Engineering and Applied Science Research
https://www.tci-thaijo.org/index.php/easr/index
Published by the Faculty of Engineering, Khon Kaen University, Thailand

Forecasting dengue severity using machine learning and environmental predictors in Chanthaburi, Thailand

Pitcha Ratanawong*[1)], Pakorn Ditthakit[2)], Pachanut Nunthaitaweekul[3)], Phuong Trang Huynh[4)] and Uruya Weesakul[5)]

[1)]Department of Civil Engineering, Faculty of Engineering, Thammasat School of Engineering, Thammasat University, Pathum Thani 12120, Thailand
[2)]Center of Excellence in Sustainable Disaster Management, School of Engineering and Technology, Walailak University, Nakhon Si Thammarat 80160, Thailand
[3)]Faculty of Nursing, Chulalongkorn University, Bangkok 10330, Thailand
[4)]Independent Consultant, Bangkok 10110, Thailand
[5)]Thammasat University Research Unit in Climate Change and Sustainability, Department of Civil Engineering, Faculty of Engineering, Thammasat School of Engineering, Thammasat University, Pathumthani 12120, Thailand

Received 9 September 2025
Revised 7 November 2025
Accepted 21 November 2025

Abstract

Dengue fever is a vector-borne disease whose dynamics are substantially driven by climate change and environmental variability. To this end, this study proposed the establishment and assessment of machine learning models for predicting the severity of dengue incidents in Chanthaburi province, Thailand, based on climatic and environmental variables from 2018–2022. Both classification accuracy and confusion matrix analyses were used to compare the implementation of four machine learning algorithms: Random Forest, Gradient Boosting, Extra Trees, and CatBoost. The highest-performing model was able to predict a sum score of 0.50 correctly 86.70% of the time, which suggests that the developed system has good predictive ability, especially in homing in on low-severity dengue cases. Still, challenges associated with recognizing high-severity cases persist. Shapley Additive Explanations (SHAP) sensitivity analysis revealed that specific air pollutant levels (PM10, PM2.5), time-lag parameters, and indicators such as temperature and humidity, particularly during certain periods, were significant predictors of dengue severity. Our results reveal the intricate relationships between environmental variables and the pattern of dengue transmission, arguing for a judicious use of machine learning tools as evidence-based support to inform disease control policies. Future studies that consider other variables, longer time series data, and advanced modeling techniques are needed to increase the accuracy of predictions, especially with respect to improving sensitivity for high-severity dengue outbreaks.

Keywords: Chanthaburi, Climate change, Dengue fever, Machine learning, Particulate matter, SHAP

1. Introduction

Climate change is reshaping the landscape of public health, particularly through its impact on vector-borne diseases such as dengue fever. In tropical and subtropical regions, shifts in climate (such as rising temperatures, increased humidity, and changes in rainfall) create ideal conditions for *Aedes* mosquitoes, the primary carriers of dengue, to thrive and spread. These environmental changes can expand the geographic range of mosquito populations and prolong their breeding seasons, thereby increasing the risk and intensity of outbreaks [1].

The World Bank Group and Asian Development Bank [2] report that Thailand is considered highly vulnerable to climate variability and change due to increasing natural hazards, including heavy rainfall, floods, and droughts. With Thailand being in a tropical zone, temperature, precipitation, and pollution may affect disease patterns and health-related issues. Recent trends illustrate this growing threat [2]. For example, several countries in Southeast Asia have reported dramatic increases in dengue infections following periods of unusual heat and rainfall. Indonesia, for instance, saw a surge in dengue cases in early 2024 that coincided with a hotter-than-average season [3]. These developments underscore how climate change acts as a catalyst in the spread of dengue.

In Thailand, the average temperature is projected to increase by 0.95°C–3.23°C above the 2005 baseline by the 2090s [2]. Furthermore, climate change contributes to poor air quality. The 6th Annual World Air Quality Report revealed that some cities in Thailand exceeded WHO PM2.5 guidelines by more than 10 times. As the climate crisis alters weather patterns, it disrupts wind and rainfall, affecting how pollutants are spread. [4].

In a prior study, we conducted a literature review to find and summarize existing published studies from 2013 onward that examine the effects of climate change on human health and diseases in Thailand [5]. This current study aims to continue analyzing climate change and health, focusing on dengue in Chanthaburi Province, Thailand, using pre-existing data from the Thai Ministry of Public

*Corresponding author.
Email address: taita.p@gmail.com
https://doi.org/10.64960/easr.2026.263838

Health and the Thai Meteorological Department. Dengue and Chanthaburi province were chosen because the data from this province was the most robust after our initial review of all available data. Chanthaburi province is in eastern Thailand, on the Gulf of Thailand's shore. We investigated the relationship between climate change and dengue using a quantitative approach. The weather factors we analyzed include temperature, humidity, precipitation, PM2.5, and PM10.

Dengue fever is endemic throughout Thailand [6], and climate is a significant factor in dengue activity transmission. Several studies have shown that fluctuations in temperature, rainfall, and humidity lead to increases in Aedes mosquito populations. For instance, Rahman et al. demonstrated that geographical variation resulted in significantly increased vulnerability for urbanized areas with diverse climatic conditions that facilitate dengue outbreaks in Thailand [7]. Furthermore, Chumpu et al. [8] reported strong relationships between provincial-level weather conditions and dengue fever outbreaks caused by delayed effects of rain or temperature in Southeast Asia from 2001 to 2014. Another study found that incorporating meteorological indicators such as maximum temperature and relative humidity into hybrid models significantly improves the accuracy of dengue predictions in Thailand’s high-risk regions [9]. Hii et al. [10] discovered that the relationship between climatic anomalies and dengue epidemics show a strong correlation between seasonal rainfall and Aedes aegypti proliferation. These findings are reinforced by recent research demonstrating that climate-induced shifts in mosquito ecology, particularly in terms of larval development and adult survival, are key determinants of transmission risk [11, 12]. Climate variables not only serve as predictive indicators but also play a crucial role in comprehending the spatiotemporal patterns of dengue transmission in Thailand. Therefore, climate-sensitive data is imperative to improve surveillance and direct proactive public health measures of dengue incidents.

Although time-series forecasting methods like ARIMA and SARIMA can capture the seasonality of dengue outbreaks, they are limited in dealing with nonlinear relationships [13]. More recent advances have introduced machine learning and hybrid models that aim for higher accuracy for early warning systems. Kesorn et al. [14] constructed a model integrating support vector machines with Aedes aegypti infection rates and achieved greater predictive accuracy, particularly for areas suffering the same climate and geographical conditions. Similarly, Jain et al. [15] reported improved dengue forecasting performance by using machine learning to combine meteorological, socio-economic, and surveillance data. These conclusions argue for the development of data-driven, climate-sensitive predictive models to form a key part of Thailand's dengue early warning systems.

Public health prediction models that examine the relationship between diseases or health conditions and climate change are vital for understanding and mitigating the health impacts of a warming planet. These models integrate climate data with epidemiological information to predict how shifts in climate may influence the incidence, spread, and severity of various health conditions. For example, they can forecast the geographical expansion of vector-borne diseases like dengue as warmer temperatures and increased rainfall create favorable conditions for mosquito breeding. Similarly, models can assess how extreme weather events, such as heatwaves or hurricanes, exacerbate cardiovascular and respiratory conditions [16-18].

Molina-Guzmán et al. [19] conducted a global systematic review of studies using modeling methods to determine the impacts of climate and environmental variations on the incidence of vector-borne and infectious diseases, and identified over ten different modeling methods used to predict vector distribution. They concluded that as global temperatures continue to rise, the frequency of extreme events is likely to increase as well. This result underscores the importance of understanding the impact of climate and environmental changes on the occurrence of infectious diseases as a key public health priority.

