Using machine learning models for near-term forecasting of dengue cases in Quezon City, Philippines

Andrea Nicole G. Senson[1], Rafael Ma. Vicente D. Tagulao[1], Francis Ira S. Velaque[1], Rolando V. Cruz[2], Wilson J. Villaganes, Jr.[2], Johnedel N. Mendoza[2], Elvira P. De Lara-Tuprio[1], Timothy Robin Y. Teng*[1], and Mark Anthony C. Tolentino[1]

[1]Department of Mathematics, Ateneo de Manila University, Loyola Heights, 1108 Quezon City, Philippines
[2]Quezon City Epidemiology and Surveillance Division, Mayaman Street corner Kalayaan Avenue, 1101 Quezon City, Philippines

ABSTRACT

Dengue is a vector-borne illness that has been a long-standing health concern in the Philippines. The disease is endemic in all regions of the country, especially in highly urbanized areas like the National Capital Region (NCR), and particularly in Quezon City (QC). The local government of QC, through the Quezon City Epidemiology and Surveillance Division (QCESD), uses a real-time monitoring system that tracks epidemic levels, reports cases by barangay, and identifies high-risk areas requiring public health interventions. The QCESD continues to enhance its system by finding ways to generate actionable insights from the epidemiological data collected. This study aims to contribute to this effort by supplementing surveillance with dengue forecasting tools. Specifically, machine learning models capable of predicting dengue incidence using epidemiological, meteorological, environmental, and socioeconomic data were implemented. Different machine learning models were explored and their performance assessed in terms of predictive accuracy and robustness. Aside from case forecasts, the factors with the greatest influence on predicted case trajectories were also identified.

INTRODUCTION

Dengue fever is a vector-borne infection caused by any of four flaviviruses. Transmission to humans occurs when they are bitten by infected female mosquitoes within the genus Aedes, primarily Aedes aegypti. Most people who get dengue are either asymptomatic, or develop flu-like symptoms such as high fever, body aches and nausea. While many recover within 2 weeks, others may develop severe symptoms which would require hospitalization. Some of the most severe dengue cases may also lead to death (WHO 2024; St. Georgiev 2009).

Dengue is now endemic in more than one hundred countries (WHO 2024), with Southeast Asia being among the seriously affected. Studies have shown that climate change, along with socio-environmental factors such as population growth and mobility, pollution, and environmental degradation, has significantly influenced the patterns and intensity of dengue transmission (Colón-González et al. 2013; Rodrigues et al. 2015). Changing temperature and precipitation patterns, as well as densely populated environments, create conditions favorable to the spread of dengue-carrying mosquitoes. The impact of environmental variables, particularly air pollution, on predicting dengue incidences has also drawn significant attention (Lu et al. 2023).

In the Philippines, dengue is a major public health problem and is endemic in all regions of the country (Undurraga et al. 2017), especially in highly urbanized areas like Quezon City (QC), the most populous city in the National Capital Region (NCR). In recent years, dengue incidence in NCR has increased significantly, with cases from January to October 2024 surging by 34% compared to the same period in the previous year. Quezon City accounted for 6,208 cases, or 26% of the region's total (Montemayor 2024).

The Quezon City Epidemiology and Surveillance Division (QCESD) closely monitors dengue cases and other notifiable diseases in the city. Its dashboard serves as a real-time monitoring

*Corresponding author
Email Address: **tteng@ateneo.edu**
Date received: 15 August 2025
Date revised: 15 April 2026
Date accepted: 04 May 2026
DOI: https://doi.org/10.54645/202619SupZDP-53

KEYWORDS

dengue, forecasting, Quezon City, Philippines, machine learning, meteorological factors

system that enables users to track epidemic levels, view reported cases by barangay, and identify high-risk areas requiring public health interventions (see Figure 1). The system is supported by the Early Warning Surveillance Unit (EWSU) and Quick Response Team (QRT), which verify reported events, conduct epidemiological investigations, and respond to potential outbreaks.

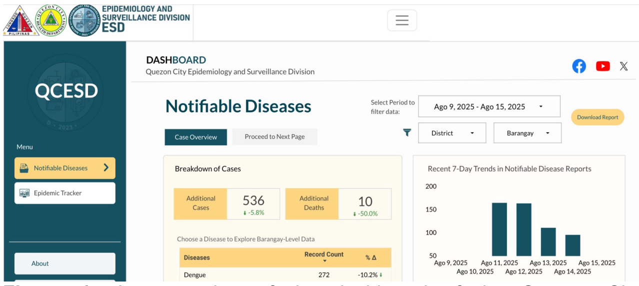


Figure 1: A screenshot of the dashboard of the Quezon City Epidemiology and Surveillance Division.

A dashboard developed by a local government unit (LGU) like QC is a valuable tool for real-time disease tracking and response, and could be enhanced by accompanying disease forecasting tools. Combining real-time surveillance with predictive analytics enables health authorities to take a more proactive approach to disease prevention and public health response. Thus, this study aims to establish the potential of machine learning tools that can supplement the real-time monitoring system of QCESD with near-term forecasts of dengue incidence.

In recent years, there has been an increase in studies on detecting and forecasting dengue outbreaks, largely driven by advances in data science tools such as machine learning, as well as greater access to weather, environmental, and epidemiological data. Researchers in Southeast Asian and Latin American countries are particularly active in these dengue studies since these regions experience frequent and widespread dengue outbreaks, which lead to significant public health challenges (Zhao et al. 2020; Tuan 2024; Al Mobin 2024; Yavari Nejad and Varathan 2021; Da Silva et al. 2025). In the Philippines, for instance, some studies employed statistical and machine learning tools in forecasting dengue cases using weather factors such as temperature, relative humidity, and precipitation, as well as data on previous dengue cases (Carvajal et al. 2018; Ligue and Ligue 2022).

A wide range of prediction models is currently in use globally, with machine learning (ML) algorithms receiving considerable attention for developing predictive models (Leung et al. 2023; Andrade Girón et al. 2025; Thirugnanam and Hussain 2023). Traditional methods, which often rely on statistical models and expert knowledge, utilize historical trends and predefined relationships to generate forecasts. While these statistical approaches can effectively capture temporal patterns, they often struggle to model the dynamic relationships between meteorological, environmental, and socioeconomic factors that influence dengue transmission. As a result, their predictive accuracy and adaptability in dengue forecasting remain limited in practice (Co et al. 2017; Carvajal et al. 2018). In contrast, ML tools are able to capture these relationships and uncover hidden patterns that traditional approaches might miss, thereby improving the effectiveness of dengue forecasting.

Machine learning models that have been employed in dengue forecasting include Random Forests, Support Vector Machines (SVM), Long Short-Term Memory (LSTM), and other artificial neural network models (Kuo et al. 2024; Tian et al. 2024, Yavari Nejad and Varathan 2021; Chen et al. 2025). In addition, ensemble models, which aggregate outputs from multiple algorithms, have been shown to perform better than individual models. For instance, Ong et al. (2023) showed that XGBoost (eXtreme Gradient Boosting), AdaBoost (Adaptive Boosting), and Random Forests generally performed better than the other ML algorithms (logistic regression, decision tree, SVM, Naïve Bayes) in predicting dengue outbreaks. A study by Sebastianelli et al. (2024) used an ensemble model consisting of CatBoost model, SVM, and LSTM with results fused together by inputting their outputs to a Random Forest model.

Given the preceding discussion and in line with the objective of establishing the potential of ML models for generating near-term forecasts of dengue incidence for QC, in this study we implemented different ML models that were calibrated to QC data and compared in terms of their predictive accuracy. Specifically, models capable of predicting dengue incidence using epidemiological, meteorological, environmental, and socioeconomic data were explored. Aside from case forecasts, the contributions of each factor or input were also analyzed, and those with the highest contributions to case forecasts were identified. Thus, the results of the study can contribute to the growing literature on the use of ML tools particularly within the public health ecosystem, focusing on dengue in the local context of QC.

This paper is organized as follows: the second section presents the data collected, their sources, and how they have been preprocessed prior to model building. The section also presents a brief description of the methodology and the machine learning models implemented. The third section presents the results as well as the corresponding interpretations and insights. The last section concludes the paper with a summary of the method and results.

MATERIALS AND METHODS

Framework

To achieve our objective, it was important to implement a range of ML models so that their suitability and predictive accuracy for near-term dengue forecasting in QC could be assessed. For this purpose, we adapted the framework and selection of ML models used by Guarneros-Nolasco et al. (2021). The entire process, from data gathering to model building and interpretation, is summarized in Figure 2. Moreover, the nine ML models we implemented and evaluated are indicated in the same figure. In alignment with the operational structure of QCESD, the ML models were implemented at the district (see Figure 3) level. In practice, this level of spatial granularity allows for more targeted planning and response against dengue and is compatible with the available datasets needed for the models.

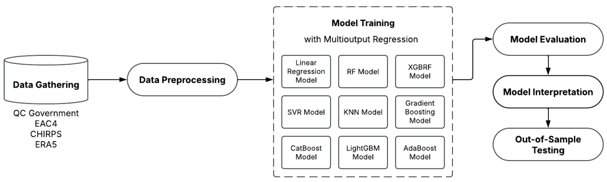


Figure 2: Overview of the Machine Learning pipeline.

