



Leveraging Machine Learning Algorithms for Early Detection and Prediction of Dengue Fever

Marwa M. Eid^{1,*}

Christos Gatzoulis²

Osama Al Abedallat²

¹ Faculty of Artificial Intelligence, Delta University for Science and Technology, Mansoura, Egypt

² School of ICT, Faculty of Engineering, Design and Information & Communications Technology (EDICT), Bahrain Polytechnic, Isa Town, Bahrain

Emails: mmm@ieee.org · christos.gatzoulis@polytechnic.bh · osama.alabedallat@polytechnic.bh

Received: September 07, 2023 Revised: December 13, 2023 Accepted: January 15, 2024 ★ Corresponding author

ABSTRACT

This paper explores the potential of machine learning (ML) in revolutionizing the screening and prognosis of dengue fever, a pervasive viral illness transmitted through the bite of infected mosquitoes prevalent in tropical and subtropical areas. Historically, traditional approaches to monitoring outbreaks have been hampered by a lack of precision and timing, creating an opportunity for machine learning to rectify datasets and uncover patterns that enhance accuracy. The paper introduces Random Forests, Support Vector Machines, Neural Networks, and combined classification models, along with their advantages, disadvantages, and the potential for incorporating external data such as climatic factors, population data, real-time Twitter data, etc. The results demonstrate significant increases in accuracy from the models, but it is clear that their applicability is contingent on localized approaches suitable for the regions. This underscores the importance of the quality and completeness of data used in the models. Current research indicates that data availability and the refinement of these models require a collective approach. The work underscores the potential of ML to redefine the paradigm of outbreak management in dengue and other vector-borne diseases, offering hope for improved public health worldwide.

Keywords: Dengue prediction ▪ Machine learning ▪ Outbreak detection ▪ Climatic data ▪ Neural networks ▪ Public health

1. INTRODUCTION

Dengue fever is an acute viral disease transmitted by *Aedes* mosquitoes, which affects public health, especially in tropical and subtropical areas. Dengue fever, which causes outbreaks in approximately 390 million people annually, imposes considerable pressure on global healthcare due to its high transmission level and periodically recurrent nature. It is usually transmitted through *Aedes aegypti* mosquitoes, and the regions with such a climate are very much at risk. The best way to deliberately control and prevent dengue outbreaks is to identify the carrier early and accurately forecast the outbreak to prevent the further spread of the virus. While traditional

approaches to surveillance are helpful, they do not offer quick, localized intelligence that could avoid complications in the event of large epidemics, and so require superior prediction methods [1].

In the last few years, much effort has been put into applying machine learning (ML) techniques to predict and detect dengue disease. These algorithms can dissect complex data using conventional statistical measures to provide inherent and non-apparent relationships or patterns. As a result, prior dengue cases, climate variables, population density, and other indicators can be amalgamated to predict outbreaks using machine learning models more accurately. This advancement

can define new possibilities in public health management by turning from a reactive approach to a proactive, predictive one [2, 3].

The use of climate data, particularly temperature, humidity, and rainfall, within model forecasts has been incredibly effective. Since *Aedes* mosquito breeding and activities are influenced by climatic factors, integrating climatic parameters improves the model's predictive potential. Research has shown that employing more climate-specific models may prove more accurate in regions such as Brazil, Peru, and Thailand. Some of these models may work depending on the local setting, suggesting that implementation is best done locally instead of replicating models developed in another context [4].

Several ML techniques studied in dengue prediction include Random Forest, Support Vector Machines (SVM), Neural Networks, and combinations. Random Forest has been attributed to dealing well with large datasets with many features, making it convenient to study the intricate connections between climate parameters and dengue occurrence. Blended methods, where different machine learning approaches are incorporated, including artificial neural networks with optimization algorithms, represent a promising direction in classification accuracy improvement. These models help determine high-risk time intervals and zones to direct the management of available human and financial public health resources [5].

However, some issues limit the formulation of efficient predictive models for dengue. One major concern is tied to data reliance and availability. When there is no data or insufficient data, the required predictive model is affected by poor accuracy and performance. Further, models trained in one geographical region may not work well in another because climatic, demographic, and healthcare-related factors differ between locations. Therefore, future work should integrate new data sources, use models that consider regional characteristics, and improve data quality to enhance predictive accuracy.