Dengue outbreak prediction involves discovering the complicated relationship between climate, socioeconomic conditions, and environmental variables. Numerous studies have revealed that various meteorological variables (i.e., temperature, precipitation, and relative humidity) are highly correlated with the temporal and geographical distribution of dengue in Thailand. Two studies developed a dengue outbreak predicting model by integrating disease monitoring data with meteorological and socio-economic variables. These studies showed the importance of using delayed weather observations and urbanization in the new model [15, 20]. Chumpu et al. [8] discovered a significant correlation with meteorological data and provincial-level dengue incidence. Kerdprasop et al [21]. utilized machine learning and remote-sensing data for predicting dengue incidence. They found that the developed models frequently outperform conventional models, even in the absence of specific knowledge regarding the underlying association between dengue cases and several environmental factors. Abdulsalam et al. [22] found a strong correlation between climatic indicators, including cloud cover, temperature, and precipitation, and the incidence of dengue fever epidemics in southern Thailand. Such evidence indicates that local climatic adaptation must be included in predictive models. Furthermore, novel techniques like support vector machines (SVM) or quadratic discriminant functions have demonstrated efficacy in predicting dengue morbidity, even when incorporating climatic and entomological data into the study, as do Random Forests and hybrid methods like Chi-squared Automatic Interaction Detection (CHAID) [13, 14, 23]. These results show that physically based computational models that use climate data are becoming more useful for making dengue forecasts with realistic short-term time and space limits. This is especially true in countries like Thailand that are at risk for health problems caused by climate change.

Recently, several studies have used conventional and advanced machine learning (ML) and deep learning models for dengue outbreak prediction, with their primary emphasis on climate/meteorological drivers. High-end predictive performance of advanced algorithms such as LSTM, XGBoost, and hybrid approaches was noticed in countries like India [24, 25], Vietnam [26], Bangladesh [27], Pakistan [28], Brazil, and China [9, 29-31]. Adjusting meteorological factors with dengue surveillance data improves forecasting [9, 25, 26, 28]. In the context of early warning systems, novel methods such as feature selection through wrappers [28], ensemble methods [29], and hybrid models to deal with zero-inflated data [27] perform very well. Specifically, in Jaipur, India, 1D-CNN achieved high predictive accuracy [24], while in Vietnam, the model that performed best was negative binomial regression (NB) [25]. Rainfall was identified as a critical factor in ANN and XGBoost, which provided useful results for Bangladesh [25]. Such studies offer proof-of-concept demonstrations of the utility of AI-based models in predicting dengue epidemics and support their application for timely interventions across endemic settings [30]. By leveraging advanced statistical techniques and machine learning, prediction models provide actionable insights for policymakers to develop targeted interventions, allocate resources effectively, and enhance public health resilience against the multifaceted impacts of climate change.

Machine learning-based models, especially random forest, are more effective than conventional statistical models for forecasting dengue incidence [31]. Similarly, a comparative study in Bangladesh concluded that the XGBoost model outperformed Artificial Neural Network (ANN) and linear regression models, demonstrating superior accuracy in predicting dengue cases based on climate data [30]. In Jaipur, India, a study evaluated several deep learning algorithms, finding that a one-dimensional convolutional neural network (1DCNN) model showed the highest accuracy and robustness for predicting dengue cases, even with smaller datasets, when compared to ANN and Long Short-Term Memory (LSTM) models [24]. Hybrid models that combine different algorithms, such as IHLOA (Improved Horned Lizard Optimization Algorithm) with SVM, KNN, or RF, demonstrate superior accuracy and stability compared to single models. These models accurately predict dengue outbreaks by analyzing the correlation and lag effects of climatic factors on dengue incidence [9]. Ensemble methods like XGBoost are known for their ability to handle complex, non-linear data effectively [30].

There is strong consensus across the literature that climatic factors are crucial for accurate dengue forecasting. Precipitation is consistently identified as a primary predictor of dengue incidence across multiple studies [26, 30]. Other significant climatic factors include temperature, rainfall, relative humidity, and wind speed, which influence mosquito breeding, survival, and dispersal [9, 26, 31]. Both minimum and maximum temperatures are significant predictors. High temperatures are noted to accelerate mosquito reproduction and growth. Rainfall is identified as a key factor, often the most significant predictor, as it creates breeding sites for mosquitoes. Relative humidity is consistently shown to influence dengue incidence, as it affects mosquito survival and viral propagation. Wind speed is also noted as an influential factor, though its effect can be complex, potentially disrupting mosquito flight or aiding in their dispersal. Additionally, meteorological factors exhibit a delayed (lag) effect on dengue incidence. For instance, a lag of 3-4 months for factors like dew point temperature and relative humidity has been observed, reflecting the time required for mosquito life cycles and virus transmission dynamics to manifest in human cases [9].

The primary goal of developing these predictive models is to establish effective early warning systems. These systems can anticipate outbreaks, allowing health authorities to implement preventive measures, manage resources, and raise community awareness in a timely manner [26, 31]. The use of data-driven machine learning models allows for more accurate and reliable forecasts compared to traditional methods. This enables proactive public health strategies, such as targeted vector control and resource allocation, based on scientific evidence rather than reactive responses [26]. Given that dengue is endemic in Thailand, our study aims to achieve two objectives: (1) to investigate the relationship between climate change and dengue outbreaks in Thailand; (2) to utilize artificial intelligence models to enhance prediction accuracy.

The research questions are: (1) Do changes in climate patterns (specifically: temperature, humidity, precipitation, PM2.5, and PM10) impact dengue in Thailand?; and (2) Is there a possibility of using a machine learning system to predict the pattern?

2. Materials and methods

The research methodology involves five main steps: (1) considering a study area affected by dengue incidence, (2) collecting relevant data, (3) exploring various predictive models and developing the selected model using Python programming, (4) evaluating its accuracy using appropriate performance metrics, and (5) conducting sensitivity analyses. The detailed steps of the research methodology are outlined below.

2.1 Study area

Chanthaburi province (Figure 1), located in eastern Thailand, was selected due to its high dengue incidence rates and its varied climatic patterns influenced by both coastal and inland geography. The province's environmental characteristics make it an ideal case study for examining the relationship between vector-borne diseases and climate-related factors [32-34]. Also, the data on dengue and Chanthaburi province were the most robust after our initial review of all available data.

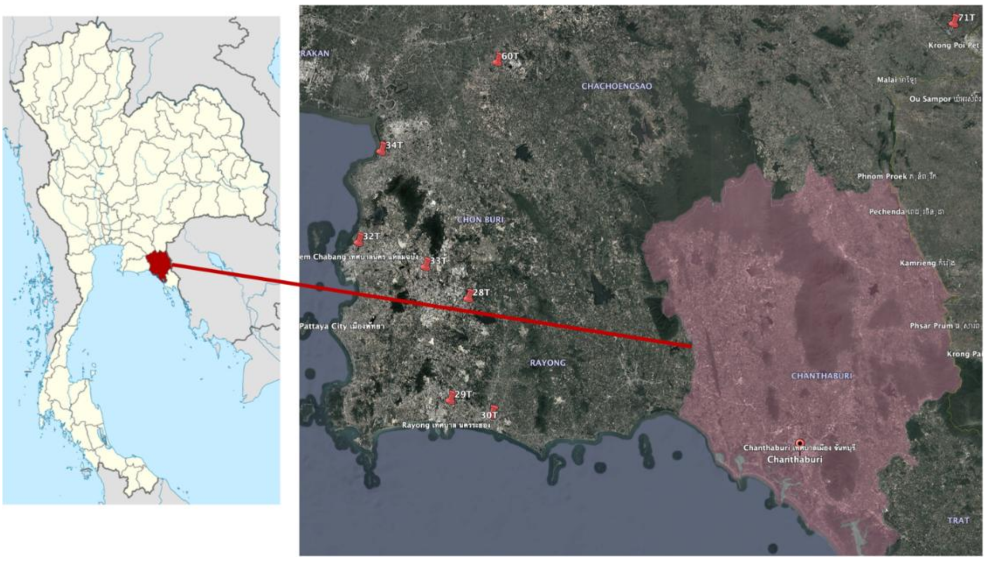


Figure 1 Map of Chantaburi, Thailand, and locations of weather stations

2.2 Data collection

2.2.1 Climate and environmental data

Monthly weather data were obtained from the Thai Meteorological Department for the period from January 2018 to December 2022, comprising a total of 60 datasets matching the existing health data. The environmental variables included in the analysis were as follows: precipitation (mm/month), temperature (°C), including minimum, maximum, and dry bulbs, relative humidity (%), and particulate matter, including PM2.5 and PM10 (μg/m³).

The first three climate variables—precipitation, temperature, and relative humidity—were collected from a weather station located in Chanthaburi province. PM10 data were retrieved from the following monitoring stations surrounding Chanthaburi: 28T and 29T in Rayong Province, 33T and 34T in Chonburi Province, 60T in Chachoengsao Province, and 71T in Sa Kaeo Province. For PM2.5, data were obtained from station 30T in Rayong province and station 32T in Chanthaburi province. The weather stations shown in Figure 1 were selected based on the availability of the data and the proximity to Chanthaburi province. The environmental variables were selected based on prior literature that highlights their potential influence on mosquito breeding, virus replication, and disease transmission.