Data Gathering and Preprocessing

Data collected and preprocessed prior to model building consist of epidemiological data, socioeconomic data, meteorological data, and environmental data.

Epidemiological data

Historical data of dengue cases in Quezon City from 2010 to 2024 were obtained from QCESD. The dataset was composed of weekly case counts from 142 barangays. Each of these barangays belongs to one of the six districts in QC (see Figure 3). To perform analysis and forecasting at the district level, data from barangays belonging to each district were aggregated. This necessitated barangay-to-district mapping, which was made possible through a fuzzy matching approach using the Levenshtein distance to overcome errors in spelling, abbreviations, and formatting inconsistencies.

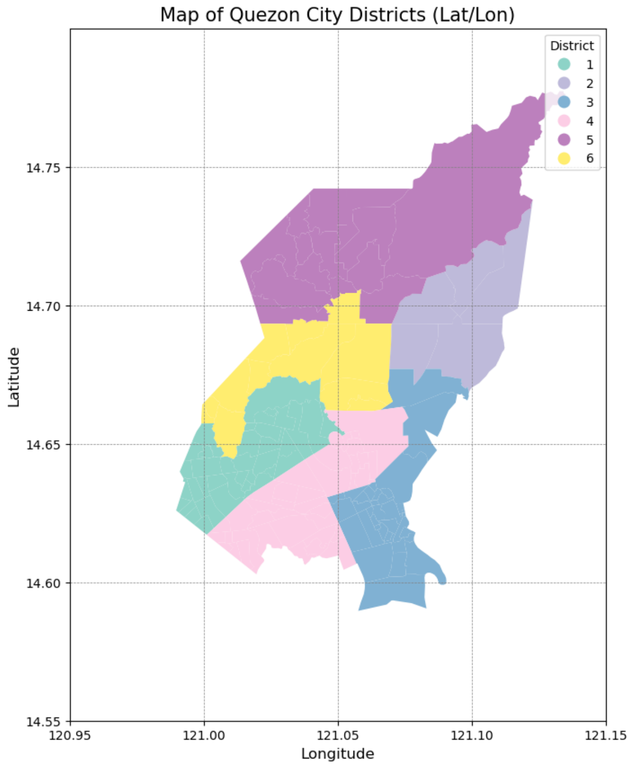


Figure 3: Map showing the six districts of Quezon City.

Population density
Population density is the only socioeconomic indicator used in this study. This data was also provided by the QCESD in the form of shapefiles that contained annual population density measurements from 2010 to 2023. The data was aggregated at the district level, which aligned with the required spatial resolution.

Rainfall
Historical data of rainfall in QC was obtained from the Climate Hazards Group InfraRed Precipitation with Station (CHIRPS, https://www.chc.ucsb.edu/data/chirps3). CHIRPS provides satellite- and station-blended precipitation estimates daily with a 0.05° × 0.05° spatial resolution. To estimate district-level precipitation from the CHIRPS gridded dataset, a spatial aggregation method was implemented. This was done by overlaying district boundaries on the rainfall grid, and assigning weights based on the fraction of the area covered by each district on a grid cell. Daily rainfall for each district was then obtained by summing the precipitation values multiplied by their corresponding weights. Mathematically, we express daily rainfall for district d at time t by $R_{d,t} = \sum_g [w_{d,g} \cdot P_{g,t}]$, where $P_{g,t}$ is the amount of rainfall on grid g at time t, and $w_{d,g}$ is the ratio of the area covered by district d on grid g to the area of grid g. Weekly data were obtained afterwards by taking the sum of daily rainfall readings.

Temperature and relative humidity
Temperature and relative humidity data were obtained from ERA5, developed by the European Centre of Medium-range Weather Forecasts or ECMWF (https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels? tab=overview). ERA5 has hourly estimates given at a 0.25° × 0.25° spatial resolution. Nearest-neighbor interpolation was used to aggregate gridded data to obtain district-level temperature and relative humidity estimates. Minimum, maximum, and mean temperatures per week, as well as average relative humidity per week, were then calculated per district.

Environmental data
Environmental data used in this study include particulate matter, specifically PM_{10} and $PM_{2.5}$, as well as primary air pollutants, including ozone (O_3), sulfur dioxide (SO_2), nitrogen dioxide (NO_2), and carbon monoxide (CO). These are converted from mass fractions (kg/kg) to parts per billion by volume (ppbv) using the formula presented in Tonga (2024). Data were obtained from the ECMWF Atmospheric Composition Reanalysis 4 (EAC4) through the Copernicus Atmosphere Monitoring Service (CAMS).

Machine Learning Models
We developed four forecasting ML models, namely n-week ahead models, where $n = 1, 2, 3, 4$, for each of the six districts of Quezon City. For each week t, the n-week ahead model predicts the dengue cases for week $t + n$. As depicted in Figure 4, the inputs to the n-week ahead model correspond to the values, for weeks t, $t - 1$, …, $t - 8$, of the following 13 variables: dengue cases, precipitation, minimum temperature, mean temperature, maximum temperature, relative humidity, PM_{10}, $PM_{2.5}$, ozone, carbon monoxide, sulfur dioxide, nitrogen dioxide, and population density. The choice of the maximum lag of 8 weeks for the input variables is consistent with the one used for a study on local forecasting of dengue cases in Singapore (Chen et al. 2018). In summary, the n-week ahead model has 117 input variables with the output variable corresponding to predicted dengue cases n weeks from the current time.

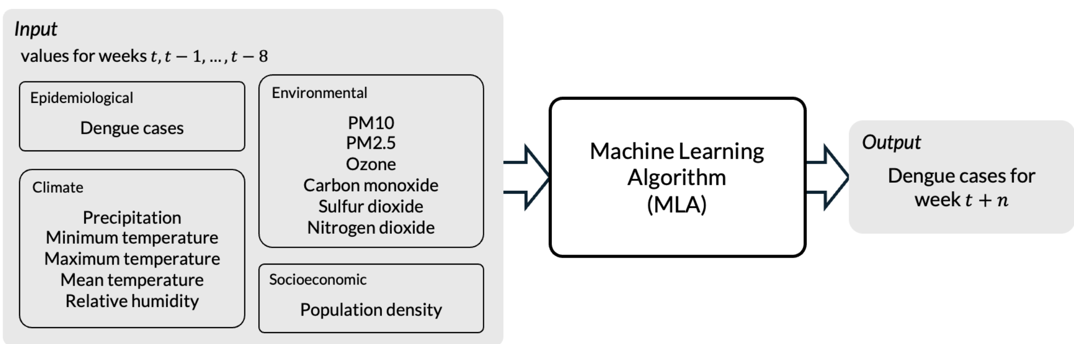


Figure 4: The n-week ahead forecasting ML model developed for each of the six districts of Quezon City and for each n = 1, 2, 3, 4.

To guide model selection for each district (as discussed in the next subsection), nine different machine learning algorithms (MLAs) were independently used to train each n-week ahead model. Following the approach of Guarneros-Nolasco et al. (2021), the nine MLAs used were: Linear Regression, Random Forest, XGBoost Random Forest (XGBRF), Support Vector Regression (SVR), K-Nearest Neighbors Regressor (KNN), Gradient Boosting, LightGBM, CatBoost, and AdaBoost. Among these, Random Forest and XGBRF are tree-based methods while Gradient Boosting, LightGBM, CatBoost, and AdaBoost are ensemble methods. A review of these algorithms can also be found in the study of Guarneros-Nolasco et. al. (2021), which shows the different advantages and limitations of these MLAs. Trying out this relatively large variety of MLAs allowed for the finding and selection of the most appropriate MLA for each Quezon City district.

Given that each of the six Quezon City districts has four forecasting models, with each one to be trained using nine different MLAs, the study employed MultiOutputRegressor from the scikit-learn Python package for better efficiency (scikit-learn Developers 2024). Instead of training models to predict the 1-week ahead, 2-week ahead, 3-week ahead and 4-week ahead dengue cases individually, the MultiOutputRegressor predicts these multiple targets simultaneously, fitting one regressor per target. Each model is trained independently on the same input features but different target variables.

Among the nine MLAs, three (Linear Regression, SVR, and KNN) were implemented with their default scikit-learn parameters. For models with stochastic components, reproducibility was prioritized by fixing random states. Random Forest, XGBRF, Gradient Boosting, and AdaBoost were all configured with `random_state=42` to ensure consistent results across multiple executions. The LightGBM implementation included additional parameters beyond the random state: `verbose=-1` to suppress training output messages, `deterministic=True` to ensure consistent results even in parallel computing environments, and `force_col_wise=True` to enforce column-wise feature splitting. Similarly, the CatBoost Regressor was parameterized with `silent=True` to minimize output during model training, `random_seed=42` to control all random operations, and `thread_count=1` to ensure deterministic behavior by restricting parallel computing that might introduce variability.