2. LITERATURE REVIEW

Dengue fever is an infectious viral disease transmitted through the bite of infected *Aedes* mosquitoes, focusing on tropical and subtropical areas. Early and reliable prediction of dengue incidence is essential in controlling the disease's spread. This literature considers state-of-the-art trends in applying various machine learning approaches to predicting the frequencies of dengue.

As explained in [8], the study bridges a significant gap in identifying and categorizing recent emergences of dengue fever (DF), as well as providing a distinction between DF and dengue hemorrhagic fever (DHF). Based on microarray technology, the study focuses on the choice of genes to improve classifier performance. A hybrid model combining artificial neural networks (ANN) with a modified cuckoo search (MCS) algorithm was developed to classify subjects into DF, DHF, and convalescent or standard cases. The results proved the effectiveness of the proposed ANN-MCS model over the traditional ANN-CS model.

Da Silva et al. [9] investigated the relationship between climate and dengue incidence by analyzing meteorological data

and dengue cases from Natal, Brazil; Iquitos, Peru; and Barranquilla, Colombia. Using Random Forest, the study examined scenarios that included dengue cases alone, climate and dengue cases, and humidity with dengue cases. The research established that the predictive capability of climatic data is location-dependent, with humidity being a particularly influential forecast variable.

Sebastianelli et al. [10] introduced a machine learning ensemble model designed for predicting the dengue incidence rate (DIR) in Brazil among individuals below 19 years of age. The model uses spatial and temporal data to provide state-level one-month-ahead DIR forecasts. The study demonstrates how such a framework can be applied to other countries, such as Peru, and highlights the value of public health forecasting systems for local authorities.

Addawe et al. [11] used Random Forest, linear regression, SVM, and gradient boosting to forecast dengue cases in Baguio City, Philippines. Random Forest with meteorological and temporal features achieved the best performance. Souza et al. [12] explored manifold learning on climate data to forecast dengue outbreaks in Brazilian cities, showing that manifold learning can be effective for coarse outbreak predictions.

Clinical applications have also received attention. Laoprasopwattana et al. [13] differentiated dengue from Chikungunya and other acute febrile illnesses in children using blood-test indicators and machine learning classifiers. Logistic regression with second-degree polynomial terms achieved promising ROC AUC values. Other studies have examined dengue survivability and hospital length of stay (LOS), showing that age, gender, admission status, assisted ventilation, hemoglobin, total leukocyte count, hematocrit, and lymphocyte count can influence LOS [14, 15, 16].

Geographical variability is critical. Sumanasinghe et al. [17] proposed regional dengue-incidence prediction models for four provinces in Thailand using Support Vector Regression. Rajendran et al. [18] used an RNN-based deep learning model for forecasting vector-borne disease incidence and found that RNNs performed well for dengue case forecasting. Guo et al. [19] developed an ensemble penalized regression algorithm using climate and social-media surveillance data in Guangzhou, China, to improve timeliness and accuracy.

Further studies used Random Forest feature extraction, ANN models, and improved metaheuristic algorithms for dengue prediction. Daniel et al. [20] examined recurrent neural networks for feature extraction from dengue fever data. Raizada et al. [21] used ANN for vector-borne disease outbreak prediction. Ibrahim et al. [22] used ANN to analyze dengue fever and dengue hemorrhagic fever, while Mustaffa et al. [23] proposed an improved Salp Swarm Algorithm for dengue incidence prediction.

Table 1 highlights the key findings of the literature review regarding dengue fever prediction. The studies describe a range of machine learning methods for predicting dengue incidence. Across the research, ensemble models and deep learning successfully classified dengue cases; regional aspects and climatic conditions were influential features; and machine learning showed strong potential for improving dengue prevention and response.