2.2.2 Health data

We obtained monthly dengue case reports from the Thai Ministry of Public Health's national disease surveillance system, which compiles data from both inpatient and outpatient health services. Among the 298 reportable conditions monitored, this study focused on dengue cases in Chanthaburi Province from January 2018 to December 2022, yielding a total of 60 monthly records. This period represented the complete dataset available from the Ministry of Health at that time. We categorized the data into two severity levels based on incidence: low and high. Cases with values below the 50th percentile were classified as low severity (56 instances), while those at or above the 50th percentile were classified as high severity (14 instances). Because the World Health Organization and no other organization have standard guidelines based on a single 50th percentile criterion for dengue severity, we used the median as a guide for our study [35, 36].

2.3 Predictive modelling and model development

We selected four machine learning models to compare their performance in predicting the severity of dengue incidents based on observed climate and air quality data. A brief description of each model is provided below.

2.3.1 Random forest

Freiman first introduced Random Forest (RF) as an ensemble learning algorithm that is rooted in decision trees [37]. This method helps in preventing overfitting in regression tasks and handling missing data in classification tasks. RF applies the bootstrap aggregation (bagging) methodology, generating a number of data sets by drawing with replacement to create multiple training trees [38-40]. The accuracy of the model increases and overfitting is reduced when we increase the number of trees, which allows us to fit more complex non-linear relationships [41, 42]. RF is an ensemble method that comprises multiple basic units that are individual decision trees (DTs), and aggregates them together [43]. RF effectively is an ensemble of decision trees, where higher rules (trees) average up to better predictions.

2.3.2 Gradient boosting regressor

Gradient Boosting Regressor (GBR) is an ensemble learning technique that constructs a series of decision trees sequentially, where each tree aims to correct the prediction errors of the previous one [44]. As noted by Rao et al. [45], GBR utilizes the boosting technique to strengthen weak learners—typically decision trees—by iteratively improving model performance. The algorithm functions by optimizing a loss function through gradient descent, selecting subsequent models that move in the direction of the negative gradient of the loss [37]. GBR relies on an ensemble of base models and fits simple learners to different data regions [46]. Three key components form the foundation of GBR: a loss function to be minimized, a weak learner to generate predictions, and an additive model to combine learners over iterations [47]. With proper tuning, these elements help mitigate overfitting. GBR constructs the predictive model in a stage-wise manner and updates predictions by minimizing the residual errors at each iteration.

2.3.3 Extra trees

The Extremely Randomized Trees (Extra Trees or ET) algorithm is a relatively recent extension of the Random Forest (RF) method, designed to reduce the likelihood of overfitting [48]. Like RF, ET trains each base estimator using a random subset of features [49, 50]. However, ET differs by selecting split features and values entirely at random and using the entire training dataset rather than bootstrap samples, as RF does [51]. ET splits the dataset into multiple subgroups (child samples) for individual predictions, which are then averaged to enhance accuracy and mitigate overfitting. Key features of ET include random split point selection and training with the full dataset [48].

2.3.4 CatBoost

CatBoost is a gradient boosting algorithm designed for efficient, high-performance machine learning, particularly excelling in handling categorical and diverse data with minimal preprocessing. Developed by Yandex in 2017, CatBoost is an open-source algorithm that supports various data types—text, audio, images—within one framework. Its standout feature is the ability to handle categorical variables automatically, removing the need for manual encoding and reducing the risk of information loss [52]. Built on decision trees and gradient boosting, CatBoost iteratively combines weak learners to optimize a loss function. It uniquely uses oblivious decision trees, where each level splits on the same feature under the same condition. This enhances computational efficiency, simplifies model structure, and reduces overfitting. These strengths make CatBoost ideal for real-world applications requiring accuracy, speed, and minimal data preparation.

The predictive models were developed using Python programming and executed in the Google Colab environment. Given the small dataset size of only 60 instances, we employed fivefold cross-validation to ensure robust model evaluation. To optimize the hyperparameters of each predictive model, we utilized Optuna—an automatic hyperparameter optimization framework specifically designed for machine learning.

2.5 Accuracy evaluation metrics

We evaluated the performance of the machine learning models in predicting dengue incident severity using two primary metrics: accuracy and the confusion matrix.

Accuracy measures the proportion of correctly classified cases—both low and high severity—out of the total number of observations. This metric provided an overall measure of each model's predictive performance. It is calculated as:

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

where: TP = True Positives (high severity correctly predicted as high)
TN = True Negatives (low severity correctly predicted as low)
FP = False Positives (low severity incorrectly predicted as high)
FN = False Negatives (high severity incorrectly predicted as low)

Additionally, confusion matrices were generated for each model to offer a more detailed view of classification outcomes. The confusion matrix summarizes the number of true positive, true negative, false positive, and false negative predictions, enabling the identification of specific patterns of misclassification. In this study, the confusion matrix allowed assessment of how well each model distinguished between low and high dengue incident severity. All metrics were computed using functions from the scikit-learn library in Python, with results reported as mean values from cross-validation procedures to ensure robustness and reliability in performance evaluation.

2.6 Sensitivity analysis

The influence of single input parameters on the prediction results was investigated by means of a sensitivity analysis in the context of the best model performing for predictivity. The step-by-step feature influences were illustrated by Shapley Additive Explanations (SHAP), a model-agnostic explanatory approach. SHAP values calculate the average contribution of each feature to the model output and then explain both global and local interpretability. Two types of SHAP plots were used: the summary bar plot, which shows the mean absolute SHAP value for each feature, and the Beeswarm plot, which displays the distribution and directionality of the features' effects for every observation. This study detailed the relative importance of meteorological and environmental predictors for dengue incidence prediction and how modifications to each variable corresponded with different predicted results. This knowledge is critical both in understanding what models do and designing public health surveillance decision-making processes.

3. Results and discussion

3.1 Data analysis

All climate and health datasets were aggregated monthly to ensure consistent temporal resolution. Descriptive statistical analyses were conducted to examine trends and variability in both the climate and dengue datasets. Based on the descriptive statistical analysis in Table 1 and boxplots (Figure 2) of the climate and air quality data categorized by dengue incident severity (High vs. Low), several key patterns emerge. Overall, high-severity months show higher minimum temperatures but similar maximum temperatures, along with higher humidity, more rainfall, and lower particulate levels. This suggests that warmer and more humid conditions may support mosquito breeding and virus transmission [53-55].

Table 1 Descriptive statistical analysis

Climate data	Descriptive statistical analysis					
	Max	**Min**	**Mean**	**SD**	**Skewness**	**Kurtosis**
Rainfall (mm)	806.60	0.00	244.28	221.09	0.85	-0.26
MinTemp (°C)	26.40	20.60	24.25	1.25	-0.76	0.27
MaxTemp (°C)	34.90	30.70	32.86	1.05	-0.06	-0.90
Dry Bulb Temp (°C)	30.00	25.90	28.06	0.84	-0.26	0.03
Relative Humidity (%)	89.00	62.00	78.27	6.71	-0.68	-0.37
PM10_28T (μg/m³)	85.54	24.17	43.42	14.74	1.12	0.80
PM10_29T (μg/m³)	81.66	23.07	43.63	14.36	0.78	-0.19
PM10_33T (μg/m³)	96.07	26.25	45.46	17.14	1.33	1.54
PM10_71T (μg/m³)	74.78	18.40	37.95	14.72	0.65	-0.44
PM10_34T (μg/m³)	73.10	5.90	32.86	16.28	0.53	-0.22
PM10_60T (μg/m³)	81.76	23.61	45.43	15.51	0.75	-0.12
PM2.5_30T (μg/m³)	42.53	7.06	17.83	8.62	0.96	0.24
PM2.5_32T (μg/m³)	46.00	7.18	20.05	9.72	1.09	0.53

From the distribution characteristics, most distributions are well-behaved. Temperatures are close to symmetric (MaxTemp skew −0.06; kurtosis −0.90), and relative humidity is slightly left-skewed (−0.68). Rainfall shows a moderate right skew (0.85) with light tails (kurtosis −0.26). Particulate matters are less balanced: PM10 (measured from station 33T) has a strong right tail and heavier tails overall (skew 1.33; kurtosis 1.54), with PM2.5 (measured from station 32T) and PM10 (measured from station 28T) also showing upper-tail elongation, pointing to occasional spikes. Together, the climate variables usually stay within a narrow range, with intermittent extremes. Any relation to dengue severity should be tested with groupwise comparisons or multivariable models, but the summaries align with the view that warm, humid conditions can aid transmission while particulate peaks are episodic. This supports building climate-aware early-warning tools and evaluating them alongside particulate indicators.

