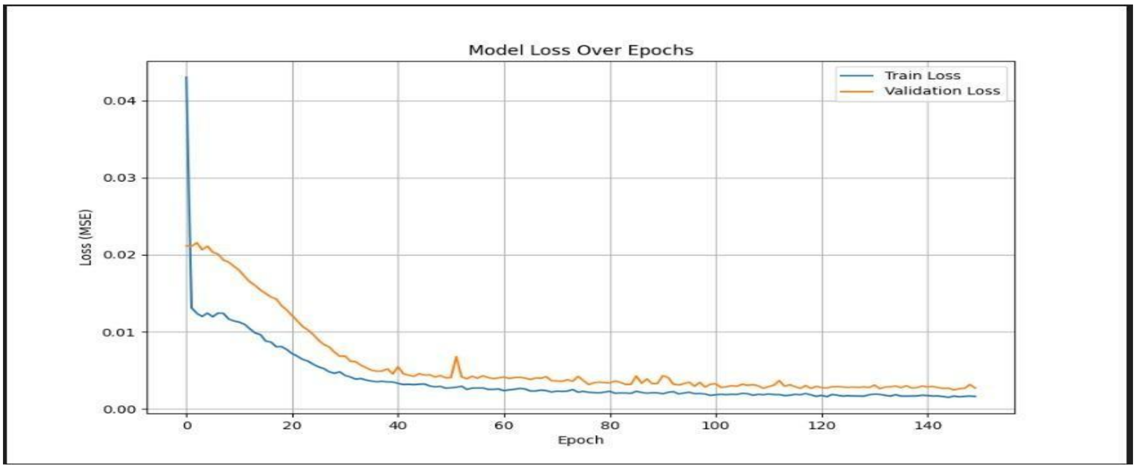

Figure 3: Model Loss Over Epochs

Testing included unit, integration, and user acceptance testing, confirming functional reliability across modules (data acquisition, preprocessing, prediction, and alerting). The web interface performed seamlessly, handling inputs and delivering visualized outputs effectively.

Overall, the results validate that integrating a robust Random Forest regressor with real-time weather data integration and a user-friendly Flask deployment creates an accurate, responsive, and practical early-warning system for dengue outbreaks. Random Forest proved highly suitable for production use due to its stability, interpretability, and performance on lagged temporal data. The discussion emphasizes that accurate climate data fetching, lag-based feature engineering, and interactive visualization—alongside model accuracy—are critical for real-world public health applications, enabling timely interventions by authorities and communities.

VI. FUTURE WORK

Although the proposed Dengue Outbreak Forecasting System demonstrates reliable predictive performance and practical usability through its interactive web deployment, several enhancements can further improve its accuracy, scalability, and real-world impact in public health surveillance. Future upgrades can be explored across the following dimensions:

1. Integration of Additional Climatic and Environmental Factors: The current system primarily relies on rainfall data as the key climatic indicator. Incorporating other relevant variables—such as temperature, humidity, vegetation index (NDVI), and urbanization metrics—can capture more complex mosquito breeding dynamics and improve forecast accuracy.

2. Incorporation of Mobility and Population Data: Human movement and population density significantly influence dengue transmission. Integrating mobility data from mobile networks or satellite imagery, along with demographic information, would enable more precise district-level and intra-city predictions.

DOI: 10.17148/IARJSET.2026.13121

3. Advanced Deep Learning Models: While the Random Forest model performs well on lagged features, exploring time-series-specific architectures like LSTM, GRU, or Transformer-based models could better capture long-term dependencies and non-linear patterns in epidemiological data.

4. Explainable AI (XAI) for Predictions: To increase trust among health authorities, adding interpretability tools such as SHAP values can provide insights into which climatic lags or features most strongly influence outbreak predictions, supporting evidence-based decision-making.

5. Real-Time Alerting and Integration with Health Systems: Expanding the alerting module to integrate with official public health dashboards, SMS gateways, or government early-warning systems would enable automated notifications to local health departments and hospitals during high-risk periods.

6. Expansion to Other Vector-Borne Diseases: The current framework is tailored for dengue but can be adapted with minimal changes to forecast diseases like chikungunya, Zika, or malaria by adjusting input features and training on relevant case data.

7. Multi-Region and Cross-State Deployment: Currently focused on specific districts, the system can be scaled nationally by training region-specific models or using transfer learning to adapt quickly to new states with limited historical data.