Table 1. Summary of Literature Review

Study	Objective	Methodology	Key Findings
[8]	Classify dengue cases into DF, DHF, and convalescent/standard groups	Bag-of-features, ANN-MCS	ANN-MCS outperforms traditional ANN-CS in classification
[9]	Investigate the relationship between climate and dengue incidence	Random Forest, meteorological data analysis	Predictive capability of climate data is location-dependent
[10]	Predict dengue incidence rate in Brazil	Ensemble model with spatial and temporal data	The ensemble model provides accurate DIR forecasts
[11]	Forecast dengue cases in Baguio City, Philippines	Random Forest, linear regression, SVM, gradient boosting	Random Forest with meteorological and time features performs best
[12]	Forecast dengue outbreaks in Brazilian cities	Manifold learning, signal analysis	Manifold learning is effective for coarse dengue outbreak predictions
[13]	Differentiate between dengue and Chikungunya in children	ML classifiers, blood-test data	L2 logistic regression achieves high classification accuracy
[14]	Predict dengue disease survivability	Deep learning methods	Deep learning methods achieve high classification accuracy
[15]	Predict hospital LOS for dengue patients	Logistic regression, Random Forest	Age, gender, and medical factors influence LOS
[16]	Predict hospital LOS for dengue patients	Logistic regression, Random Forest	Logistic regression with elastic net performs best
[17]	Predict dengue incidence in Thailand	Support Vector Regression	Regional factors influence dengue incidence
[18]	Forecast dengue incidence using deep learning	RNN-based model	RNN outperforms other classifiers in dengue forecasting
[19]	Forecast dengue outbreaks in China	Ensemble penalized regression algorithm	EPRA provides timely and accurate forecasts
[20]	Improve feature extraction for dengue prediction	Random Forest	RF-based feature extraction improves efficiency and accuracy
[21]	Predict vector-borne disease outbreaks	ANN algorithm	ANN achieves high prediction accuracy for Chikungunya, malaria, and dengue
[22]	Predict the day of fever resolution in dengue patients	ANN	ANN accurately predicts the day of fever resolution
[23]	Improve dengue incidence prediction	Improved SSA algorithm (iSSA)	iSSA achieves higher accuracy and faster convergence

Specifically, this literature review explores the increasing prominence of machine learning methods in dengue fever studies. The reviewed studies show that applying machine learning in dengue prediction can significantly enhance public health interventions. Prospective studies may build upon existing frameworks, consider more elaborate datasets, and examine data credibility and meaningfulness issues.

3. CONCLUSION

Using machine learning algorithms for the early diagnosis and prognosis of dengue fever is a significant innovation in the public health approach. Since the dengue threat remains high, especially in warm and subtropical areas, efficient and timely prediction has never been more critical. Various examples have shown that machine learning models help analyze big data, providing insights into patterns that the human eye cannot find and delivering predictions that other statistical models may not provide.

These models enable incorporating multiple inputs such as climatic data, demographics, and new data streams such as micro-blogging to present a holistic view of optimizing dengue control. The review covers Random Forests, Support Vector Machines, hybrid methods, and ensemble techniques, all offering different enhancements to prediction. Neverthe-

less, it is distinct that none of these models or approaches can be used consistently across geographical contexts. Therefore, localized data, region-specific model design, and validation across settings are necessary.

The integration of machine learning in dengue prediction represents a promising pathway for strengthening outbreak preparedness and public health decision-making. Data quality, data completeness, and collaboration among researchers, clinicians, governments, and public health agencies remain crucial. The potential to adapt these models to other vector-borne diseases offers a powerful tool for public health systems worldwide.

REFERENCES

- [1] M. Cabrera, J. Leake, J. Naranjo-Torres, N. Valero, J. C. Cabrera, and A. J. Rodríguez-Morales, "Dengue Prediction in Latin America Using Machine Learning and the One Health Perspective: A Literature Review," *Trop. Med. Infect. Dis.*, vol. 7, no. 10, p. 322, 2022.
- [2] A. Abdulrahman et al., "Exploring Optimization Algorithms: A Review of Methods and Applications," *Journal of Artificial Intelligence and Metaheuristics*, vol. 7, no. 2, pp. 08–17, 2024.

