Dengue Risk Detection and Observation System

Hung Wei Lee
Academy of Innovative Semiconductor and Sustainable Manufacturing
National Cheng Kung University
Tainan, Taiwan
m26121061@gs.ncku.edu.tw

Pei-Xuan Li
Department of Electrical Engineering
National Cheng Kung University
Tainan, Taiwan
n28121107@gs.ncku.edu.tw

Po-Jui Lai
Department of Electrical Engineering
National Cheng Kung University
Tainan, Taiwan
n26130702@gs.ncku.edu.tw

Tzu-Chang Lee
Department of Urban Planning
National Cheng Kung University
Tainan, Taiwan
jtclee@mail.ncku.edu.tw

Chih Ching Tsao
Department of Electrical Engineering
National Cheng Kung University
Tainan, Taiwan
e14086622@gs.ncku.edu.tw

Zheng Lu
Department of Mechanical Engineering
National Cheng Kung University
Tainan, Taiwan
e14086266@gs.ncku.edu.tw

Hsun-Ping Hsieh*
Department of Electrical Engineering[1]
Academy of Innovative Semiconductor and Sustainable Manufacturing[2]
National Cheng Kung University
Tainan, Taiwan
hphsieh@mail.ncku.edu.tw

Ally Chang
Department of Electrical Engineering
National Cheng Kung University
Tainan, Taiwan
e24106783@gs.ncku.edu.tw

***Abstract*—This study aims to develop a dedicated dengue fever prediction and monitoring system for the Tainan City Government to predict dengue fever outbreaks using advanced AI technologies. We compared statistical models, linear models, machine learning (ML), and deep learning (DL) models to construct the system. We found that the Graph WaveNet (Gwinnet) model which is based on graph neural networks, performed best for predicting the total egg count. In contrast, the gradient boosting machine learning algorithm (XGBoost) was most effective for predicting the positivity rate. Using these optimal models, we successfully forecasted the total egg count and positivity rate in all regions of Tainan. The system provides a clear and user-friendly interface for the government to quickly view the relationship between the risk areas and spatially influenced factors of dengue. Using the developed real-time risk warning and monitoring system, the efficiency and effectiveness of dengue fever prevention are improved. The potential of AI technology in public health is confirmed, and the system provides a reference for future epidemic prevention efforts.**

Keywords—spatial-temporal analysis, dengue prediction, dengue observation system

I. INTRODUCTION

In early June 2023, Tainan City reported its first case of dengue fever, marking the onset of the dengue outbreak in Taiwan. By the end of June, the number of cases nationwide surged, with clustered outbreaks in Tainan City and Yunlin County. By mid-July, the national case count reached the highest level for the same period in nearly a decade, with weekly new cases continuously rising. Tainan City, in particular, experienced a severe outbreak, with increasing community transmission risk. The favorable summer climate and temperature in the southern region provided an ideal environment for mosquito breeding, accelerating the outbreak. The flight range of the mosquito vector and population movement further spread the disease. Additionally, the difficulty in identifying the infection sources of confirmed cases increased the complexity of epidemic prevention efforts. In response to the rapid increase in dengue cases, we provided the government with relevant information for dengue prevention by formulating effective response measures based on publicly available data in Tainan. These measures include prioritizing areas for mosquito spraying and disinfection and enhancing epidemic monitoring.

We identified high-risk areas for dengue fever to provide accurate predictions of mosquito egg density for the government. Compared with other studies [9,12], we used a wider variety of models and spatial-temporal features to predict total egg count and positivity rate accurately. The model distinguishes between high, medium, and low dengue risk areas as a systematic platform to observe and monitor dengue fever outbreaks and provides prediction results and different influencing factors.

II. METHODOLOGY

In collaboration with the Tainan City Government Dengue Fever Prevention Center, we obtained data on total egg count and positivity rate from ovitraps across Tainan City. We aimed to predict these values using various models by incorporating weather factors. The data were analyzed to explore basic linear regression, and create decision trees and long short-term memory (LSTM) [5] models using the time series data, and graph convolutional networks (GCNs) [10] to understand temporal components based on Tainan's regional adjacency. We implemented GCNs combined with LSTMs (GNN-LSTMs) and the state-of-the-art model, Graph WaveNets (GWNet) [6] for evaluation. Our research process is as follows:

A. Data Collection and Processing

We obtained statistical data from ovitraps in all regions of Tainan City from the Dengue Fever Prevention Center, including total egg count and positivity rate. To better handle the total egg count data, we took the natural logarithm (*ln*) before inputting these values into the model for prediction, reducing the impact of data skewness and outliers and improving prediction performance. Data preprocessing was conducted by collecting weather factors [1], including (minimum) temperature, relative humidity, rainfall, and air pressure, which significantly impact mosquito breeding and virus transmission and normalizing the data to ensure data completeness and consistency.