Training, Evaluation, and MLA Selection

A training-testing paradigm was then applied to the forecasting models. For the training phase, the dataset used consisted of dengue cases data over the 11-year period from January 2010 to December 2021. On the other hand, for the testing phase, the dataset used also consisted of dengue cases data but for the one-year period from January to December 2022. It should be noted, however, that the training dataset had missing entries corresponding to a significant number of weeks in 2012 and 2014. Consequently, the data entries for the four weeks prior and the eight weeks following the missing periods also had to be excluded to maintain integrity of the data. In total, each district had 562 valid weeks, and the training and testing datasets corresponded to approximately 90% and 10% of the entire dataset, respectively. The MinMaxScaler function from scikit-learn was also applied to scale both the training and testing datasets.

Hyperparameter tuning was subsequently performed using a randomized search approach. It should be noted that we did not pursue cross-validation for this procedure, and we used a strict temporal train-test split instead. This is because the correlation between the train and test phases may become unreasonable when cross-validation techniques are applied to time series data. Consequently, this could lead to poor estimates of generalization error (Cerqueira et al. 2020). Moreover, while time-series-aware alternatives such as scikit-learn's TimeSeriesSplit preserve temporal ordering, they require equally spaced, contiguous observations across splits (scikit-learn Developers 2024). This requirement is violated in our dataset, where valid observations are non-contiguous due to the removal of weeks surrounding missing case report periods in 2012 and 2014. Hence, any folding scheme that is vulnerable to misleading fold boundaries risks undermining the integrity that the approach is intended to ensure.

Refer to Table 1 for the search spaces utilized for each hyperparameter across the nine machine learning algorithms.

Table 1: Hyperparameter description and search spaces.

Model	Hyperparameter Description	Search Space
Linear Regression	N/A	N/A
Random Forest	Number of trees in the forest	100, 200, 300
	Maximum depth of each tree	10, 20, None
	Minimum samples required to split a node	2, 5
XGBRF	Number of trees to be built	100, 200, 300
	Maximum depth of each tree	3, 6, 10
	Step size shrinkage rate to prevent overfitting	0.01, 0.05, 0.1
SVR	Regularization parameter; controls trade-off between margin and misclassification	0.1, 1, 10
	Width of the insensitive margin around predictions	0.01, 0.1, 0.5
KNN	Number of nearest neighbors to consider	3, 5, 7, 10

	Weighting function for neighbors; uniform gives equal weight, distance weighs closer neighbors more	uniform, distance
Gradient Boosting	Number of boosting stages	100, 200, 300
	Maximum depth of each tree	3, 5, 8
	Shrinkage factor applied to each tree's contribution	0.01, 0.05, 0.1
LightGBM	Number of boosting iterations	100, 200, 300
	Maximum depth of each tree; -1 means no limit	5, 10, -1
	Shrinkage factor applied to each tree's contribution	0.01, 0.05, 0.1
CatBoost	Number of boosting iterations	300, 500, 700
	Maximum depth of each tree	6, 8, 10
	Shrinkage factor applied to each tree's contribution	0.01, 0.05, 0.1
AdaBoost	Number of weak learners to combine	50, 100, 200
	Weight applied to each classifier's contribution at each iteration	0.01, 0.1, 0.5, 1.0

To assess the performance of the forecasting models under the different MLAs, eight different evaluation metrics were generated. The first two were Mean Squared Error (MSE) and coefficient of determination (R^2), which are both widely used measures (Tatachar 2021). The remaining six metrics were error measures adopted from the evaluation framework proposed by Tabataba et al. (2017): Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), Symmetric Mean Absolute Percentage Error (sMAPE), Median Absolute Percentage Error (MdAPE), and Median Symmetric Absolute Percentage Error (MdsAPE). Recall that R^2 values usually fall between 0 and 1, with higher values suggesting better model performance. On the other hand, for the other seven metrics, which are error-based, lower values indicate better model performance.

Model Interpretation and Out-of-sample Testing

This next portion of the methodology focuses on the outputs of the forecasting models equipped with the best MLA identified for each of the six Quezon City districts. First, outputs during the testing period (2022) were subjected to further analysis. Secondly, in line with one of the work's objectives, a SHapley Additive Explanations (SHAP) analysis (over the training period 2010-2021) was also conducted to understand the influence of different features on the models' predictions. This method (Merrick and Taly 2020) provides a way of quantifying each feature's contribution to the models' outputs; in turn, this offers insights into the factors driving predictions across different districts. More specifically, the SHAP score of each feature at each time step may reflect a 'positive effect' (i.e., the feature is associated with an increase in the particular target variable) or a 'negative effect' (i.e., the feature is associated with a decrease in the value of the target variable).

The same forecasting models (i.e., those using the best MLA for each district) then underwent out-of-sample testing. Recall that

For each of the six Quezon City districts and each of the nine MLAs, only one set of evaluation metrics was computed over the testing period. This meant that the metrics were computed so that the multitarget outputs (i.e., one-week ahead to four-week ahead predictions) from MultiOutputRegressor were all appropriately processed and incorporated. The nine MLAs for each district were then ranked based on their performance with respect to the different evaluation metrics. To synthesize the performance results across the multiple error measures, a Consensus Ranking approach as described by Wang and Wang (2001) was then adopted. Through this approach, rankings were aggregated by averaging evaluation ranks to produce a balanced assessment of the different MLAs. As each MLA would have a single Consensus Ranking score for each district, the best MLA for each district could be selected for further analysis, interpretation, and out-of-sample testing of the corresponding forecasting models.

these models were trained and tested using 2010-2021 and 2022 datasets, respectively. These models, with no additional training or calibration, were then used to predict 1-week to 4-week ahead dengue cases (recall Figure 4) for each week of the year 2023. These predictions were then evaluated by comparing them to actual data from that year.

RESULTS AND DISCUSSION

Evaluation and Selection of Models

The evaluation metrics were derived for each of the nine ML models across six districts over the test period. The corresponding metrics for the best- and worst-performing ML models are tallied in Table 2. It can be observed that tree-based/ensemble models (Random Forest, LightGBM, XGBRF, and Gradient Boosting) outperformed other models across all districts, suggesting greater reliability in capturing dengue case trends. On the other hand, simpler models such as KNN and Linear Regression tend to exhibit poorer performance, as indicated by higher error magnitudes and negative R^2 scores. Focusing on MAE, RMSE and MSE, we see that the best- performing models for Districts 1, 3, 4, 5 and 6 had comparable error values. In particular, the MAE values for the aforementioned districts were within (4.7, 5.4), while the corresponding RMSE values were within (6.4, 7.0). On the other hand, the best-performing model for District 2 recorded the highest level of errors, with MSE and RMSE of 84.96 and 9.22, respectively.

MAPE and sMAPE values, which measure prediction errors as a proportion involving actual values, were found to be lowest in District 1. Districts 2, 4, 5, and 6 have comparable MAPE values, while District 3 has a relatively higher MAPE value of 84.44%. Similarly, Districts 2, 4, 6 have comparable sMAPE values, while Districts 3 and 5 have relatively higher sMAPE values at around 50%. The MdAPE values were consistently lower than MAPE for all districts, with most values not exceeding 35% (other than District 3). This implies that extreme prediction errors occur less frequently for the best-performing models in these districts. On the other hand, the best-performing model for District 3 has the highest MdAPE value of 44.24%, while also having the second highest sMAPE value. These indicate that the models for District 3 struggle to capture actual case trends. Overall, these results illustrate that the forecasting models' predictive ability can vary across the different districts.

Finally, the highest R^2 scores were attained by the best models for Districts 1, 4, and 6 at 0.73, 0.69, and 0.69, respectively. These scores reveal the models' capability to explain the variations in disease incidence data. Slightly lower but comparable R^2 values were recorded for Districts 2, 3, and 5 at 0.61 0.51, and 0.65, respectively. Meanwhile, the lowest R^2 scores were recorded in District 5, where the worst-performing models achieved highly negative values; most notably, Linear Regression recorded an R^2 of -3.20 and a MAPE of 267.64%.

Table 2: Performance metrics of top and worst models per district.

District	Model	MAE	MSE	RMSE	MAPE	MdAPE	sMAPE	MdsAPE	R^2
1	Random Forest	4.72	47.06	6.86	35.86%	25.68%	32.79%	27.87%	0.73
	KNN	14.25	377.92	19.44	68.6%	74.77%	106.28%	111.6%	-1.16
2	LightGBM	6.00	84.96	9.22	49.93%	27.55%	37.32%	28.87%	0.61
	KNN	14.56	402.13	20.05	95.05%	81.23%	107.08%	112.04%	-0.86
3	XGBRF	5.09	45.10	6.72	84.44%	44.24%	49.90%	42.49%	0.51
	KNN	7.30	92.63	9.62	119.40%	58.17%	65.96%	61.15%	-0.002
4	XGBRF	5.35	47.54	6.90	48.94%	28.97%	36.75%	28.08%	0.69
	KNN	13.49	317.71	17.82	65.10%	69.11%	99.8%	96.25%	-1.05
5	Gradient Boosting	5.02	44.44	6.67	51.64%	34.83%	53.13%	37.21%	0.65
	Linear Regression	18.06	527.48	22.97	267.64%	94.30%	126.66%	154.37%	-3.20
6	LightGBM	4.71	42.17	6.49	56.92%	24.78%	34.43%	22.95%	0.69
	KNN	9.64	170.93	13.07	74.42%	62.23%	71.53%	70.13%	-0.26

Table 3: Model Performance on the Train and Test Set Across Districts.

