International Journal of Future Engineering Innovations

Comparative Modeling of Dengue Incidence in Nepal Using Climate-Based Machine Learning Techniques

Pralad Kadel [1*], **Dr. Suman Thapaliya** [2], **Rohit Prasad Pandey** [3], **Dipak Adhakari** [4]

[1] Department of Information Technology, Texas College of Management and IT, Siphal, Chabahil, Kathmandu, Nepal
[2] Department of Information Technology, Lincoln University College, Petaling Jaya, Selangor, Malaysia
[3, 4] Texas College of Management and IT, Lincoln University College, Malaysia

* Corresponding Author: **Pralad Kadel**

Article Info

ISSN (online): 3049-1215
Impact Factor (RSIF): 8.25
Volume: 03
Issue: 03
May-June 2026
Received: 19-03-2026
Accepted: 20-04-2026
Published: 18-05-2026
Page No: 83-88

Abstract

Dengue fever has become an increasingly significant public health concern in Nepal, with outbreaks showing strong seasonal patterns influenced by climatic conditions. Temperature, rainfall, and relative humidity affect mosquito breeding dynamics and virus transmission, making climate variables important predictors of dengue incidence. While previous studies in Nepal have primarily examined statistical relationships between climate variability and dengue cases, predictive modeling approaches remain limited. This study develops and compares climate-based machine learning models to predict dengue incidence in Nepal.

Monthly dengue case data and climatic variables, including temperature, rainfall, and relative humidity, were collected for the period 2022–2025. To capture delayed climatic effects on vector ecology and disease transmission, one-month and two-month lagged climate variables were incorporated into the modeling framework. Comparative prediction models were developed using ensemble machine learning techniques, including Random Forest and Gradient Boosting, to model nonlinear relationships between climate variables and dengue incidence.

Model performance was evaluated using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) to assess predictive accuracy. The results demonstrate that machine learning models can effectively capture the complex relationships between climatic factors and dengue incidence. Incorporating lagged climate predictors improved forecasting performance, highlighting the importance of delayed climatic influences on dengue transmission dynamics.

The findings demonstrate the potential of climate-based machine learning approaches for dengue prediction in Nepal and highlight their usefulness for developing early warning systems to support public health preparedness and outbreak prevention.

Keywords: Dengue incidence, climate variability, machine learning, Random Forest, Gradient Boosting, time-lagged predictors, dengue forecasting, Nepal

1. Introduction

Dengue fever has emerged as one of the most important mosquito-borne diseases affecting tropical and subtropical regions worldwide. The disease is transmitted primarily by *Aedes aegypti* mosquitoes and has expanded rapidly due to urbanization, population mobility, and climate variability [1] Climatic conditions play a critical role in shaping dengue transmission dynamics because mosquito development, survival, and viral replication are strongly influenced by environmental factors such as temperature, rainfall, and humidity [2, 3]. Consequently, dengue incidence often exhibits strong seasonal patterns linked to climatic variability.

In Nepal, dengue has transitioned from sporadic outbreaks to a recurring public health challenge over the past decade. Earlier cases were largely reported in the Terai lowland region, but recent outbreaks have increasingly affected hill districts and urban centers, including Kathmandu Valley. National surveillance records indicate substantial outbreaks during recent years, highlighting the growing vulnerability of Nepal to climate-sensitive vector-borne diseases. Environmental changes, urban expansion, and favorable climatic conditions have contributed to the increasing distribution of dengue vectors across diverse ecological zones [4, 5].

Previous research has established that climatic variables influence dengue transmission through both immediate and delayed mechanisms. Rainfall can create mosquito breeding habitats, temperature affects mosquito development and viral incubation periods, and humidity influences mosquito survival and activity levels [3]. However, these environmental influences often operate with short temporal delays due to mosquito life cycles and virus incubation processes.