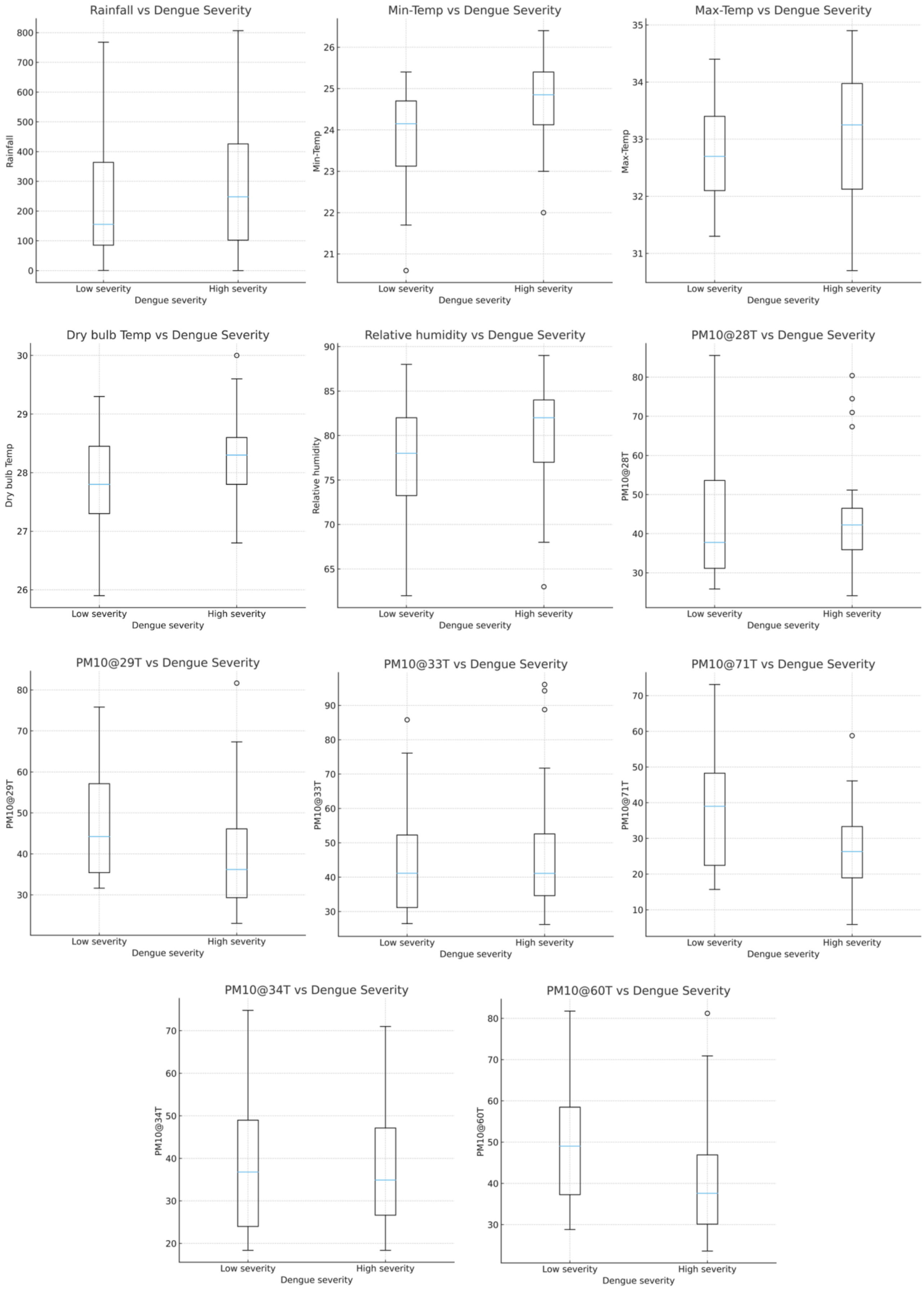


Figure 2 Boxplot of each climate data with dengue incident

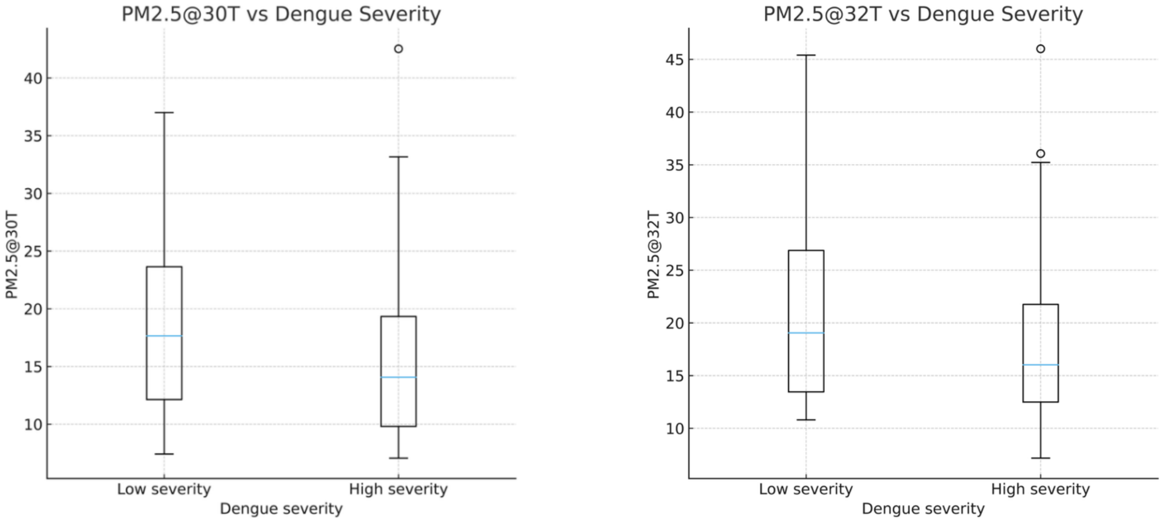


Figure 2 Boxplot of each climate data with dengue incident (cont.)

The boxplots reveal clear distinctions in air quality variables. Most PM10 and PM2.5 concentrations (across various monitoring stations) show lower median values during high-severity months than during low-severity months, although some outliers exist. This could imply that severe dengue outbreaks are not strongly correlated with elevated particulate matter levels, or that other climatic factors (such as temperature, and humidity) play a more direct role in dengue dynamics.

3.2 Optimal hyperparameters of the predictive model

To enhance predictive performance, the hyperparameters of all machine learning models were optimized using Optuna. We performed 50 trials for each model, employing cross-validation to identify parameter settings yielding the highest mean classification accuracy. Table 2 shows defined parameter search spaces and optimal hyperparameters for machine learning models. Hyperparameters, such as the number of estimators, tree depth, learning rate, and minimum samples per split or leaf, strongly shape model accuracy and generalization. Defaults can be serviceable, but systematic tuning often improves predictive performance and limits overfitting, especially in complex or high-dimensional settings. Research underscores dataset-specific tuning, advanced search strategies (e.g., Bayesian optimization, genetic algorithms), and the effects of hyperparameter choices on interpretability and computational efficiency [17, 56-58]. These hyperparameters reflect the complexity required by each model to capture the non-linear relationships between climatic variables and dengue incidence in Chanthaburi province.