	MAE		R^2	
District	Train	Test	Train	Test
1	3.24	4.72	0.9581	0.7309
2	7.30	6.00	0.7592	0.6060
3	2.34	5.09	0.9734	0.5110
4	7.99	5.35	0.8049	0.6924
5	3.31	5.02	0.9915	0.6515
6	6.91	4.71	0.7028	0.6876

The best models for each district were subsequently analyzed for overfitting. This was done by comparing MAE and R^2 scores for training and test datasets as presented in Table 3. It can be seen that similar MAE values were obtained for both the training and testing period across all districts, with the greatest discrepancy being observed in District 3. In relation to goodness-of-fit, higher R^2 scores were derived from the training data, with half of the districts tallying values greater than 0.9. On the other hand, the R^2 scores derived from test data were expectedly lower across all districts, with the greatest discrepancies being observed in Districts 3 and 5. Overall, while there were a few signs of overfitting based on the computed metrics, particularly in District 3, the associated best models in most districts were able to generalize moderately well and exhibited a decent balance between training and testing performance.

Results of Forecasting Models

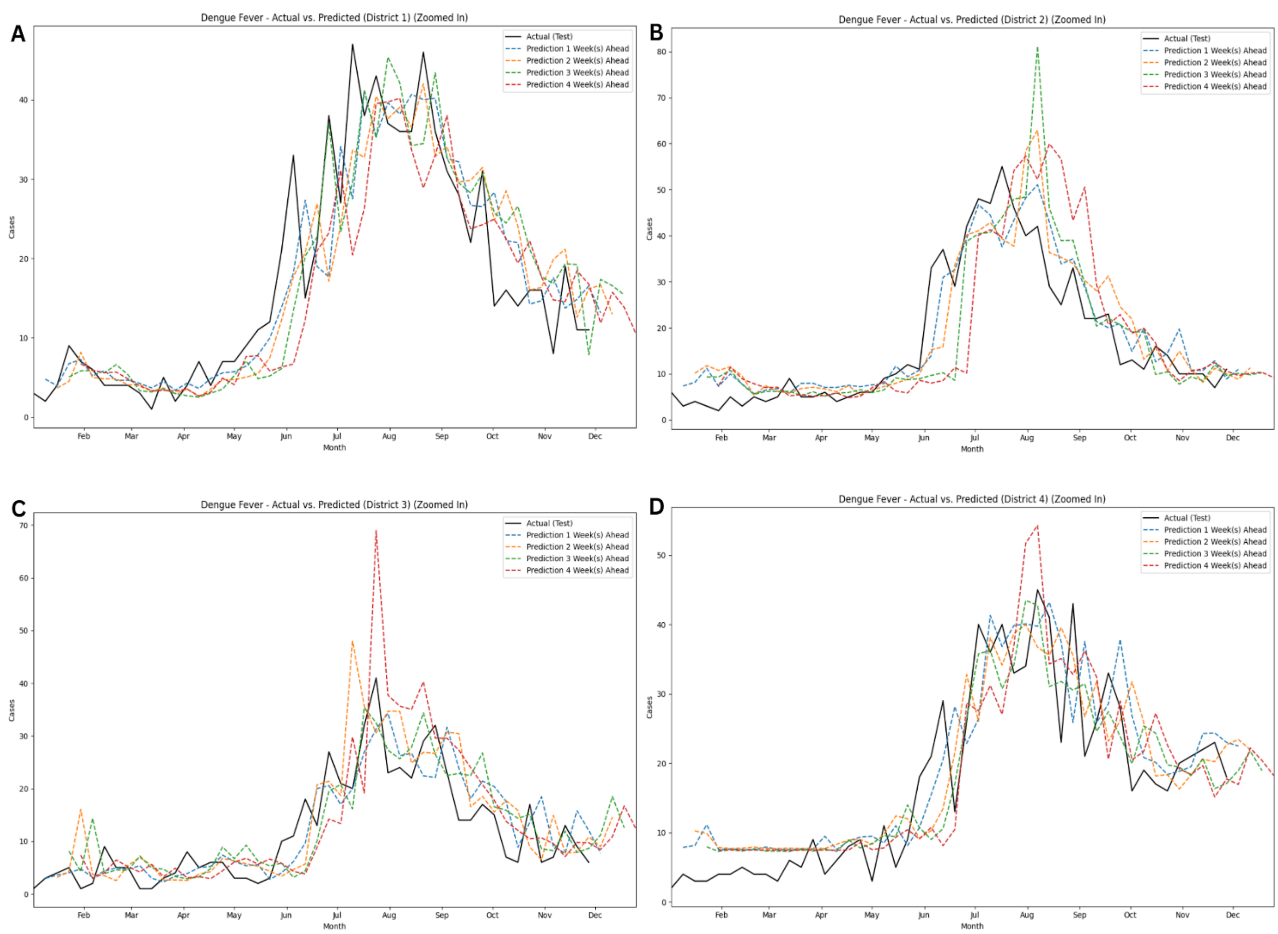
Testing Period Predictions

Based on the results in Table 2 on the evaluation metrics per district, the following top performing models were chosen for forecasting dengue incidence: the Random Forest model for District 1, the LightGBM model for Districts 2 and 6, the XGBRF model for Districts 3 and 4, and the Gradient Boosting model for District 5. We then consider n-week ahead models for each district, where $n \in \{1,2,3,4\}$. This gives us a total of twenty-four models for this part of the study. The generated model outputs, along with the actual case numbers, over the testing period (2022) are presented in Figure 5.

Comparison of actual cases and n-week ahead forecasts demonstrate that the models were able to follow general incidence patterns and variations in the test dataset for each district. The models perform well in predicting the number of cases when the trend remained stable. However, it can also be observed that prediction accuracy suffered over certain periods. For example, the predicted case numbers around June were significant underestimations across Districts 1 to 5. Significant overestimations were also observed when true case counts began to decline, particularly in Districts 2, 3, 5, and 6. In this case, the 3- and 4-week ahead models produced greater overestimations compared to the 1- and 2-week ahead forecasts. As expected, the n-week ahead models generally exhibited poorer performance as the forecast horizon was extended (i.e., n was increased).

A notable pattern in the predictions is that each n-week ahead model appears to be shifted forward in time, and is hence indicative of the delay in predicting seasonal peaks and downturns. This is typically observed in tree-based models such as Random Forest (Zhao et al. 2020) where prediction outputs tend to lag behind actual numbers, especially when there is a sudden increase or decrease in cases. Boosting models such as CatBoost and Gradient Boosting may offer improvements in predictive performance, but they are still prone to delayed responses when sudden changes in disease incidence occur.

Generally, the forecasts of 1-week ahead models for all districts had the least deviations from actual data compared to other forecast horizons. The accuracy dropped for the forecasts produced by the 2-week ahead model, with frequent overestimations observed during the peak periods in Districts 2, 3, 5, and 6. A similar pattern was also observed in the outputs of the 3-week ahead models. Lastly, the 4-week ahead models exhibited the poorest performance, showing the largest deviations across multiple periods in most districts.



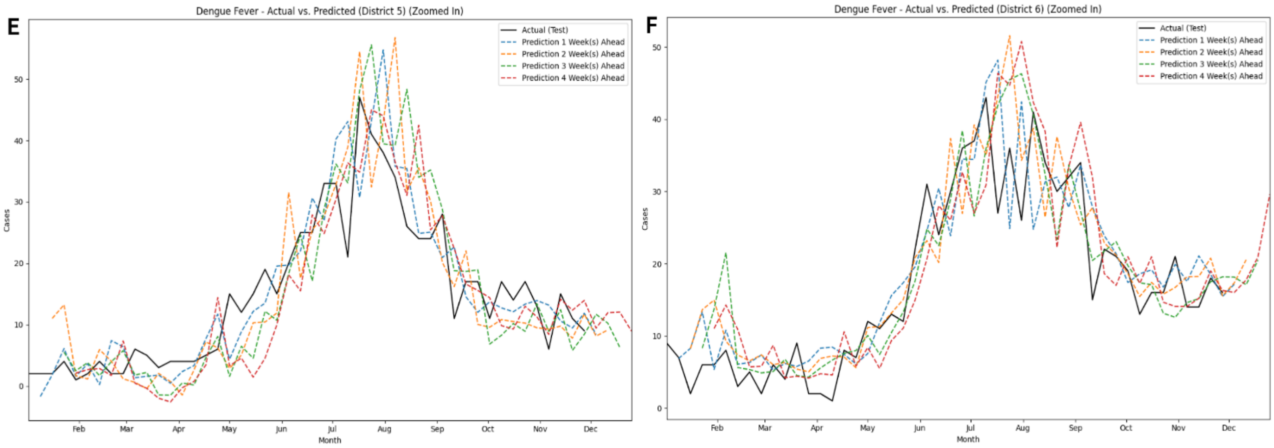


Figure 5: Weekly dengue cases on the district level with actual and predicted cases.