Machine learning techniques have recently gained attention for infectious disease forecasting because of their ability to capture complex and nonlinear relationships between environmental predictors and disease incidence. Ensemble models such as Random Forest and Gradient Boosting have demonstrated strong performance in epidemiological prediction tasks, while deep learning approaches such as Long Short-Term Memory (LSTM) networks can model temporal dependencies in sequential data. These techniques provide promising tools for improving dengue forecasting accuracy when compared with traditional statistical approaches.

Despite increasing international research on dengue prediction, relatively limited studies in Nepal have applied machine learning methods to forecast dengue incidence using climatic predictors. This study therefore aims to develop and compare climate-based machine learning models for predicting dengue incidence in Nepal using monthly data from 2022–2025. By evaluating the performance of Random Forest, Gradient Boosting, and LSTM models, the study seeks to identify effective approaches for climate-informed dengue forecasting that may support early warning and public health preparedness.

2. Literature Review

2.1. Climate and Dengue Transmission

Climatic conditions are widely recognized as key determinants of dengue transmission dynamics. Environmental variables such as temperature, rainfall, and relative humidity influence mosquito breeding, survival, and virus development within the vector. Temperature affects mosquito growth rate, biting activity, and the extrinsic incubation period of the dengue virus, thereby influencing transmission intensity [2, 3]. Rainfall contributes to the formation of breeding habitats through water accumulation in natural and artificial containers, particularly in urban and peri-urban environments. Relative humidity also affects mosquito survival and activity, increasing the likelihood of human–vector contact.

In addition to immediate climatic influences, several studies have shown that climatic effects on dengue incidence often occur with temporal delay. These delayed responses arise from biological processes such as mosquito breeding cycles and viral incubation periods within the mosquito. Climate change and increasing temperature variability may further alter vector distribution and disease transmission patterns globally [6].

Understanding these climate–disease interactions is therefore important for developing forecasting systems that can anticipate outbreak risk.

2.2. Machine Learning Approaches in Dengue Forecasting

Predictive modeling has become an increasingly important tool for dengue surveillance and outbreak prediction. Traditional statistical models have been widely used to examine relationships between climate variables and dengue incidence. However, dengue transmission is influenced by complex nonlinear interactions among environmental, biological, and socio-ecological factors, which may not always be fully captured by conventional statistical methods. Machine learning approaches offer improved capability for modeling nonlinear relationships and complex interactions among predictors. Ensemble algorithms such as Random Forest and Gradient Boosting are particularly effective in handling high-dimensional data and identifying important predictors influencing disease transmission. These models have demonstrated promising performance in epidemiological prediction tasks because they can capture nonlinear relationships between climate variables and dengue incidence [7].

Deep learning methods have also been applied in recent years for infectious disease forecasting. Long Short-Term Memory (LSTM) networks are especially suitable for time-series prediction because they can capture temporal dependencies and sequential patterns in data. When applied to climate-sensitive diseases such as dengue, LSTM models can learn patterns of seasonal variation and delayed environmental effects that influence outbreak dynamics.

2.3. Research Gap in Nepal

Nepal has experienced a rapid increase in dengue transmission over recent years, with outbreaks expanding from the Terai region to hill districts and urban centers. Climatic variability and environmental change have been identified as important factors contributing to the expanding distribution of dengue vectors in the country [4], [5]. Recent studies have also suggested that increasing temperature suitability may elevate dengue transmission risk in mid-hill regions of Nepal [8].

Despite growing concern over dengue outbreaks, most research in Nepal has focused on epidemiological trends, climate suitability, or outbreak descriptions rather than predictive modeling approaches. Consequently, there remains limited evidence on the application of machine learning techniques for forecasting dengue incidence using climate data in the Nepalese context.

To address this gap, the present study develops and compares climate-based machine learning models for predicting dengue incidence in Nepal. By incorporating climatic variables and lagged predictors, the study aims to evaluate the predictive performance of different machine learning techniques and explore their potential for supporting climate-informed dengue early warning systems.