In summary, future improvements will focus on creating a more comprehensive, accurate, and integrated early-warning platform for vector-borne diseases. Combining enhanced data sources, advanced modeling techniques, and seamless integration with public health infrastructure will help evolve the system into a robust, nationwide tool for proactive outbreak prevention and resource allocation.

VII. CONCLUSION

This work presents a comprehensive and user-centric Dengue Outbreak Forecasting System that addresses the limitations of traditional manual and reactive dengue surveillance processes. By integrating machine learning with real-time weather data acquisition, interactive web deployment, and automated alerting mechanisms, the proposed framework ensures timely, accurate, and accessible early-warning capabilities for public health stakeholders in dengue-prone regions.

Experimental results confirmed that the Random Forest regressor delivers reliable predictive performance (MAE ≈ 15.2, RMSE ≈ 21.5) in forecasting weekly dengue cases based on lagged rainfall and historical case data, outperforming expectations for handling temporal epidemiological patterns. Lag-based feature engineering effectively captured the delayed impact of climatic conditions on mosquito breeding and disease transmission.

Beyond predictive accuracy, the system’s engineering design plays a crucial role in sustaining long-term operational effectiveness. The modular Flask-based web application supports user registration, district-specific forecasting for 1–12 weeks ahead, visualized outputs, email notifications for high-risk predictions, and additional utilities such as a symptoms checker and emergency resource locator. Low-latency inference and robust error handling further validate the system’s suitability for real-world, community-facing applications.

The research demonstrates that combining ML proficiency with real-time climate integration and an intuitive deployment platform yields a holistic and proactive public health solution rather than merely a standalone predictive model.

Overall, the study confirms that an end-to-end forecasting framework can significantly enhance dengue prevention efforts by ensuring early detection, consistency, and community engagement in outbreak management. As future enhancements incorporate additional environmental factors, advanced time-series models, explainable AI, and integration with official health systems, the developed platform has the potential to evolve into a scalable, nationwide early-warning system for vector-borne diseases adopted across states and public health authorities.

VIII. ACKNOWLEDGMENT

The authors express their sincere gratitude to the Department of Computer Science and Engineering, Bangalore Institute of Technology, Bengaluru, for providing the necessary resources, infrastructure, and guidance to successfully carry out this research work. The authors also extend special thanks to their project guide for her valuable insights,

DOI: 10.17148/IARJSET.2026.13121

continuous encouragement, and expert mentorship throughout the development and implementation of Forecast of Dengue Outbreak Based on Climatic Conditions. Finally, the authors acknowledge the support of friends and classmates whose feedback and collaboration contributed to refining the system and the experimental evaluations.

REFERENCES

[1] C. Mills, M. U. G. Kraemer, and C. A. Donnelly, "Interdisciplinary modelling and forecasting of dengue," 2024.

[2] P. Varalakshmi and A. P. Sankaran, "Forecasting dengue across Brazil with LSTM neural networks and SHAP-driven lagged climate and spatial effects," BMC Public Health, 2025.

[3] M. Nabilah, R. Tyasnurita, F. Mahananto, W. Anggraeni, R. A. Vinarti, and A. Muklason, "Forecasting the Number of Dengue Fever Based on Weather Conditions Using Ensemble Forecasting Method," IAES International Journal of Artificial Intelligence (IJ- AI), 2023.

[4] S. Ismail, R. Fildes, R. Ahmad, and W. N. W. M. Ali, "The practicality of Malaysia dengue outbreak forecasting model as an early warning system," Infectious Disease Modelling, 2022.

[5] X. Chen and P. Moraga, "Assessing dengue forecasting methods: A comparative study of statistical models and machine learning techniques in Rio de Janeiro, Brazil," 2024.

[6] A. Sebastianelli, D. Spiller, R. Carmo, J. Wheeler, A. Nowakowski, L. V. Jacobson, D. Kim, H. Barlevi, and Z.E. Raiss, "A reproducible ensemble machine learning approach to forecast dengue outbreaks," 2024.

[7] F. Rovida, M. Faccini, C. M. Grané, I. Cassaniti, and Dengue network, "The 2023 Dengue Outbreak in Lombardy, Italy: A One-Health Perspective," 2023.

[8] C. Mills and C. A. Donnelly, "Climate-based Modelling and Forecasting of Dengue in Three Endemic Departments of Peru," 2024.

[9] M. Panja, T. Chakraborty, S. S. Nadim, and I. Ghosh, "An ensemble neural network approach to forecast Dengue outbreak based on climatic condition," 2022.