SHAP Heatmaps

A feature's SHAP value corresponds to its contribution to the target variable. In this study, the target is denoted by *f(x)* and pertains to the 1-week, 2-week, 3-week, and 4-week ahead predictions. All of the features' contributions, when added, make up the model's prediction. Visually, the SHAP values of all input features over the training phase (year 2010-2021) can be represented by a SHAP heatmap. The heatmaps for all districts and different forecast horizons are presented in Figures 6, 7, 8 and 9. The red bands indicate positive SHAP values, while blue bands indicate negative SHAP values. Moreover, darker colored bands correspond to SHAP values with greater magnitude. The black bands at the end of each row represent the global importance of the corresponding feature, which is based on the average of the absolute value of SHAP scores. The features are arranged in decreasing order of global importance, with the bottom row constituting the combined contributions of the lower-ranked features.

Among the input features, current and historical dengue cases have the highest global importance across heatmaps, indicating that recent incidence patterns are strong predictors of future disease counts. This likely suggests the temporal persistence commonly observed in dengue time series, where present case levels are strongly related to near-future transmission (Benedum 2020). The result highlights the value of timely surveillance data for short-term forecasting, although strong reliance on lagged incidence may also contribute to delayed responses during sudden outbreak increases or rapid declines.

Climatic factors also have an impact on the predictions. Temperature-related inputs appeared more prominently as the top features in multiple districts and across different forecast horizons. MaxTemp_lag_8 and MeanTemp_lag_8, for example, significantly contributed to the prediction outputs of the 4-week ahead models of Districts 2 and 4 (see Figures 9B and 9D). In addition, these features often showed seasonal effects, with red bands appearing periodically, as in the *n*-week ahead models of District 6 in Figures 6F, 7F, 8F and 9F. Precipitation-related features, while having relatively lower SHAP values, were seen in the top features of Districts 1 and 5. Finally, relative humidity also had an influence, but was restricted to specific sectors such as District 1, where it appeared as an important feature in all forecast horizons.

Aside from the epidemiological and meteorological features, population density (the lone socioeconomic variable) also exhibited some relevance to the predictions as it appeared as a higher ranked feature of Districts 2, 4 and 5. Particulate matter and air pollutants, on the other hand, had limited influence on the outputs, and they are mostly seen as higher ranked features in District 3.

It should be noted that while it is essential to recognize the importance of the top ranked features in incidence prediction, the collective contribution of the less important features often exceeded many of the features discussed earlier. When the forecast horizon was extended, the global importance of the weaker features became more prominent, as seen in the 3-week and 4-week ahead models in Figures 8 and 9, respectively. Hence, their relevance in disease forecasting cannot be set aside.

The SHAP results reveal several insights and implications. First, since dengue cases have the greatest contribution to case predictions, recent incidence trends may provide early warnings for upcoming outbreaks, and hence highlights the importance of real-time surveillance of disease incidence. This is thus aligned with the efforts of QCESD in sustaining and enhancing real-time monitoring of dengue cases. Second, while meteorological variables are among the top features in the various models, their contributions to case predictions are not as impactful as that of epidemiological data. Lastly, the remaining lower-ranked features form a significant contribution to case predictions, increasing in predictive strength as the forecast horizon is extended. The inclusion of such features could be helpful in generating more reliable long-term forecasts.

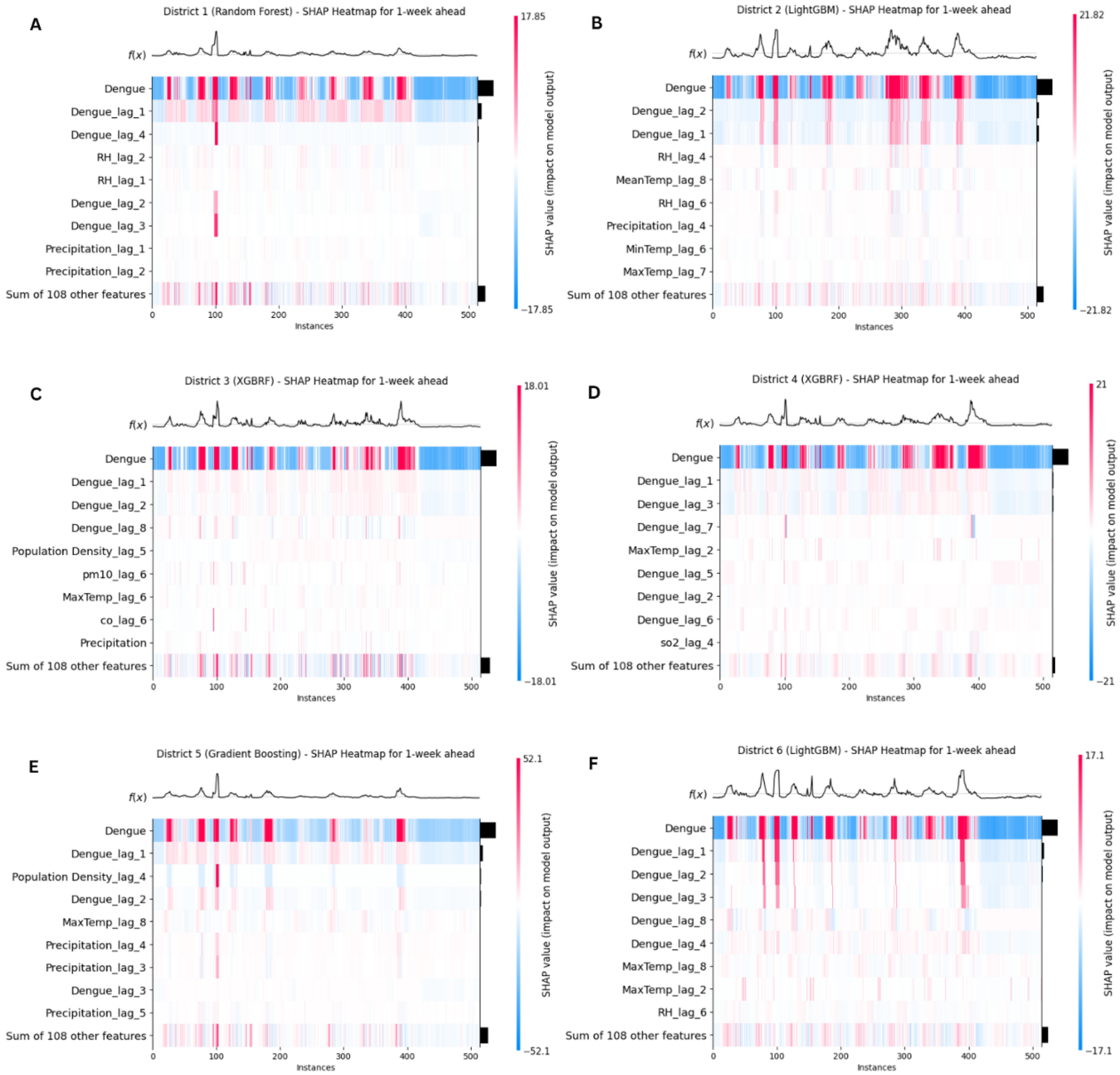


Figure 6: SHAP heatmap for the six districts, 1-week ahead dengue prediction.