3. Methodology

3.1. Study Design and Data Sources

This study employed a retrospective time-series modeling approach to predict dengue incidence in Nepal using climate-based predictors. Monthly dengue case data were obtained from national surveillance records maintained by the Epidemiology and Disease Control Division (EDCD), Ministry of Health and Population, Government of Nepal. The dataset covered the period from January 2022 to December 2025 and included district-level reported dengue cases across Nepal [9].

Meteorological data were obtained from the Department of Hydrology and Meteorology (DHM), Nepal. The climatic variables included average monthly temperature (°C), total monthly rainfall (mm), and average monthly relative humidity (%) [10]. District-level dengue case counts were aggregated to national monthly totals to align with the temporal resolution of climate records.

3.2. Data Processing and Feature Engineering

Dengue incidence data and climate variables were merged using month and year identifiers to construct a unified national monthly dataset. Climate variables were prepared using aggregation methods consistent with meteorological reporting practices. Rainfall was represented as total monthly precipitation, while temperature and humidity were represented as monthly averages.

To capture delayed climatic effects on dengue transmission, lagged climate variables were constructed. One-month ($t-1$) and two-month ($t-2$) lag variables were generated for rainfall, temperature, and humidity. These lagged predictors were aligned with dengue incidence at time *t*, allowing the models to account for biological delays associated with mosquito breeding cycles and viral incubation periods.

After lag creation, rows containing missing values resulting from temporal shifting were removed. The final analytical dataset consisted of approximately 44–46 monthly observations depending on the lag structure used for model training.

3.3. Model Development and Evaluation

Three predictive modeling approaches were applied to forecast monthly dengue incidence: Random Forest, Gradient Boosting, and Long Short-Term Memory (LSTM). Random Forest and Gradient Boosting are ensemble machine learning algorithms capable of modeling nonlinear relationships between predictors and response variables. LSTM is a recurrent neural network architecture designed for time-series prediction and capable of capturing temporal dependencies in sequential data.

The input feature set consisted of climate variables (temperature, rainfall, humidity) and their corresponding lagged variables (1-month and 2-month lags). The target variable was monthly dengue incidence.

Model performance was evaluated using two commonly used forecasting accuracy metrics: Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). These metrics quantify the difference between predicted and observed dengue case counts, providing interpretable measures of prediction accuracy. Lower values of MAE and RMSE indicate better model performance.

All analyses and model development were implemented using Python in a reproducible computational environment.

4. Results and Analysis

4.1. Comparative Model Performance

Comparative modeling was conducted to evaluate the ability of climate-based predictors to estimate monthly dengue incidence in Nepal. A chronological train–test strategy was used, with data from 2022–2024 used for training and 2025 reserved for testing. Model performance was assessed using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

Table 1: Comparative Performance of Prediction Models for 2025

Model	MAE	RMSE
Lag-1 Baseline	6.43	15.86
Linear Regression	7.84	21.09
Random Forest	7.02	16.46
Gradient Boosting	7.322	17.291
LSTM	8.27	20.30
Hybrid Model	5.83	14.93

Table 1 presents the performance of the evaluated models. The Lag-1 baseline model, which used the previous month's dengue cases as the predictor, achieved relatively strong performance (MAE = 6.43; RMSE = 15.86), indicating substantial short-term persistence in dengue transmission. Linear Regression produced the highest prediction errors (MAE = 7.84; RMSE = 21.09), suggesting that a simple linear combination of climatic and lagged variables was insufficient to capture the underlying dynamics of dengue incidence.

Among the machine learning models, Random Forest performed better than Linear Regression (MAE = 7.02; RMSE = 16.46), indicating that nonlinear modeling improved predictive accuracy. In contrast, the LSTM model showed comparatively weaker performance (MAE = 8.27; RMSE = 20.30). This may reflect the relatively short length of the monthly time series, which limited the ability of the deep learning model to learn complex temporal patterns effectively.