Table 2 Hyperparameters for machine learning models

Model	Parameter	Range	Optimal Value
Random Forest	n_estimators	50–200	51
	max_depth	3–20	19
	min_samples_split	2–10	4
	min_samples_leaf	1–10	10
Gradient Boosting	n_estimators	50–200	64
	learning_rate	0.01–0.30	0.14
	max_depth	3–20	10
	min_samples_split	2–10	5
	min_samples_leaf	1–10	5
Extra Trees	n_estimators	50–200	50
	max_depth	3–20	15
	min_samples_split	2–10	4
	min_samples_leaf	1–10	10
CatBoost	iterations	50–200	69
	learning_rate	0.01–0.30	0.19
	depth	3–10	5

3.3 Model performance comparison

In this study, a total of four models (Random Forest, Gradient Boosting, Extra Trees, and CatBoost) were developed to forecast the dengue incidence severity in Chanthaburi province based on weather and environmental factors. Validation of each model: mean classification accuracy by fivefold cross-validation. Table 3 shows the results that all models had a high prediction rate with mean accuracy from 0.80 to 0.85. CatBoost, performed the best with a mean accuracy score of 0.85. The accuracies of Gradient Boosting and Random Forest were 0.82 and 0.80, respectively, where they also performed better than the other classifiers mentioned and close to Extra Trees at 0.83. These results also reinforce the compatibility of CatBoost for modeling highly non-linear ecological and epidemiological structures (which in part could be attributed to its capability of dealing with categorical variables—when compared with other boosting algorithms), in addition to helping to control against overfitting through the ordered boosting techniques. This minor variance seen in model performance could imply that ensemble learning methods are reliably accurate in forecasting dengue outbreaks from climate variables but that advanced boosting techniques like CatBoost may offer an additional edge with high predictive precision. High performance of the CatBoost Model has indicated its viability as an alert tool to prevent dengue outbreaks and will help public health authorities with better resource planning and timely intervention.

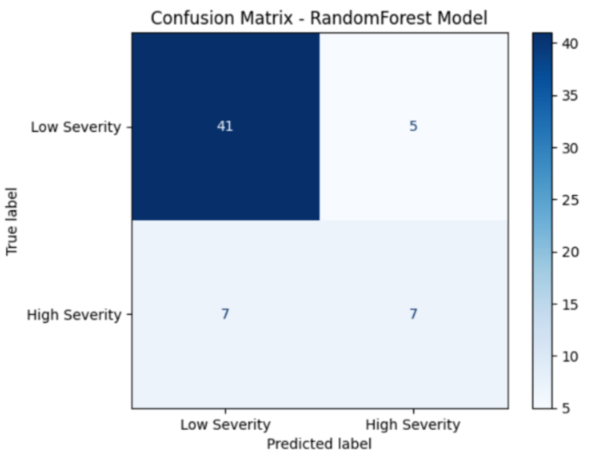
Table 3 Mean accuracy for predicting dengue incidence severity in Chanthaburi province by model.

Model	Mean Accuracy
Random Forest	0.80
Gradient Boosting	0.82
Extra Trees	0.83
CatBoost	0.87

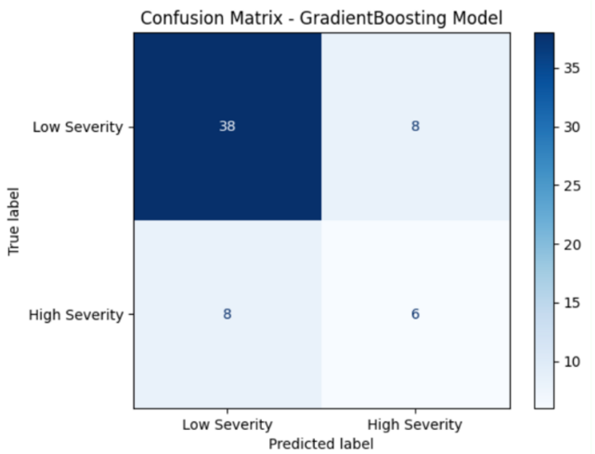
A confusion matrix analysis (Figure 3) was conducted for each model to further elucidate their classification performance. The Random Forest model correctly classified 41 out of 46 instances of low incident severity, but misclassified 5 as high severity, indicating a few false positives. For high-severity cases, it accurately identified 7 out of 14 instances, while 7 were misclassified as low-severity, yielding a recall of 50.00% for high-severity and an overall accuracy of approximately 80.00%.

The Gradient Boosting model showed slightly lower performance, correctly predicting 38 low-severity cases while misclassifying eight as high-severity. For high-severity, only 6 out of 14 cases were correctly identified, and 8were misclassified as low severity, resulting in a recall of 42.90% for high-severity and an overall accuracy of around 82.00%.

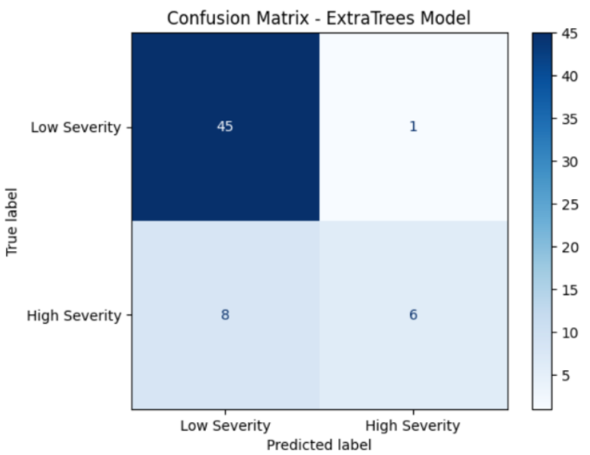
The Extra Trees model demonstrated improved accuracy, correctly classifying 45 low-severity cases with only 1 false positive, reflecting high specificity. However, for high-severity instances, 6 out of 14 were correctly predicted, while eight were misclassified as low severity, similar to previous models, resulting in a recall of 42.90% for high-severity and an overall accuracy of approximately 83.30%.



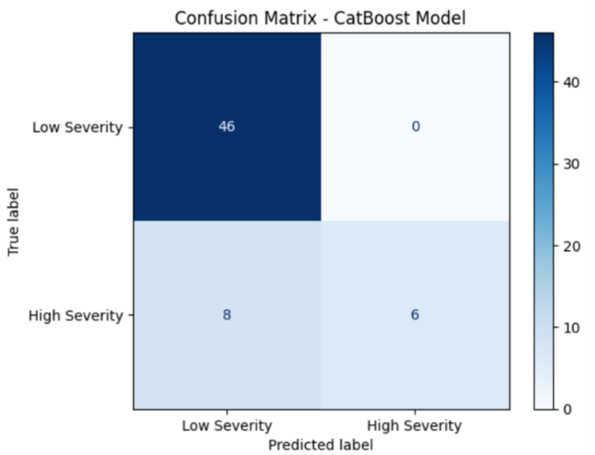
(a)



(b)



(c)



(d)

Figure 3 Confusion matrices illustrating the performance of four ML models

Among all models, CatBoost achieved the strongest performance, correctly classifying all 46 instances of low severity without any false positives, thus demonstrating perfect specificity for low severity detection. However, for high-severity cases, only 6 out of 14 instances were accurately identified, while 8 were misclassified as low-severity, yielding a recall of 42.90% for the high-severity class. Despite this limitation in sensitivity, CatBoost maintained perfect precision for predicting high severity, indicating no false positive predictions, and achieved the highest overall accuracy of 86.70% among the evaluated models. These findings collectively suggest that while all models performed well in correctly identifying low-severity dengue cases, distinguishing high-severity events remains a challenge, with CatBoost offering the best balance between specificity and overall predictive performance.

3.4 Influence of each input variable on the model predictions

We evaluated the influence of individual input variables on CatBoost's predictions for dengue incident severity using Shapley Additive Explanations (SHAP). Two types of SHAP plots were generated to interpret model behavior: a bar plot summarizing mean feature importance values and Beeswarm plot illustrating the distribution and directionality of each variable's effect on the model output.

Figure 4 (a) presents the SHAP bar plot, showing the mean absolute SHAP values for all input features. This plot reveals that variables related to particulate matter concentrations, particularly PM10 measured at stations 34T, 32T, and 29T, had the highest average impact on model predictions. These findings suggest that air pollution levels, especially specific lagged values of PM10 and PM2.5, are significant predictors in forecasting dengue severity in Chanthaburi province. Minimum temperature (MinTemp) also showed considerable influence, followed by PM10 at station 28T, and relative humidity (RH). Other variables, including rainfall, PM10 aggregated values, and dry bulb temperature, contributed less prominently to the model's output.

Figure 4 (b) depicts the SHAP Beeswarm plot, which provides a detailed visualization of how each feature's values influence the model's predictions across all observations. In this plot, each dot represents a single SHAP value for a specific feature and observation. The horizontal spread indicates the magnitude of the feature's impact, while color represents the actual feature value (from low in blue to high in pink). For example, higher values of PM10 at station 34T and PM2.5 at station 32T generally pushed predictions towards higher dengue severity, as indicated by SHAP values skewed to the right. Conversely, certain lower values of MinTemp and RH were associated with predictions of lower severity, reflected by negative SHAP values. This visualization confirms that both the magnitude and direction of individual feature contributions vary across observations, highlighting the complex relationships between climate, environmental variables, and dengue severity.