B. Model Selection and Implementation

We employed basic regression models and statistical models such as the Linear model, the Ridge model [3], the Lasso model [11], and the Elastic model [15]. Machine learning methods such as Support Vector Machine (SVM) [4] and XGBoost [2] were also used for prediction. Then, the Long Short-Term Memory (LSTM) model [5] was incorporated to learn temporal variations over short and long periods. Finally, the GCN [10] was utilized to capture the correlations between regions in Tainan, combining graph and temporal models to achieve spatio-temporal analysis.

C. Model Evaluation

To evaluate the performance of the models, we employed root mean square error (RMSE) as the evaluation metric. We established baseline models that predict either all zeros or the previous week's data and compared their performance to the predictions from chosen models. Finally, we optimized the performance of each model through cross-validation and hyperparameter tuning.

D. Result Analysis

The GWNet model showed the best prediction performance for the prediction task of the total egg count, as it reasonably explained that the spread of vector mosquitoes was related to whether the regions were adjacent. In contrast, the XGBoost model excelled at predicting the positivity rate, suggesting decision tree models were well-suited for the scenario. Based on the experimental results, we chose the best-performing model to predict the total egg count and the positivity rate and integrated the results into the front-end system.

E. System Feature

1) Weather Data Integration

Based on the research on dengue fever and climate [13], we identified weather factors that were previously considered relevant and selected the best features using feature importance analysis, including Station pressure (hPa), Temperature (°C), Minimum temperature (°C), Relative humidity(%), and Precipitation (mm). These features are normalized and lagged by one week before being input into the models. Feature importance analysis revealed that "minimum temperature" is the most crucial factor, significantly impacting dengue fever outbreak predictions.

2) Egg Trap Monitoring Data

We obtained the positivity rate and the total egg count (ln-ln-transformed) from ovitraps set up in various regions by the Tainan City Dengue Fever Prevention Center. These data serve as temporal quantitative data for dengue fever outbreaks, providing a vital basis for model training and prediction.

F. Model and Results

We developed two baseline methods, "Guess 0" and "Guess Last Week". Next, we employed linear models, machine learning (ML) models [8], and deep learning (DL) models [7] for predictions. We then evaluated the prediction performance using RMSE. The results are presented in Table 1. GWNet demonstrated superior performance in predicting the total egg count. Conversely, XGBoost exhibited the best performance in predicting the positivity rate.

TABLE I. EXPERIMENTAL RESULTS

Models	RMSE	
	ln (Total egg count)	***Positivity Rate***
Guess Zero	7.5248	28.5899
Guess Last Week	0.5343	8.7386
Bayesian	0.4878	7.6518
Linear	0.4882	7.6523
Ridge	0.4853	7.6503
Lasso	0.4842	7.6353
Cross Decomposition	0.4872	7.6496
Elastic net	0.4834	7.652
SVM	0.4818	<u>7.2658</u>
XGBoost	<u>0.4746</u>	**7.0071**
LSTM	0.5253	8.4432
GWNet	**0.4146**	7.5811

1) GWNet Model Advantages

GWNet excels in capturing both spatial and temporal features within the data. Spatially, it effectively models the correlations between different regions. This capability reflects the transmission patterns of dengue fever risk, as mosquito populations can travel and spread the disease across geographic areas. Temporally, by incorporating convolutional neural networks (CNNs), GWNet learns the evolving patterns of egg counts over time. This combined spatio-temporal understanding significantly enhances the accuracy of GWNet's total egg count predictions, ultimately contributing to a more robust assessment of dengue fever risk.

2) XGBoost Model Advantages

XGBoost is adept at unraveling the intricacies of positivity rate data. Its strength lies in handling high-dimensional datasets and complex relationships between features. This allows XGBoost to effectively identify a multitude of factors influencing positivity rates. Furthermore, by integrating multiple decision trees, XGBoost can capture the non-linear characteristics inherent in positivity rate data. This combined capability leads to more accurate predictions, providing valuable insights into the factors driving dengue fever outbreaks.

In summary, GWNet and XGBoost demonstrate distinct advantages in different prediction tasks, particularly excelling in forecasting the total egg count and the positivity rate. Therefore, we obtained the respective predictive outcomes through a frontend interface, supplemented with additional relevant features, enabling users to intuitively observe dengue risk areas. The architecture of the two best prediction models is shown in Fig. 1.