The Hybrid model achieved the best overall performance, with the lowest MAE (5.83) and RMSE (14.93). This result indicates that combining short-term epidemic persistence with climate-based residual correction provided more accurate predictions than either conventional regression or stand-alone machine learning models.

4.2. Predicted versus Observed Dengue Cases

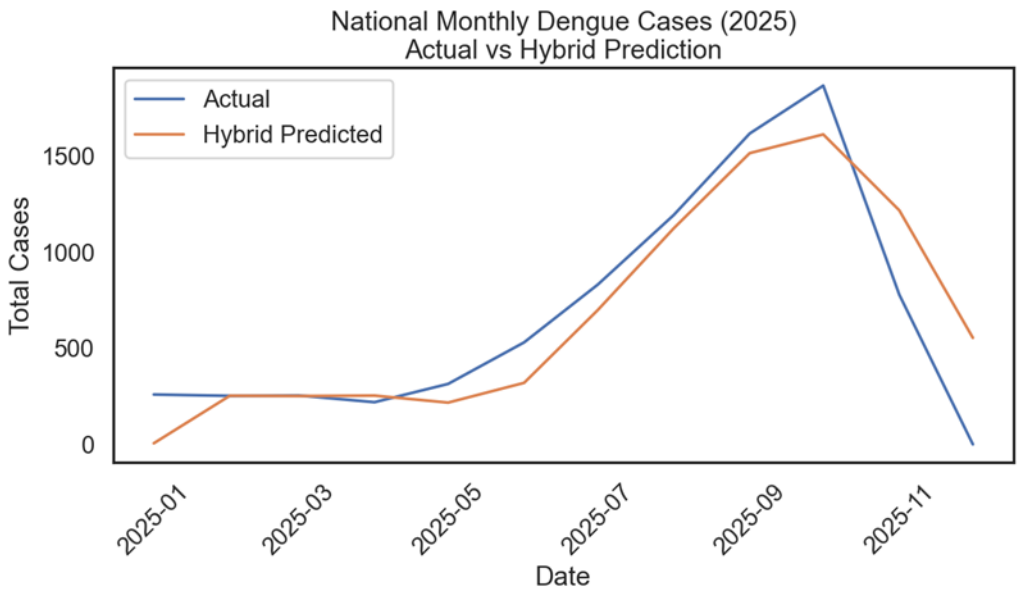

Fig 1: Observed and Predicted Monthly Dengue Cases for 2025 Using the Hybrid Model

Figure 1 shows the comparison between observed monthly dengue cases and predictions from the Hybrid model for the test year 2025. The model closely followed the observed seasonal pattern, including the mid-year increase and the peak period around September–October. Although small deviations were observed during the decline phase, the model reproduced the general outbreak timing and magnitude with reasonable accuracy.

This finding suggests that the Hybrid approach was able to capture both short-term epidemic momentum and climate-related variation in dengue incidence. The improved performance relative to the Lag-1 baseline indicates that climatic predictors added useful information beyond simple persistence alone.