- [3] V. H. Nguyen et al., “Deep learning models for forecasting dengue fever based on climate data in Vietnam,” *PLoS Negl. Trop. Dis.*, vol. 16, no. 6, p. e0010509, 2022.
- [4] F. I. Abdulsalam, P. Antunez, S. Yimthiang, and W. Jawjit, “Influence of climate variables on dengue fever occurrence in the southern region of Thailand,” *PLOS Global Public Health*, vol. 2, no. 4, p. e0000188, 2022.
- [5] S. Patil and S. Pandya, “Forecasting Dengue Hotspots Associated With Variation in Meteorological Parameters Using Regression and Time Series Models,” *Front. Public Health*, vol. 9, p. 798034, 2021.
- [6] A. L. Buczak et al., “Ensemble method for dengue prediction,” *PLoS One*, vol. 13, no. 1, p. e0189988, 2018.
- [7] M. A. Majeed et al., “A Deep Learning Approach for Dengue Fever Prediction in Malaysia Using LSTM with Spatial Attention,” *Int. J. Environ. Res. Public Health*, vol. 20, no. 5, p. 4130, 2023.
- [8] S. Anno et al., “Challenges and implications of predicting the spatiotemporal distribution of dengue fever outbreak in Chinese Taiwan using remote sensing data and deep learning,” *Geo-spatial Information Science*, vol. 27, no. 4, pp. 1155–1161, 2024.
- [9] S. T. da Silva et al., “When climate variables improve the dengue forecasting: a machine learning approach,” *European Physical Journal Special Topics*, pp. 1–15, 2024.
- [10] A. Sebastianelli et al., “A reproducible ensemble machine learning approach to forecast dengue outbreaks,” *Scientific Reports*, vol. 14, no. 1, pp. 1–17, 2024.
- [11] J. C. Addawe, J. D. L. Caro, and R. A. B. Juayong, “Machine Learning Methods for Modeling Dengue Incidence in Local Communities,” *Lecture Notes in Networks and Systems*, vol. 556, pp. 392–400, 2023.
- [12] C. Souza, P. Maia, L. M. Stolerma, V. Rolla, and L. Velho, “Predicting dengue outbreaks in Brazil with manifold learning on climate data,” *Expert Syst. Appl.*, vol. 192, p. 116324, 2022.
- [13] K. Laoprasopwattana et al., “Differential diagnosis of Chikungunya, dengue viral infection and other acute febrile illnesses in children,” *Pediatr. Infect. Dis. J.*, vol. 31, no. 5, pp. 459–463, 2012.
- [14] T. J. Schaefer, P. K. Panda, and R. W. Wolford, “Dengue Fever,” *BMJ Best Practice*, pp. 5–6, 2024.
- [15] H. J. M. Hendri and H. Sulaiman, “Predictive modeling for dengue patient’s length of stay (LoS) using big data analytics (BDA),” *Lecture Notes on Data Engineering and Communications Technologies*, vol. 5, pp. 12–19, 2018.
- [16] S. A. Kularatne and C. Dalugama, “Dengue infection: Global importance, immunopathology and management,” *Clinical Medicine*, vol. 22, no. 1, p. 9, 2022.
- [17] N. Sumanasinghe et al., “Data driven prediction of dengue incidence in Thailand,” *Advances in Intelligent Systems and Computing*, vol. 566, pp. 95–107, 2018.
- [18] N. M. Rajendran et al., “Communicable Disease Prediction Using Machine Learning and Deep Learning Algorithms,” *Lecture Notes in Networks and Systems*, vol. 757, pp. 979–992, 2023.
- [19] P. Guo et al., “An ensemble forecast model of dengue in Guangzhou, China using climate and social media surveillance data,” *Science of The Total Environment*, vol. 647, pp. 752–762, 2019.
- [20] J. Daniel et al., “Recurrent Neural Networks for Feature Extraction from Dengue Fever,” 2022.
- [21] S. Raizada, S. Mala, and A. Shankar, “Vector-Borne Disease Outbreak Prediction Using Machine Learning Techniques,” *EAI/Springer Innovations in Communication and Computing*, pp. 227–241, 2021.
- [22] F. Ibrahim et al., “A novel dengue fever (DF) and dengue haemorrhagic fever (DHF) analysis using artificial neural network (ANN),” *Comput. Methods Programs Biomed.*, vol. 79, no. 3, pp. 273–281, 2005.
- [23] Z. Mustaffa et al., “Predictive Analysis of Dengue Outbreak Based on an Improved Salp Swarm Algorithm,” 2020.