These SHAP analyses underscore the importance of air quality variables and specific climatic factors in predicting dengue severity, offering valuable insights for public health surveillance and early warning systems in the context of Chanthaburi province.

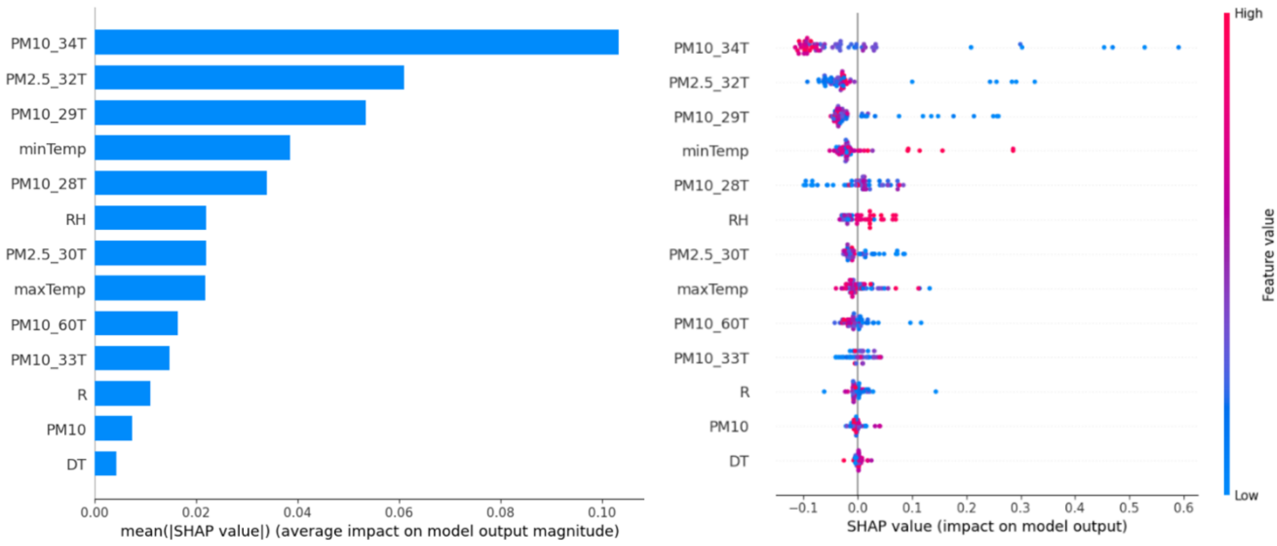


Figure 4 SHAP plots feature the importance for predicting dengue severity using CatBoost in Chanthaburi province: (a) mean absolute SHAP values ranking feature influence; (b) Beeswarm plot visualizing individual feature impacts and value effects on predictions.

4. Discussion

Climate factors, especially temperature and precipitation, are crucial determinants of dengue disease transmission. Their distinct effects differ by place and are shaped by local environmental and socioeconomic conditions. Research throughout Southeast Asia and China consistently identifies temperature and precipitation patterns as critical determinants for dengue outbreaks, influencing mosquito vector development, survival, and the extrinsic incubation period of the virus [59-63]. Temperature is a primary factor influencing dengue transmission [62]. An increase in weekly average temperatures is consistently associated with a higher risk of dengue. For example, a 1°C rise was found to increase dengue risk by up to 24.00% in parts of Laos and 18.90% in parts of Thailand [59]. An analysis in Laos and Thailand found that weekly average temperatures below 28°C or 29°C increased dengue risk, while temperatures above this threshold had a weaker or decreasing effect [59]. Our findings are consistent with these previous studies. In Chanthaburi province high dengue severity was associated with slightly higher minimum and maximum temperatures, as well as higher dry bulb temperatures and relative humidity.

Rainfall's effect on dengue is more complex than temperature's effect. Moderate rainfall can create abundant breeding sites for mosquitoes, increasing their population and disease risk [59]. In Laos and Thailand, it was found that a weekly cumulative rainfall of up to 60 mm increased dengue risk by 1.80-3.20% [59]. Conversely, heavy rainfall can have a negative effect by flushing out mosquito larvae and pupae from breeding sites, temporarily reducing the mosquito population [61, 63]. Research in northern Thailand found a negative association between total rainfall and dengue morbidity in a multiple regression analysis [63]. Higher relative humidity is also a critical factor, correlating with increased dengue cases. It supports the development, feeding activity, and survival of Aedes mosquitoes [61]. Relatedly, in our study, rainfall contributed less prominently to the model's prediction than did temperature and relative humidity (RH).

However, the impact of climatic conditions is not consistent. Dengue outbreaks in China are predominantly fueled by imported cases from Southeast Asia, which are affected by climatic conditions in the originating countries [60]. The impact of climate is interconnected with land use and land cover (LULC). The process of urbanization and the proliferation of rubber plantations might establish conducive microclimates and breeding environments for dengue vectors, hence exacerbating the risks linked to climatic influences [59]. Conversely, wetlands, especially after flooding events, might be adversely related to dengue risk due to the flushing of breeding sites and the presence of natural predators [59].

The relationship between particulate matter (PM10 and PM2.5) and dengue incidence has been explored in several studies, revealing some significant associations and potential mechanisms. Studies in Singapore and Taiwan have shown that higher concentrations of PM2.5 and PM10 are positively associated with increased dengue cases. In Singapore, PM2.5 and PM10 were found to be positively associated with dengue incidence, with relative risks (RR) of 1.28 and 1.30, respectively, for the 90th percentile of PM concentrations [64]. Similarly, in Taiwan, PM10 and PM2.5 were included in models predicting dengue fever incidence, indicating their potential role in influencing dengue outbreaks [65, 66]. Air pollution, including PM2.5, can modify the effects of meteorological factors on dengue incidence. For instance, in Guangzhou, higher PM2.5 levels were found to increase the risk of dengue during low-precipitation days, while high-precipitation days showed a decreased risk with higher PM2.5 levels [67]. This suggests that PM2.5 can interact with weather conditions to influence dengue transmission dynamics. The exact mechanisms by which PM influences dengue incidence are not fully understood, but several hypotheses have been proposed. PM may affect the life cycle and behavior of mosquito vectors. For example, a study hypothesized that haze and PM could increase mosquito mortality, thereby influencing dengue transmission patterns [68]. PM exposure can affect human immune responses, potentially making individuals more susceptible to infections, including dengue [65, 69]. Interestingly, some studies have found that minimum levels of PM2.5 showed a negative contribution to monthly dengue incidence, suggesting a complex relationship that may depend on various environmental and biological factors [65]. Our study findings suggest that air pollution levels, especially specific lagged values of PM10 and PM2.5, are significant predictors in forecasting dengue severity in Chanthaburi province. However, both the magnitude and direction of PM10 and PM2.5 contributions vary across observations. This concurs with previous studies that highlight the complex relationships between climate, environmental variables, and dengue severity. Our findings can serve as a reference for practical use, particularly by organizations engaged in surveillance planning, to identify high-risk areas and enhance efforts in the prevention and control of dengue fever.

5. Conclusions

Per our findings above, we conclude that changes in climate patterns do impact dengue in Thailand, and a machine learning system can be used to predict the severity of dengue cases. Of the four machine learning approaches employed, CatBoost was the top-performing model with an overall accuracy of 87.00% in general and had considerable predictive power for low severity. Nevertheless, challenges were present in identifying high-severity incidents with accuracy, which should be improved with increased sensitivity towards more severe outbreaks.

Sampling at lags allowed for a refinement in the most appropriate time window of initial conditions, and sensitivity analysis with Shapley Additive Explanations (SHAP) showed that particulate matter concentrations—especially select lagged PM10 and PM2.5 values—were more relevant to mortality risk compared to ozone, temperature, or other covariates; meanwhile, greater optimal bandwidths emerged from sampling such concentration values. PM data collection stations not being located in Chanthaburi province (Figure 1) might have impacted the PM measurements, which would be a limitation of this study. Temperature and humidity made the most significant contribution to model predictions. These findings reveal intricate relationships involving air quality, climate factors, and dengue transmissibility for tailored public health efforts.