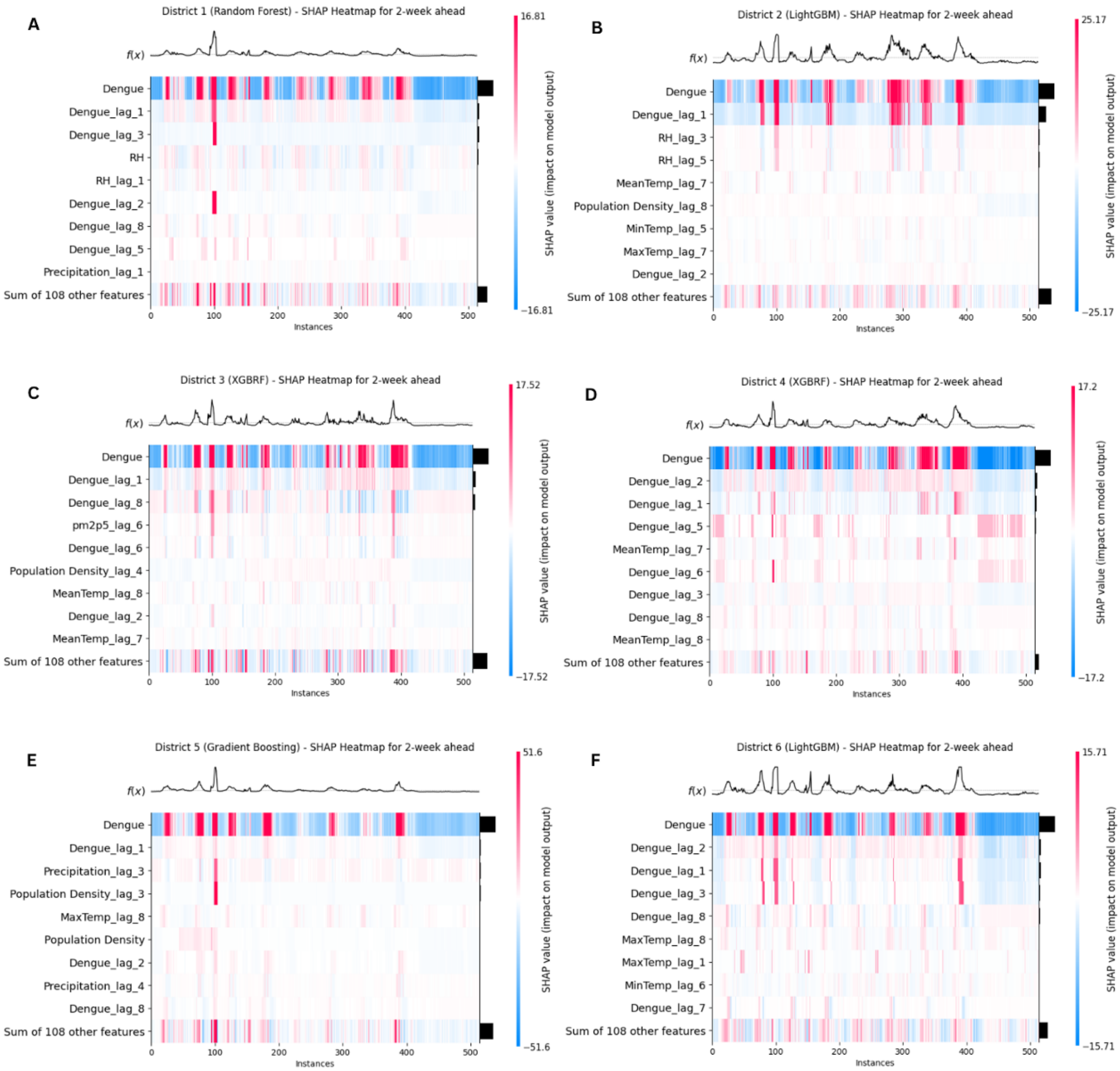


Figure 7: SHAP heatmap for the six districts, 2-week ahead dengue prediction.

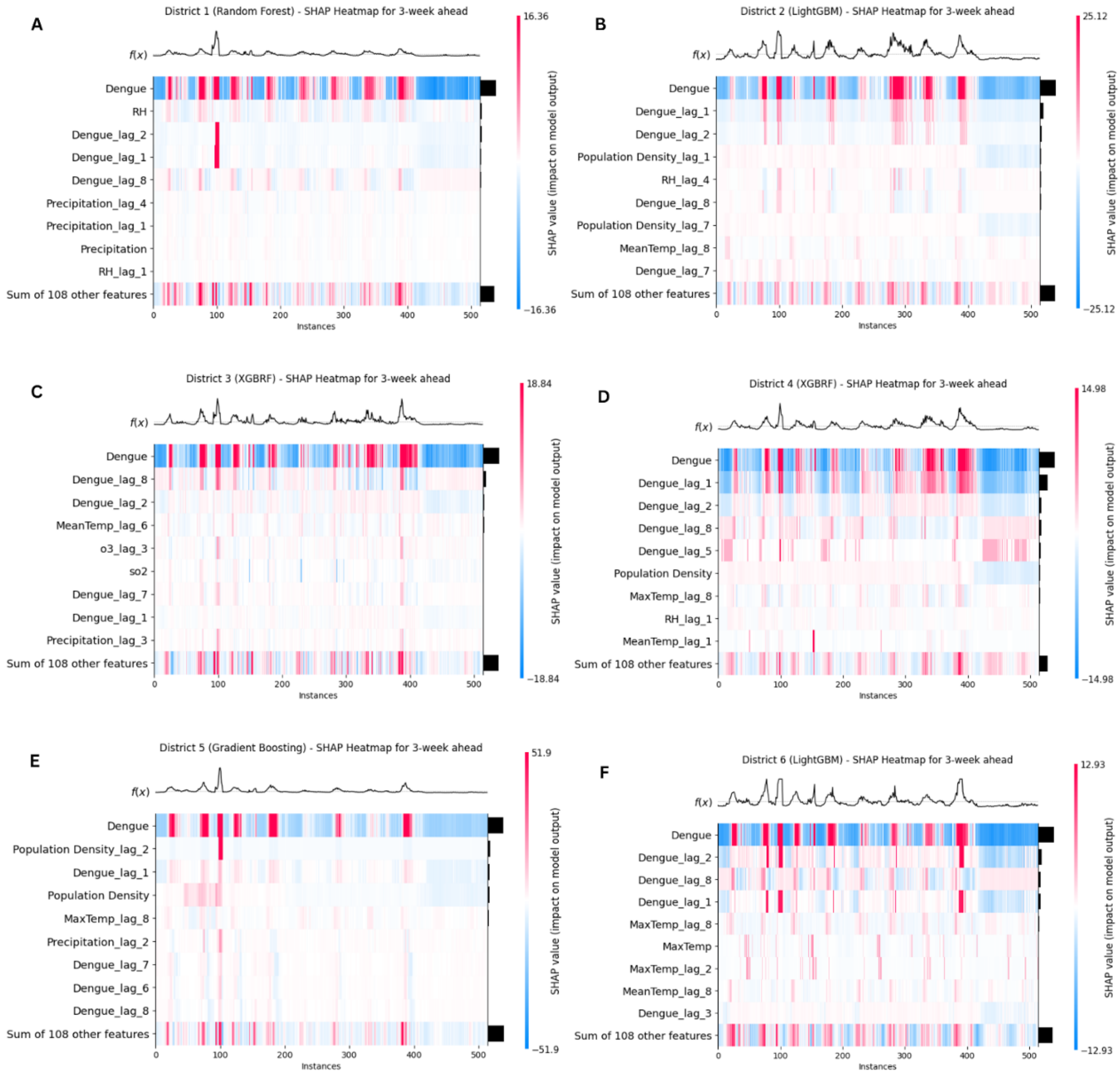


Figure 8: SHAP heatmap for the six districts, 3-week ahead dengue prediction.

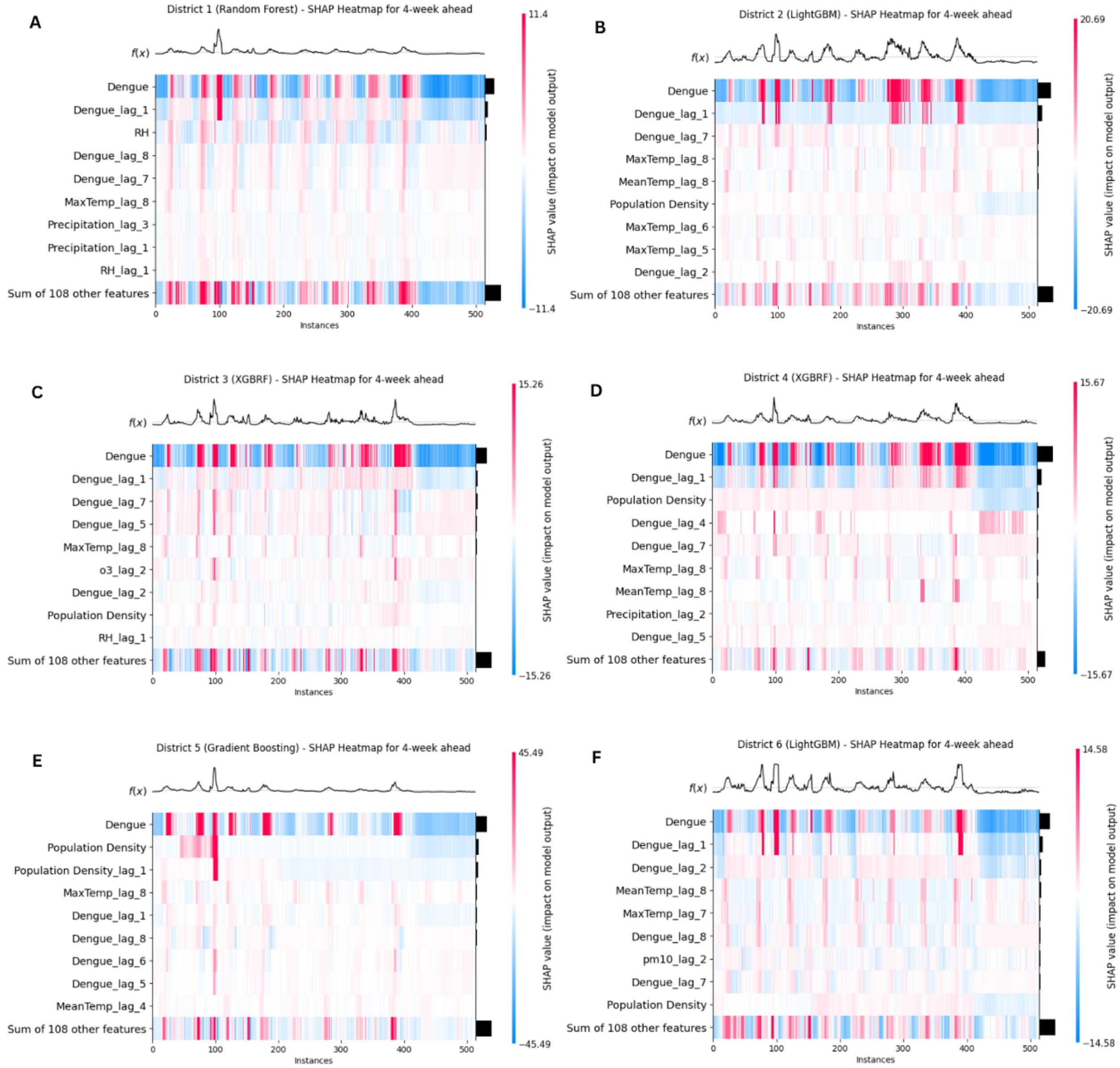


Figure 9: SHAP heatmap for the six districts, 4-week ahead dengue prediction.