4.3. Feature Importance of the Best-Performing Model

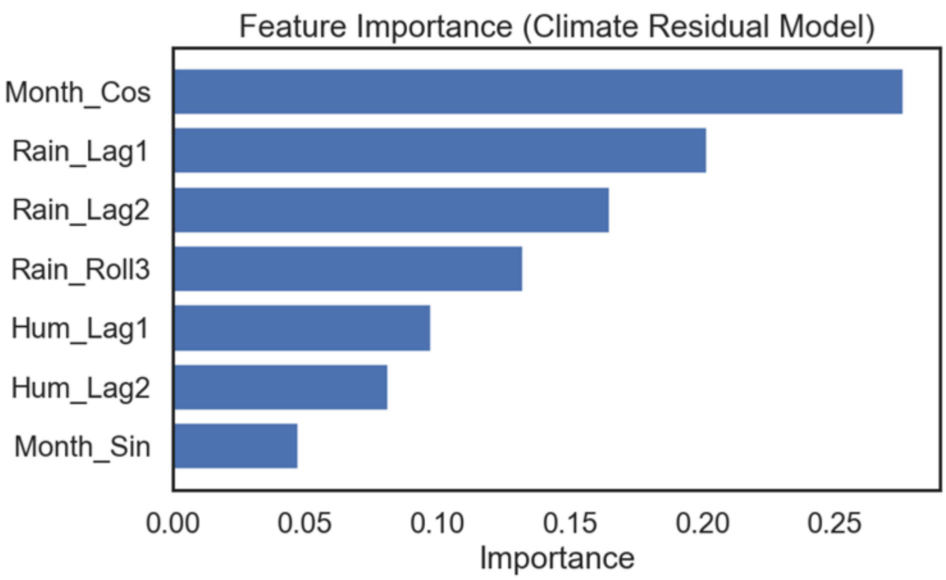

Fig 2: Feature Importance of the Hybrid Climate Residual Model

To identify the variables contributing most strongly to prediction accuracy, feature importance was examined for the Random Forest residual component of the Hybrid model. As shown in Figure 2, Month_Cos had the highest importance, indicating that seasonal timing played a major role in explaining deviations from the baseline model.

Among climatic predictors, Rain_Lag1 and Rain_Lag2 were the most influential variables. This indicates that rainfall from the previous one to two months contributed substantially to dengue prediction, consistent with delayed effects associated with mosquito breeding and transmission cycles. Rain_Roll3 also showed meaningful contribution, while humidity variables had smaller but still noticeable influence.

Overall, the feature importance results suggest that seasonal structure and lagged rainfall were the strongest contributors to improved prediction performance beyond epidemic persistence.

5. Discussion

This study compared multiple predictive approaches for forecasting monthly dengue incidence in Nepal using climate variables and lagged predictors. The results demonstrate that incorporating climate information can improve dengue forecasting performance when combined with temporal persistence in disease transmission.

The Lag-1 baseline model already showed relatively strong predictive performance, indicating that dengue incidence exhibits short-term temporal continuity. In other words, the number of dengue cases in the previous month is a strong predictor of cases in the following month. This finding is consistent with the epidemic momentum often observed in vector-borne diseases, where transmission persists across consecutive months during outbreak periods.

However, the Hybrid model outperformed all other evaluated models, achieving the lowest MAE and RMSE values. By combining a persistence baseline with climate-driven residual correction, the hybrid approach was able to capture both short-term epidemic dynamics and delayed environmental influences. This result suggests that climatic predictors provide additional explanatory power beyond simple temporal persistence.

Among the climatic variables, lagged rainfall indicators emerged as the most influential predictors in the model. Rainfall from one to two months prior contributed substantially to prediction performance, reflecting the biological delay associated with mosquito breeding and virus incubation cycles. These results are consistent with previous studies demonstrating that rainfall often exhibits delayed effects on dengue transmission due to the time required for mosquito population growth and viral development (Mordecai *et al*., 2019; Aguiar *et al*., 2022) [3, 7].

In contrast, the LSTM model did not outperform simpler machine learning or hybrid approaches. One possible explanation is the relatively short time series used in this study. Deep learning models typically require large datasets to effectively learn complex temporal patterns. With only four years of monthly observations, the dataset may have been insufficient for the LSTM model to fully capture nonlinear sequential relationships.

The findings highlight the importance of combining epidemiological persistence with climate-informed predictors for dengue forecasting. Rather than relying solely on historical case counts or purely climate-based models, hybrid approaches that integrate both sources of information may provide improved prediction accuracy.

From a public health perspective, the results suggest that rainfall patterns can serve as useful indicators for anticipating dengue outbreaks in Nepal. Since rainfall variables at one- to two-month lag were strong predictors, increases in monsoon rainfall may provide an early signal for strengthening vector control measures, surveillance activities, and healthcare preparedness.