Together, the results of this research highlight the potential for using machine learning models as tools in early warning systems to improve dengue surveillance. Further research should use more environmental and socioeconomic variables. Variables such as land use and land cover should also be considered as they are impacted by socioeconomic drivers and have been shown to be associated with dengue fever [54, 70]. Also of note is that Chanthaburi province is a border city that is impacted by issues such as immigration and environmental issues, such as heavy rainfall, marine pollution, land-use change, etc. Future studies using data from border cities should take these issues into consideration. Further research should also increase the training period of models to include earlier time series data and evaluate ensemble or hybrid modeling techniques for even higher prediction of dengue severity outbreaks. Additionally, because high severity of dengue incidence is an important issue for prediction, more categories for levels of severity should be included.

Ethical considerations

As this study uses anonymized secondary data without direct human subjects, ethical approval was deemed exempt by the Human Research Ethics Committee of Thammasat University. There is no direct benefit to the individuals represented in the dataset. However, the findings may inform public health interventions and support forecasting systems for climate-sensitive diseases.

5. Acknowledgements

The authors would like to express their sincere gratitude to the Thai Ministry of Public Health for providing dengue surveillance data and to the Thai Meteorological Department for supplying climate and environmental datasets that made this research possible.

6. References

[1] World Health Organization. Vector-borne diseases [Internet]. 2024 [cited 2025 Apr 17]. Available from: https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases.

[2] World Bank Group, Asian Development Bank. Climate risk country profile: Thailand. Thailand: World Bank Group and Asian Development Bank: 2021. Report No. TCS210314-2.

[3] Alied M, Nguyen D, Aziz JMA, Vinh DP, Huy NT. Dengue fever on the rise in Southeast Asia. Pathog Global Health. 2023;117(1):1-2.
[4] IQAir. Interactive global map of 2024 PM2.5 concentrations by city [Internet]. 2024 [cited 2025 Apr 17]. Available from: https://www.iqair.com/th-en/world-air-quality-report.
[5] Ratanawong P, Nunthaitaweekul P, Huynh PT, Weesakul U. Impact of climate change on human health in Thailand: a literature review. J. Disaster Res. 2025;20(4):423-44.
[6] Centers for Disease Control and Prevention. Thailand: CDC Yellow Book 2024 [Internet]. 2024 [cited 2024 Apr 17]. Available from: https://wwwnc.cdc.gov/travel/yellowbook/2024/itineraries/thailand#infectious.
[7] Rahman MS, Pientong C, Zafar S, Ekalaksananan T, Paul RE, Haque U, et al. Mapping the spatial distribution of the dengue vector *Aedes aegypti* and predicting its abundance in northeastern Thailand using machine-learning approach. One Health. 2021;13:100358.
[8] Chumpu R, Khamsemanan N, Nattee C. The association between dengue incidences and provincial-level weather variables in Thailand from 2001 to 2014. PLoS One. 2019;14(12):e0226945.
[9] Cheng Y, Cheng R, Xu T, Tan X, Bai Y, Yang J. Integrating meteorological data and hybrid intelligent models for dengue fever prediction. BMC Public Health. 2025;25:1516.
[10] Hii YL, Rocklöv J, Ng N, Tang CS, Pang FY, Sauerborn R. Climate variability and increase in intensity and magnitude of dengue incidence in Singapore. Glob Health Action. 2009;2(1):2036.
[11] Chakraborty S, Zigmond E, Shah S, Dayal D, Sylla M, Akorli J, et al. Thermal tolerance of *Aedes aegypti* mosquito eggs is associated with urban adaptation and human interactions. J Therm Biol. 2025;131:104167.
[12] Sharma G, Khan Z, Das D, Singh S, Singh S, Kumar M, et al. Thermal influence on development and morphological traits of *Aedes aegypti* in central India and its relevance to climate change. Parasites Vectors. 2025;18:279.
[13] Puengpreeda A, Yhusumrarn S, Sirikulvadhana S. Weekly forecasting model for dengue hemorrhagic fever outbreak in Thailand. Eng J. 2020;24(3):71-87.
[14] Kesorn K, Ongruk P, Chompoosri J, Phumee A, Thavara U, Tawatsin A, et al. Morbidity rate prediction of dengue hemorrhagic fever (DHF) using the support vector machine and the *Aedes aegypti* infection rate in similar climates and geographical areas. PLoS ONE. 2015;10(5):e0125049.
[15] Jain R, Sontisirikit S, Iamsirithaworn S, Prendinger H. Prediction of dengue outbreaks based on disease surveillance, meteorological and socio-economic data. BMC Infect Dis. 2019;19:272.
[16] Thompson MC, Palmer T, Morse AP, Cresswell M, Connor SJ. Forecasting disease risk with seasonal climate predictions. Lancet. 2000;355(9214):1559-60.
[17] Kaur I, Kumar Y, Sandhu AK, Ijaz MF. Predictive modeling of epidemic diseases based on vector-borne diseases using artificial intelligence techniques. In: Tamrakar S, Choubey SB, Choubey A, editors. Computational intelligence in medical decision making and diagnosis. Boca Raton: CRC Press; 2023. p. 81-100.
[18] Leung XY, Islam RM, Adhami M, Ilic D, McDonald L, Palawaththa S, et al. A systematic review of dengue outbreak prediction models: current scenario and future directions. PLoS Negl Trop Dis. 2023;17(2):e0010631.
[19] Molina-Guzmán LP, Gutiérrez-Builes LA, Ríos-Osorio LA. Models of spatial analysis for vector-borne diseases studies: a systematic review. Vet World. 2022;15(8):1975-89.
[20] Zhao N, Charland K, Carabali M, Nsoesie EO, Maheu-Giroux M, Rees E, et al. Machine learning and dengue forecasting: comparing random forests and artificial neural networks for predicting dengue burden at national and sub-national scales in Colombia. PLoS Negl Trop Dis. 2020;14(9):e0008056.
[21] Kerdprasop N, Kerdprasop K, Chuaybamroong P. Computational intelligence and statistical learning performances on predicting dengue incidence using remote sensing data. Adv Sci Technol Eng Syst J. 2020;5(4):344-50.
[22] Abdulsalam FI, Antunez P, Yimthiang S, Jawjit W. Influence of climate variables on dengue fever occurrence in the southern region of Thailand. PLOS Glob Public Health. 2022;2(4):e0000188.
[23] Kerdprasop K, Kerdprasop N, Chuaybamroong P. Forecasting dengue incidence with the chi-squared automatic interaction detection technique. 2019 2nd Artificial Intelligence and Cloud Computing Conference, AICCC 2019; 2019 Dec 21-23; Kobe Japan. USA: ACM; 2019. p. 37-42.
[24] Dhaked DK, Sharma O, Gopal Y, Gopal R. Prediction of dengue patients using deep learning methods amid complex weather conditions in Jaipur, India. Discov Public Health. 2025;22:58.
[25] Tuan D. Leveraging climate data for dengue forecasting in Ba Ria Vung Tau Province, Vietnam: an advanced machine learning approach. Trop Med Infect Dis. 2024;9(10):250.
[26] Mazhar B, Ali NM, Manzoor F, Khan MK, Nasir M, Ramzan M. Development of data-driven machine learning models and their potential role in predicting dengue outbreak. J Vector Borne Dis. 2024;61(4):503-14.
[27] Francisco ME, Carvajal TM, Watanabe K. Hybrid machine learning approach to zero-inflated data improves accuracy of dengue prediction. PLoS Negl Trop Dis. 2024;18(10):e0012599.
[28] Al Mobin M. Forecasting dengue in Bangladesh using meteorological variables with a novel feature selection approach. Sci Rep. 2024;14:32073.
[29] Sebastianelli A, Spiller D, Carmo R, Wheeler J, Nowakowski A, Jacobson LV, et al. A reproducible ensemble machine learning approach to forecast dengue outbreaks. Sci Rep. 2024;14:3807.
[30] Chowdhury AH. Comparison of deep learning and gradient boosting: ANN versus XGBoost for climate-based dengue prediction in Bangladesh. Health Sci Rep. 2025;8:e70714.
[31] Sutriyawan A, Rahardjo M, Martini M, Sutiningsih D, Rattanapan C, Kassim NFA. Global forecasting models for dengue outbreaks in endemic regions: a systematic review. Zh Mikrobiol Epidemiol Immunobiol. 2025;102(2):331-42.
[32] Rocklöv J, Dubrow R. Climate change: an enduring challenge for vector-borne disease prevention and control. Nat Immunol. 2020;21:479-83.
[33] The Health Aisle Team. How does climate change affect diseases like Malaria, Dengue and Chikungunya? [Internet]. n.d. [cited 2025 Apr 17]. Available from: https://www.prudential.co.th/corp/prudential-th/en/we-do-pulse/health-wellness/how-does-climate-change-affect-diseases-like-malaria-dengue-and-chikungunya/.
[34] Tewari P, Guo P, Dickens B, Ma P, Bansal S, Lim JT. Associations between dengue incidence, ecological factors, and anthropogenic factors in Singapore. Viruses. 2023;15(9):1917.