Out-of-Sample Testing

Recall that 1-week ahead to 4-week ahead forecasts were generated using the framework in Figure 4, with the chosen MLAs corresponding to the ones that achieved the best performance based on the evaluation metrics in Table 2. For out-of-sample testing, such forecasts were generated in every district and for each week of 2023. The actual data and outputs for the *n*-week ahead model, labeled as Dengue_*n* where $n \in \{1,2,3,4\}$, are presented in Figure 10, with the corresponding mean absolute errors in Table 4.

Table 4: Mean Absolute Error (MAE) for out-of-sample testing.

District	Dengue_1	Dengue_2	Dengue_3	Dengue_4
1	4.98	6.27	8.04	7.47
2	6.80	7.95	7.26	7.06
3	5.33	6.63	6.30	6.44
4	6.29	7.04	5.81	6.70
5	4.33	5.11	6.31	6.97
6	5.14	6.16	6.91	9.17

The MAE outputs for all models yielded values between 4.3 and 9.2, with most of them falling below 7. Dengue_1 consistently had lower MAE values compared to all other derived error values. This shows a slight decline in performance as the forecast horizon is extended. For Dengue_1 and Dengue_2, District 5 had the lowest error values, while District 2 had the highest. Meanwhile, for Dengue_3 and Dengue_4, Districts 3, 4 and 5 had comparable model performance, with MAE values falling within 5.8 and 7. On the other hand, District 1 and 6 performed the worst, with District 6 reaching an error value of 9.17 for Dengue_4.

Observing the generated outputs, the models typically underestimated case data, in varying degrees, within May to mid-June. Overestimations generally occurred on certain periods within June to August, and November to December; this was mostly prevalent in the outputs of Districts 3, 4 and 6. The models for District 6, moreover, were consistently overestimating case numbers throughout the entire prediction period. Despite varying performances, the resulting prediction trends across different locations still follow the general trajectory of out-of-sample dengue incidence data. This was particularly evident in the 1-week ahead

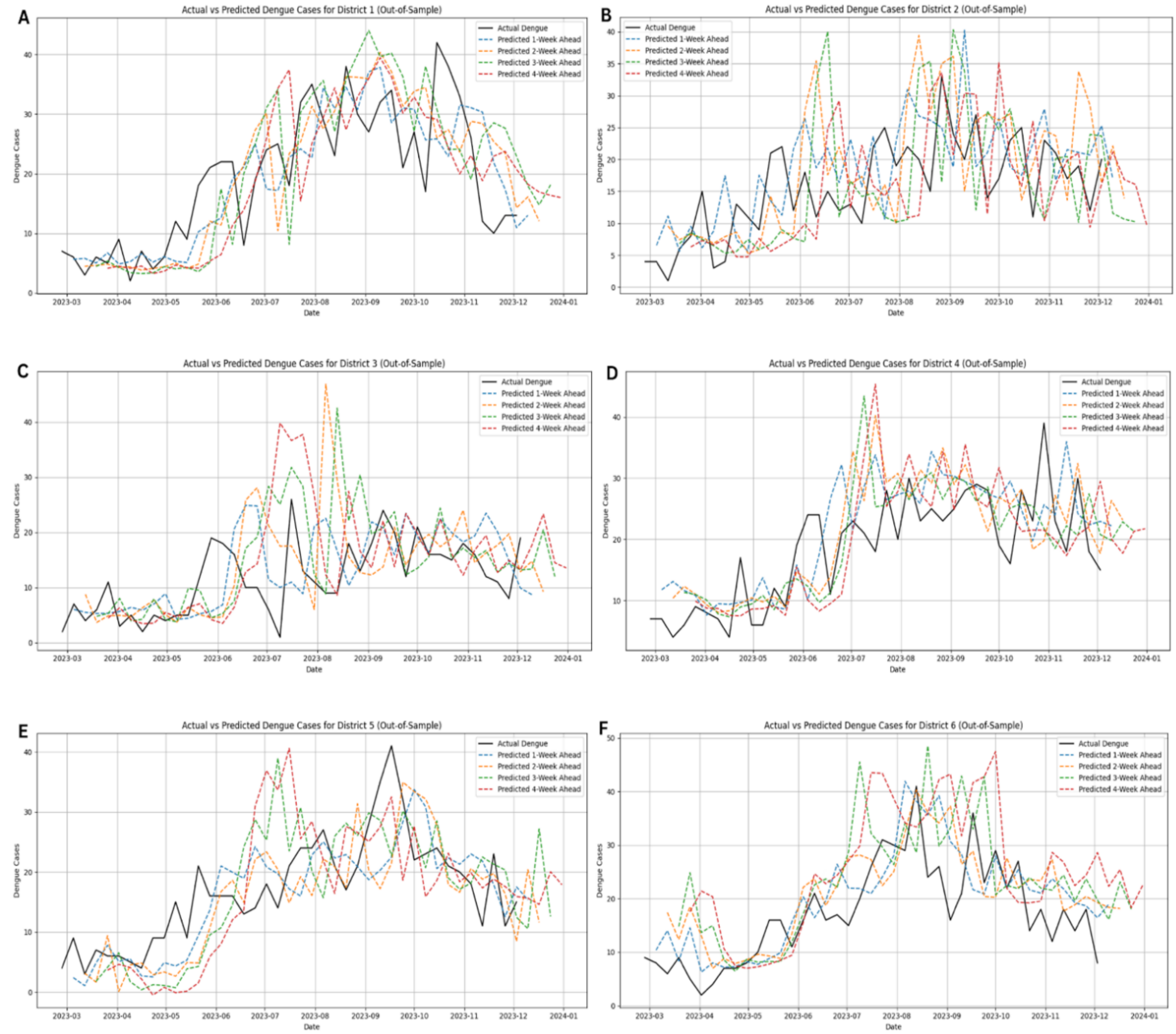


Figure 10: Dengue predictions in the out-of-sample testing for the six districts

and 2-week ahead model outputs of District 1 and 5 as seen in Figures 10A and 10E.

Overall, the results for the out-of-sample testing show that the best-performing ML models we identified through the study's framework can indeed be used for near-term dengue forecasting across the six QC districts. However, it is important to inform users of these models about the presence and extent of prediction errors, which are expected from any forecasting tool. Nevertheless, the outputs of these predictive models can enhance planning and decision-making of the QCESD when it comes to dengue mitigation and response.

CONCLUSION

Dengue remains one of the most prevalent communicable diseases in the Philippines. Quezon City, one of the country's largest cities, continues to record high case numbers, which accounted for around 25% of the National Capital Region's total cases in 2024. Thus, the Quezon City Epidemiology and Surveillance Division (QCESD) maintains a critical role in monitoring the dengue situation in the city's six districts and in implementing interventions and policies for disease mitigation and control. Essential elements of this role of the QCESD include the collection, presentation, and analysis of dengue case data, the extensive outputs of which are regularly published in the city's online disease surveillance dashboard. To enhance these efforts, the QCESD also continues to pursue new ways by which additional insights from their epidemiological data can be generated. In support of this objective, this work focused on the application of machine learning methods that could generate near-term dengue forecasts for the city.

For each of Quezon City's six districts, we implemented machine learning models that incorporate historical dengue data, meteorological variables, environmental variables, and population density as features and whose output corresponds to 1-week, 2-week, 3-week, or 4-week ahead forecasts of dengue cases. Nine different machine learning algorithms (MLAs) were then implemented to serve as forecasting models, which were then trained and tested using data on dengue cases from 2010-2021 and 2022, respectively. Using well-established evaluation metrics and a Consensus Ranking approach, the best MLA was then identified for each district. Tree-based and ensemble models outperformed all other MLAs across the six districts, with XGBRF and LightGBM typically performing the best among them. Results show that each district's models, equipped with the corresponding best MLA, can follow general incidence patterns and variations in the testing datasets and perform well in predicting the number of cases when the trend remains stable. Out-of-sample testing of the best-performing models for each district was also carried out, with results showing that while there is variability in the models' predictive accuracy, the forecasted trends could still follow the trajectory of out-of-sample dengue incidence data, illustrating that these models can indeed be used for near-term dengue forecasting across the six QC districts.

A SHapley Additive Explanations (SHAP) analysis (over the training period 2010-2021) of each district's best forecasting models was also conducted and revealed that features corresponding to historical dengue cases have the greatest contribution to case predictions. Thus, this underscores the importance of accurate epidemiological datasets in generating near-term dengue case forecasts and supports QCESD's extensive efforts for real-time monitoring of dengue cases in Quezon City. On the other hand, the SHAP analysis also shows that while the remaining features (i.e., meteorological, environmental, and socioeconomic variables) have low significance individually, their collective contribution is also comparable to that of the epidemiological features. This observation highlights the complex dynamics of dengue transmission and how some of these complexities can be unraveled using machine learning.