6. Conclusion

This study developed and compared multiple predictive models to forecast monthly dengue incidence in Nepal using climate variables and lagged predictors. The results demonstrate that climate-informed modeling approaches can improve dengue prediction accuracy when combined with short-term epidemic persistence.

Among the evaluated models, the Hybrid approach achieved the best predictive performance, outperforming Linear Regression, Random Forest, and LSTM models. The results also highlight the importance of lagged rainfall variables, which were identified as key predictors influencing dengue incidence.

These findings suggest that integrating climate indicators with epidemiological persistence can enhance dengue forecasting capability. Such models may contribute to the development of climate-sensitive early warning systems that support proactive public health preparedness in Nepal.

Although the study provides useful insights, several limitations should be acknowledged. The analysis used national-level aggregated monthly data and a relatively short time series, which may limit the ability to capture regional heterogeneity and long-term climatic variability. Future research may extend this approach using district-level datasets, longer time series, and additional environmental or socio-demographic predictors.

Despite these limitations, the results demonstrate the potential of climate-based machine learning and hybrid approaches for dengue forecasting in Nepal and provide a foundation for developing data-driven early warning tools to support dengue prevention and control strategies.

References

1. World Health Organization, "Innovative approaches for preventing and controlling dengue and other mosquito-borne diseases," World Health Organization Regional Office for the Eastern Mediterranean, Oct. 2024. [Online]. Available: https://iris.who.int/handle/10665/380260.
2. V. Megens, *Aedes Aegypti: Ecology, Control and Transmission of Disease*. Nova Science Publishers, Incorporated, 2020.
3. E. A. Mordecai *et al*., "Thermal biology of mosquito-borne disease," *Ecol. Lett.*, vol. 22, no. 10, pp. 1690–1708, 2019, doi: 10.1111/ele.13335.
4. D. Bhandari, P. Bi, J. B. Sherchand, M. Dhimal, and S. Hanson-Easey, "Climate change and infectious disease research in Nepal," *Acta Trop.*, vol. 204, p. 105337, Apr. 2020, doi: 10.1016/j.actatropica.2020.105337.
5. M. Dhimal *et al*., "Climate change and its association with the expansion of vectors and vector-borne diseases in the Hindu Kush Himalayan region: A systematic synthesis of the literature," *Adv. Clim. Change Res.*, vol. 12, no. 3, pp. 421–429, Jun. 2021, doi: 10.1016/j.accre.2021.05.003.
6. J. Rocklöv and R. Dubrow, "Climate change: an enduring challenge for vector-borne disease prevention and control," *Nat. Immunol.*, vol. 21, no. 5, pp. 479–483, May 2020, doi: 10.1038/s41590-020-0648-y.
7. M. Aguiar *et al*., "Mathematical models for dengue fever epidemiology: A 10-year systematic review," *Phys. Life Rev.*, vol. 40, pp. 65–92, Mar. 2022, doi: 10.1016/j.plrev.2022.02.001.
8. B. K. Acharya, L. Khanal, and M. Dhimal, "Increased thermal suitability elevates the risk of dengue transmission across the mid hills of Nepal," *PLoS One*, vol. 20, no. 4, p. e0322031, Apr. 2025, doi: 10.1371/journal.pone.0322031.

9. Epidemiology and Disease Control Division (EDCD), "National dengue surveillance data (2022–2025)," Ministry of Health and Population, Government of Nepal, Kathmandu, Surveillance Report, 2025.
10. Department of Hydrology and Meteorology (DHM), "Climatological records (2022–2025)," Government of Nepal, Kathmandu, Climate Data Report, 2025.

How to Cite This Article
Kadel P, Thapaliya S, Pandey RP, Adhakari D. Comparative modeling of dengue incidence in Nepal using climate-based machine learning techniques. International Journal of Future Engineering Innovations. 2026 May–Jun;3(3):83–88.