[35] Narvaez F, Gutierrez G, Pérez MA, Elizondo D, Nuñez A, Balmaseda A, et al. Evaluation of the traditional and revised WHO classifications of dengue disease severity. PLoS Negl Trop Dis. 2011;5(11):e1397.
[36] Ajlan BA, Alafif MM, Alawi MM, Akbar NA, Aldigs EK, Madani TA. Assessment of the new World Health Organization's dengue classification for predicting severity of illness and level of healthcare required. PLoS Negl Trop Dis. 2019;13(8):e0007144.
[37] Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat. 2001;29(5):1189-232.
[38] Ditthakit P, Pinthong S, Salaeh N, Binnui F, Khwanchum L, Pham QB. Using machine learning methods for supporting GR2M model in runoff estimation in an ungauged basin. Sci Rep. 2021;11:19955.
[39] Dou J, Yunus AP, Tien Bui D, Merghadi A, Sahana M, Zhu Z, et al. Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. Sci Total Environ. 2019;662:332-46.
[40] Zhu L, Qiu D, Ergu D, Ying C, Liu K. A study on predicting loan default based on the random forest algorithm. Procedia Comput Sci. 2019;162:503-13.
[41] Safari MJS. Hybridization of multivariate adaptive regression splines and random forest models with an empirical equation for sediment deposition prediction in open channel flow. J Hydrol. 2020;590:125392.
[42] Schoppa L, Disse M, Bachmair S. Evaluating the performance of random forest for large-scale flood discharge simulation. J Hydrol. 2020;590:125531.
[43] Ali M, Prasad R, Xiang Y, Yaseen ZM. Complete ensemble empirical mode decomposition hybridized with random forest and kernel ridge regression model for monthly rainfall forecasts. J Hydrol. 2020;584:124647.
[44] Otchere DA, Ganat TOA, Ojero JO, Tackie-Otoo BN, Taki MY. Application of gradient boosting regression model for the evaluation of feature selection techniques in improving reservoir characterisation predictions. J Pet Sci Eng. 2022;208:109244.
[45] Rao H, Shi X, Rodrigue AK, Feng J, Xia Y, Elhoseny M, et al. Feature selection based on artificial bee colony and gradient boosting decision tree. Appl Soft Comput. 2019;74:634-42.
[46] Alexopoulos MJ, Müller-Thomy H, Nistahl P, Šraj M, Bezak N. Validation of precipitation reanalysis products for rainfall-runoff modelling in Slovenia. Hydrol Earth Syst Sci. 2023;27(13):2559-78.
[47] Bagalkot N, Keprate A, Orderløkken R. Combining computational fluid dynamics and gradient boosting regressor for predicting force distribution on horizontal axis wind turbine. Vibration. 2021;4(1):248-62.
[48] Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Mach Learn. 2006;63:3-42.
[49] Ahmad MW, Reynolds J, Rezgui Y. Predictive modelling for solar thermal energy systems: a comparison of support vector regression, random forest, extra trees and regression trees. J Cleaner Prod. 2018;203:810-21.
[50] John V, Liu Z, Guo C, Mita S, Kidono K. Real-time Lane estimation using deep features and extra trees regression. 7th Pacific-Rim Symposium, Image and Video Technology 2015, PSIVT 2015; 2015 Nov 25-27; Auckland, New Zealand. Cham: Springer; 2016. p. 721-33.
[51] Ahmad MW, Mourshed M, Rezgui Y. Tree-based ensemble methods for predicting PV power generation and their comparison with support vector regression. Energy. 2018;164:465-74.
[52] Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. 32nd International Conference on Neural Information Processing Systems; 2018 Dec 2-8; Montréal, Canada. USA: NeurIPS; 2018. p. 6838-48.
[53] Jia X, Wang Y, Wang Z, Ji X. New insights into the main meteorological factors affecting mosquito densities: time lags, interaction effects, and species specificity. Environ Ecol Stat. 2025;32:1091-113.
[54] Ardiansyah Akbar K, Kumala Fatma R, Elamouri F, Rockstroh JK. Climate change and dengue fever: a 14-year study of mortality trends during 2010–2023 in Indonesia. Travel Med. Infect. Dis. 2025;67:102893.
[55] Sy AK, Biggs J, Chowdhury SK, Quinones MA, Jones-Warner W, Ashall J, et al. The impact of altitude and population density on dengue incidence and force of infection across the Philippines: implications for climate-adapted surveillance [Internet]. SSRN [Preprint]. 2025 [cited 2025 Apr 17]. Available from: https://doi.org/10.2139/ssrn.5400338.
[56] Hancock JT, Khoshgoftaar TM. CatBoost for big data: an interdisciplinary review. J Big Data. 2020;7:94.
[57] Dunias ZS, Van Calster B, Timmerman D, Boulesteix AL, Van Smeden M. A comparison of hyperparameter tuning procedures for clinical prediction models: a simulation study. Stat Med. 2024;43(6):1119-34.
[58] Nishio M, Nishizawa M, Sugiyama O, Kojima R, Yakami M, Kuroda T, et al. Computer-aided diagnosis of lung nodule using gradient tree boosting and Bayesian optimization. PLoS ONE. 2018;13(4):e0195875.
[59] Zafar S, Rocklöv J, Paul RE, Shipin O, Rahman MS, Pientong C, et al. Landscape and climatic drivers of dengue fever in Lao People's Democratic Republic and Thailand: a retrospective analysis during 2002–2019. Landsc Ecol. 2025;40:102.
[60] Sang S, Wang Y, Liu Q, Chen P, Li C, Zhang A. Predicting dengue incidence in high-risk areas of China through the integration of Southeast Asian and local meteorological factors. Ecotoxicol Environ Saf. 2025;290:117751.
[61] Kaewhan S, Junpha J, Pimpeach W. The regional distribution of dengue fever in Thailand and other emerging countries in Southeast Asia: a literature review. Toxicol Environ Health Sci. 2025;17:327-34.
[62] Liu M, Zhang Y. Impact of climate change on dengue fever: a bibliometric analysis. Geospat Health. 2025;20:1301.
[63] Sapbamrer P, Chaichanan P, Chiablam S, Pimonsree S, Thongtip S. Association of meteorology and air quality with dengue fever incidence in Upper Northern Thailand. EnvironmentAsia. 2025;18(1):164-73.
[64] Mailepessov D, Ong J, Aik J. Influence of air pollution and climate variability on dengue in Singapore: a time-series analysis. Sci Rep. 2025;15:13467.
[65] Lu HC, Lin FY, Huang YH, Kao YT, Loh EW. Role of air pollutants in dengue fever incidence: evidence from two southern cities in Taiwan. Pathog Global Health. 2023;117(6):596-604.
[66] Kuo CY, Yang WW, Su ECY. Improving dengue fever predictions in Taiwan based on feature selection and random forests. BMC Infect Dis. 2024;24:334.
[67] Ju X, Zhang W, Yimaer W, Lu J, Xiao J, Qu Y, et al. How air pollution altered the association of meteorological exposures and the incidence of dengue fever. Environ Res Lett. 2022;17:124041.
[68] Massad E, Coutinho FAB, Ma S, Burattini MN. A hypothesis for the 2007 dengue outbreak in Singapore. Epidemiol Infect. 2010;138:951-7.

[69] Cano-Granda DV, Ramírez-Ramírez M, Gómez DM, Hernandez JC. Effects of particulate matter on endothelial, epithelial and immune system cells. Rev Bionatura. 2022;7(1):4.

[70] Andrade AC, Falcão LAD, Borges MAZ, Leite ME, Espírito Santo MMD. Are land use and cover changes and socioeconomic factors associated with the occurrence of dengue fever? a case study in Minas Gerais State, Brazil. Resources. 2024;13(3):38.