The study contributes to the growing literature on the use of ML tools particularly within the public health ecosystem. While the study and its results are entirely focused on QC, the framework and methodology employed can be adapted for other cities or regions to support the potential integration of ML-based predictions in disease surveillance and response.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to the reviewers for their valuable comments and suggestions.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

CONTRIBUTIONS OF INDIVIDUAL AUTHORS

ANGS, RMVDT and FISV were responsible for data gathering and preprocessing, model development, coding and implementation, analysis of results, and writing of the paper. RVC, WJV and JNM contributed to data gathering and preprocessing, analysis of results, and writing of the paper. EPDLT, TRYT and MACT coordinated with QCESD, and contributed to model development, analysis of results, and writing of the paper.

REFERENCES

Al Mobin M. Forecasting dengue in Bangladesh using meteorological variables with a novel feature selection approach. Sci Rep 2024; 14, 32073. https://doi.org/10.1038/s41598-024-83770-0.

Amazon Web Services. CatBoost Hyperparameters - AWS Documentation. March 2024. https://docs.aws.amazon.com/sagemaker/latest/dg/catboost-hyperparameters.html.

Andrade Girón DC, Marín Rodriguez WJ, Lioo-Jordan FdM, Ausejo Sánchez JL. Machine learning and deep learning models for dengue diagnosis prediction: a systematic review. Informatics 2025; 12(1):15. https://doi.org/10.3390/informatics12010015.

Benedum CM, Shea KM, Jenkins HE, Kim LY, Markuzon N. Weekly dengue forecasts in Iquitos, Peru; San Juan, Puerto Rico; and Singapore. PLOS Negl Trop Dis. 2020; 14(10):e0008710.

Carvajal TM, Viacrusis KM, Hernandez LFT et al. Machine learning methods reveal the temporal pattern of dengue incidence using meteorological factors in Metropolitan Manila, Philippines. BMC Infect Dis 2018: 18, 183. https://doi.org/10.1186/s12879-018-3066-0.

Cerqueira V, Torgo L, Mozetič I. Evaluating time series forecasting models: an empirical study on performance estimation methods. Machine Learning. 2020; 109:1997–2028.

Chen X, Moraga P. Forecasting dengue across Brazil with LSTM neural networks and SHAP-driven lagged climate and spatial effects. BMC Public Health 2025; 25, 973. https://doi.org/10.1186/s12889-025-22106-7.

Chen Y, Ong JHY, Rajarethinam J, et al. Neighbourhood level real-time forecasting of dengue cases in tropical urban Singapore. BMC Med 2018; 16:129. https://doi.org/10.1186/s12916-018-1108-5.

Co J, Tan JA et al. Dengue spread modeling in the absence of sufficient epidemiological parameters: comparison of SARIMA and SVM time series models. Proceedings of the RAIS Conference: The Future of Ethics, Education and Research 2017. http://dx.doi.org/10.2139/ssrn.3086161.

Colón-González FJ, Fezzi C, Lake IR, Hunter PR. The effects of weather and climate change on dengue. PLOS Negl Trop Dis 2013; 7(11), e2503. https://doi.org/10.1371/journal.pntd.0002503.

Da Silva, ST, Gabrick EC, Protachevicz PR et al. When climate variables improve the dengue forecasting: a machine learning approach. Eur. Phys. J. Spec. Top 2025; 234, 555–569. https://doi.org/10.1140/epjs/s11734-024-01201-7.

Guarneros-Nolasco LR, Cruz-Ramos NA, Alor-Hernández G, Rodríguez-Mazahua L, & Sánchez-Cervantes JL. Identifying the main risk factors for cardiovascular diseases prediction using machine learning algorithms. Mathematics 2021; 9(20):2537. https://doi.org/10.3390/math9202537.

Kuo CY, Yang WW, Su EC. Improving dengue fever predictions in Taiwan based on feature selection and random forests. BMC Infect Dis 2024; 24(Suppl 2):334. https://doi.org/10.1186/s12879-024-09220-4.

Leung XY, Islam RM, Adhami M, Ilic D, McDonald L, Palawaththa S, Diug B, Munshi SU, Karim MN. A systematic review of dengue outbreak prediction models: current scenario and future directions. PLOS Negl Trop Dis 2023; 17(2):e0010631. https://doi.org/10.1371/journal.pntd.0010631.

Ligue KDB, Ligue KJB. Deep learning approach to forecasting dengue cases in Davao City using long short-term memory (LSTM). Phil J Sci 2022; 151 (3): 797-812. https://doi.org/10.56899/151.03.01.

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 Glob Health. 2023; 117(6):596-604. https://doi.org/10.1080/20477724.2022.2135711.

Madewell ZJ, Rodriguez DM, Thayer MB, Rivera-Amill V, Paz-Bailey G, Adams LE, Wong JM. Machine learning for predicting severe dengue in Puerto Rico. Infect Dis Poverty 2025; 14(01):50-66. https://doi.org/10.1186/s40249-025-01273-0.

Merrick L, Taly A. The explanation game: explaining machine learning models using Shapley values. In: Holzinger A, Kieseberg P, Tjoa A, Weippl E. (eds) Machine Learning and Knowledge Extraction. CD-MAKE 2020. Lect Notes Comput Sci. Cham: Springer, 2020: 12279. https://doi.org/10.1007/978-3-030-57321-8_2.

Montemayor MT. Dengue cases in NCR up by 34.47% Jan-Oct. Philippine News Agency. 2024. https://www.pna.gov.ph/articles/1237153.

Ong SQ, Isawasan P, Ngesom AMM et al. Predicting dengue transmission rates by comparing different machine learning models with vector indices and meteorological data. Sci Rep 2023; 13, 19129. https://doi.org/10.1038/s41598-023-46342-2.

Rodrigues Md, Marques GRAM, Serpa LLN et al. Density of Aedes aegypti and Aedes albopictus and its association with number of residents and meteorological variables in the home environment of dengue endemic area, São Paulo, Brazil. Parasites Vectors 2015; 8, 115. https://doi.org/10.1186/s13071-015-0703-y.

Scikit-learn Developers. MultiOutputRegressor. 2024. Available from: https://scikit-learn.org/stable/modules/generated/sklearn.multioutput.MultiOutputRegressor.html.

Scikit-learn Developers. TimeSeriesSplit. 2024. Available from: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection. TimeSeriesSplit.html.

Sebastianelli A, Spiller D, Carmo R et al. A reproducible ensemble machine learning approach to forecast dengue outbreaks. Sci Rep 2024; 14, 3807. https://doi.org/10.1038/s41598-024-52796-9.

St. Georgiev V. Viral hemorrhagic fever: flaviviruses. National Institute of Allergy and Infectious Diseases, NIH: Volume 2: Impact on Global Health. Springer Science & Business Media. 2009: 265-271. https://doi.org/10.1007/978-1-60327-297-1_23.

Tabataba F, Chakraborty P, Ramakrishnan N, et al. A framework for evaluating epidemic forecasts. BMC Infect Dis 2017; 17:345. https://doi.org/10.1186/s12879-017-2365-1.

Tatachar AV. Comparative assessment of regression models based on model evaluation metrics. Int J Res Eng Technol 2021; 8(9):853–860.

Thirugnanam A, Hussain FBJ. Exploring machine learning algorithms for the prediction of dengue: A comprehensive review. Revue d'Intelligence Artificielle 2023; 37 (5): 1281-1290. https://doi.org/10.18280/ria.370521.

Tian N, Zheng J-X, Li L-H, Xue J-B, Xia S, Lv S, Zhou X-N. Precision prediction for dengue fever in Singapore: a machine learning approach incorporating meteorological data. Trop Med Infect Dis 2024; 9(4):72. https://doi.org/10.3390/tropicalmed9040072.

Tonga LP. CCHAIN Data Extraction Documentation. Online Document. 2024. https://drive.google.com/file/d/1uEa9rvS0daSEzJI51K4Wcpeaq49SopMJ/.

Tuan DA. 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. https://doi.org/10.3390/tropicalmed9100250.

Undurraga EA, Edillo FE, Erasmo JNV, Alera MTP, Yoon IK, Largo FM, Shepard DS. Disease burden of dengue in the Philippines: adjusting for underreporting by comparing active and passive dengue surveillance in Punta Princesa, Cebu City. Am J Trop Med Hyg 2017 Apr; 96(4):887-898. https://doi.org/10.4269/ajtmh.16-0488.

Wang R & Wang S. How does consensus scoring work for virtual library screening? An idealized computer experiment. J Chem Inf Comput Sci 2001; 41(5):1422–1426. https://doi.org/10.1021/ci010025x.

World Health Organization. Dengue and severe dengue. April 2024. https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue.

Yandex LLC. CatBoost Hyperparameters Guide. March 2024. https://catboost.ai/docs/en/concepts/parameter-tuning.

Yavari Nejad F, Varathan KD. Identification of significant climatic risk factors and machine learning models in dengue outbreak prediction. BMC Med Inform Decis Mak 2021; 21(1):141. https://doi.org/10.1186/s12911-021-01493-y.

Zhao N, Charland K, Carabali M, Nsoesie EO, Maheu-Giroux M, Rees E, Yuan M, Garcia Balaguera C, Jaramillo Ramirez G, Zinszer K. 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. https://doi.org/10.1371/journal.pntd.0008056.