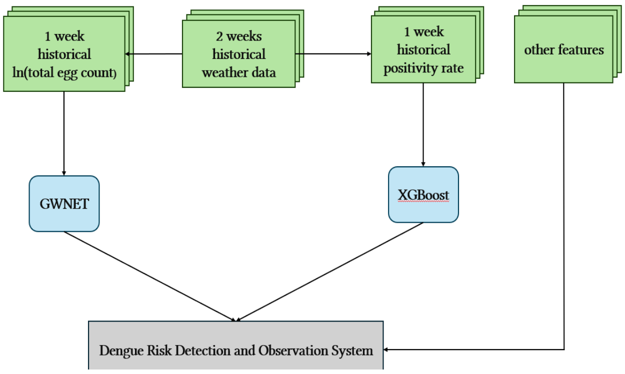

Fig. 1. Architecture of developed system.

III. IMPLEMENTATION

Figures 2–4 show the dengue fever prediction and monitoring system designed for the Tainan City Government. This system predicts the trends of dengue fever outbreaks, providing important and indispensable application value for dengue fever prevention efforts. Using the dengue fever prediction system, users view the predicted values and detailed spatial features of various regions, thus enabling more effective epidemic control.

A. User Application

The web page provides the staff and managers from the Tainan City Government with an efficient, simple, and user-friendly tool for monitoring and handling dengue fever outbreaks.

1) Interactive Map

The right side of the web page features an interactive map of Tainan City. Users can view dengue fever data and other detailed information for different regions. The map can be zoomed in or out using the mouse wheel, quickly locating the desired area (Fig. 2).

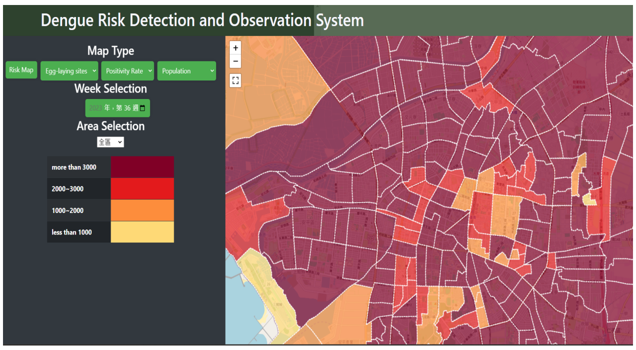

Fig. 2. Population observation.

2) Data Display

The left side of the web page displays dengue fever prediction data and other detailed information, including the total egg count and positivity rate for each region. Users can choose to view related feature maps such as population density (Fig. 2). We selected several significant land use-related factors such as building height to help the government formulate strong policies to combat dengue fever [14].

3) Risk Level Classification Standard

In Fig. 3, the lower left part of the web page is a four-quadrant chart indicating the risk level classification standard for dengue fever. The standard is determined based on the egg count and positivity rate, with the horizontal axis representing the positivity rate, divided by 40%, and the vertical axis representing the egg count, divided by ln(500). Risk levels are classified as follows.

- High-risk areas: positivity rate > 40% and egg count > ln(500)
- Medium-risk areas: positivity rate > 40% or egg count > ln(500)
- Safe areas: positivity rate < 40% and egg count < ln(500)

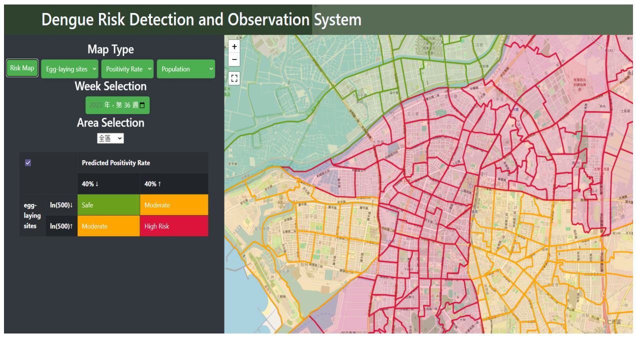

Fig. 3. Dengue fever risk zone observation map.

Through intuitive regional selection, users can access the dengue fever risk prediction results for each area, along with trend charts for total egg count and positivity rates. This enables relevant government departments to promptly understand the dynamics of dengue fever outbreaks and implement corresponding preventive measures. As shown in Fig. 4, the system provides a user-friendly interface for visualizing dengue fever risk predictions and trends.

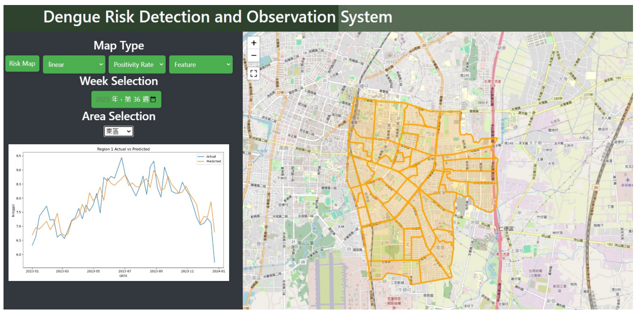

Fig. 4. The dengue front-end system.

IV. CONCLUSION

We developed a dengue fever prediction and monitoring system for the Tainan City Government by combining the image processing model GWNet with the decision tree model XGBoost. This model effectively learns regional adjacency and temporal correlations to predict total egg count and positivity rates in the Tainan region. The system's user-friendly interface enables comprehensive monitoring and aids the government in formulating effective response measures. The system provides timely risk warnings and improves the efficiency of dengue fever prevention. Advanced AI and data analysis can significantly enhance disease prediction accuracy and prevention precision. The system will be further optimized by integrating more data sources and advanced models. It can be applied to other cities with dengue fever risk. The system equips the Tainan City Government with a robust tool for dengue fever prevention and showcases the potential of AI in public health. Continuous research and improvement are necessary to enhance the scientific and practical aspects of disease prediction and monitoring.

Acknowledgment

This research was partially supported by the National Science and Technology Council (NSTC) under Grants 111-2636-E-006 -026 -, 112-2221-E-006 -100 - and 112-2221-E-006 -150 -MY3.

References

[1] Seah Annabel, Joel Aik, Lee-Ching Ng, Clarence C. Tam. "The Effects of Maximum Ambient Temperature and Heatwaves on Dengue Infections in the Tropical City-State of Singapore – A Time Series Analysis."

[2] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA, 785–794. https://doi.org/10.1145/2939672.2939785

[3] Marquardt, Donald & Snee, Ron. (1975). Ridge Regression in Practice. American Statistician - AMER STATIST. 29. 3-20. 10.1080/00031305.1975.10479105.

[4] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt and B. Scholkopf, "Support vector machines," in IEEE Intelligent Systems and their Applications, vol. 13, no. 4, pp. 18-28, July-Aug. 1998, doi: 10.1109/5254.708428.

[5] Majeed Mokhalad, Helmi Zulhaidi Mohd Shafri, Zed Zulkafli, Aimrun Wayayok. "A Deep Learning Approach for Dengue Fever Prediction in Malaysia Using with Spatial Attention." Int J Environ Res Public Health, 2023. https://pubmed.ncbi.nlm.nih.gov/36901139/

[6] Dhankhar Nischay, Rajneesh Tiwari, Ayushman Buragohain. "GWNET: Detecting Gravitational Waves using Hierarchical and Residual Learning based 1D CNNs."

[7] Zhao N., Charland, K., Carabali, M., Nsoesie, E.O., Maheu-Giroux, M., Rees, E., 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, e0008056.

[8] Aleixo Robson, Fabio Kon, Rudi Rocha, Marcela Santos Camargo, Raphael Y. De Camargo. "Predicting Dengue Outbreaks with Explainable Machine Learning."

[9] Yip Stan, Norziha Che Him, Nur Izzah Jamil, Daihai He, Sujit K. Sahu. "Spatio-Temporal Detection for Dengue Outbreaks in the Central Region of Malaysia Using Climatic Drivers at Mesoscale and Synoptic Scale."

[10] Zhang, S., Tong, H., Xu, J. et al. Graph convolutional networks: a comprehensive review. Comput Soc Netw 6, 11 (2019). https://doi.org/10.1186/s40649-019-0069-y

[11] Safi, Samir & Alsheryani, Mouza & Alrashdi, Maitha & Suleiman, Rawan & Awwad, Dania & Abdalla, Zainab. (2023). Optimizing Linear Regression Models with Lasso and Ridge Regression: A Study on UAE Financial Behavior during COVID-19. Migration Letters. 20. 139-153. 10.59670/ml.v20i6.3468.

[12] Chuang Ting-Wu, Ka-Chon Ng, Thi Luong Nguyen, and Luis Fernando Chaves. "Epidemiological Characteristics and Space-Time Analysis of the 2015 Dengue Outbreak in the Metropolitan Region of Tainan City, Taiwan."

[13] Martheswaran T.K., Hamdi, H., Al-Barty, A. et al. "Prediction of Dengue Fever Outbreaks Using Climate Variability and Markov Chain Monte Carlo Techniques in a Stochastic Susceptible-Infected-Removed Model." Sci Rep 12, 5459 (2022). https://doi.org/10.1038/s41598-022-09489-y

[14] Hsiu Yang, Thi-Nhung Nguyen, and Ting-Wu Chuang. "An Integrative Explainable Artificial Intelligence Approach to Analyze Fine-Scale Land-Cover and Land-Use Factors Associated with Spatial Distributions of Place of Residence of Reported Dengue Cases." Tropical Medicine and Infectious Disease, 2023.

[15] Hui Zou, Trevor Hastie, Regularization and Variable Selection Via the Elastic Net, Journal of the Royal Statistical Society Series B: Statistical Methodology, Volume 67, Issue 2, April 2005, Pages 301–320, https://doi.org/10.1111/j.1467-9868.2005.00503.x