



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

Marwa M. Eid^{1,*}, Christos Gatzoulis², Osama Al Abedallat²

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

²School 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

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 the *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 some inherent and non-apparent relationships or patterns. As a result, what can be called 'big data'

on prior dengue cases, climate variables, population density, etc., can be amalgamated with the indicators to predict the outbreak, 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 breeds and activities are influenced or affected 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 regionals 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 worked out in another context. For example, although climate data enhanced prediction accuracy in some areas but not in others, it implies that place-related factors constitute critical variables for enhancing these prognostic instruments [4].

Several ML techniques studied in dengue prediction include Random Forest, Support Vector Machines (SVM), Neural Networks, and combinations. For example, Random Forest has been attributed to dealing well with big datasets with many features, making it convenient to study the intricate connections between climate parameters and dengue occurrence. On the other hand, using the blended methods, where different machine learning is incorporated, including the artificial neural network with optimization algorithms, represents a promising direction in classification accuracy improvement due to the integration of the given approaches. These models help determine high-risk time intervals and zones to help 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 of the major concerns is tied to data reliance and availability. When there is no data or insufficient data gathered, or in areas such as climatic conditions, slaughter population or even health records are hard to come by, then constructing sound models is a Herkules task. However, the performance of the ML models depends on the space and time distribution of dengue transmission. To achieve great certainty and reliability in size, models need locally bound data corresponding to exact environmental and demographic characteristics and prerequisites. Meeting these challenges demands both the application of advanced technologies and advancements in data gathering and data-utilization methods [6].

Afterward, methods in deep learning and ensemble approaches have brought additional enhancements to the dengue risk prediction involving several data inputs. For instance, some research has incorporated data from social applications involving disease symptoms searches and tweets about dengue symptoms with the conventional epidemiological and climatic factors to expand prediction models. These ensemble models can process information and quickly and accurately give forecasts, which can help start early intervention. The introduction of actual time sources, for example, through social media, is convenient since it enables one to receive almost contemporary reports regarding any areas of probable outbreaks and, therefore, support a fast and effective public health intervention [7].

In conclusion, using machine learning for the early diagnosis and estimation of dengue threat is possible, which makes the improvement of the population's health futureless. That way, resources can be appropriately targeted, timely vector control can be launched, and the impact of dengue transmission can be mitigated. Future work should be directed to integrating new data sources, using models that can consider regional characteristics, and improving data quality to improve predictive accuracy. With such advancements in machine learning technology, there is promise in how dengue and other vector-borne diseases can be monitored, predicted, controlled and managed to improve global health.

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 that concerns the paper bridges a significant gap in identifying and categorizing the recent emergences of DF, as well as providing a distinction between dengue fever (DF) and dengue hemorrhagic fever (DHF). Based on the development of microarray technology, the study focuses on the choice of genes to improve the performance of a classifier. The authors described a more straightforward bag-of-features modification approach to identify the most suitable genes for the classification task. Subsequently, a hybrid model combining artificial neural networks (ANN) with a modified cuckoo search (MCS) algorithm (ANN-MCS) was developed to classify subjects into three distinct groups: DF, DHF and a group consisting of convalescent and standard cases. Three of the most popular classifiers were used to compare the proposed method: MLP-FFN, ANN-CS, and ANN-PSO. During the feature selection, at the beginning, experiments set different cluster numbers using bag-of-features. Classification of the reduced dataset was then performed using the hybrid ANN-MCS model. The experimental results analyzed using the confusion matrix-derived performance measures proved the effectiveness of the proposed ANN-MCS model over the traditional ANN-CS model in enabling the effective classification of dengue.

Dengue fever is a viral vector-borne illness that has spread in many countries, affecting about 390 million people yearly and with frequent incidences developing in subtropical and tropical regions. The study referenced as [9] investigates the relationship between climate and dengue incidence by analyzing meteorological data and dengue cases from three locations: Natal of Brazil from 2016 to 2019, Iquitos of Peru from 2001 to 2012, and Barranquilla of Colombia from 2011 to 2016. The authors intend to predict potential dengue occurrences using currently available and lagged climate predictors by deploying machine learning methods, especially the random forest algorithm. They examine three scenarios for their machine learning analysis: when including only dengue cases, climate and dengue cases, humidity, and dengue cases. This research established that the predictive capability of climatic data, when used to analyze dengue, is location-dependent. For example, better results are obtained if climatic data are excluded in Natal and all climatic data are included in Iquitos. The best model in Barranquilla was the one where humidity was added to the dengue cases. The study also shows that accurate dengue case forecasting is not easy, primarily if it depends on climate factors. While such variables improve on the basic models, their benefits are locationality dependent, and humidity is the most influential forecast variable.

Dengue fever is an arboviral disease of significant and increasing importance in the tropical and subtropical regions of the world, and it continually presents a public health and economic burden. The paper marked as [10] introduces the machine learning ensemble model designed for predicting the DIR in Brazil among individuals below 19. This model uses spatial and temporal data to provide state-level one-month-ahead DIR forecasts. Further, using a dummy comparison baseline and ablation studies confirms the qualitative and quantitative benefits of the ensemble model at the granularity of 27 Brazilian Federal Units. Therefore, the researchers illustrate how this framework can be applied to other countries, such as Peru, with different epidemiological features. This forecasting system has an enormous stake for the local authorities as it provides tools to implement conservative controls. The research contributes to improving the provision of climate services for health by determining factors leading to dengue incidence in Brazil and Peru and calling for cooperation with international and public health organizations. The innovation is not only in the algorithms but also in the application of the algorithms to a domain that is frequently constrained by data availability and operation scale. Imagining well-sourced ground data with state-of-the-art data analysis, the investigation fills several of the voids of the current practices. The implementation of the model in Peru and the results obtained during the recent experience in Brazil prove the possibility of further scale and more effective use of the model. Despite the challenges encountered in elements of extreme values, particularly in areas with low DIR, such an approach reveals the accuracy of predictions, especially where accuracy is of the essence. This work adds to improving the accuracy of DIR forecasting and epitomizes a shift in applying high-level analytics toward utilizing operational models in public health; the paradigm changes are expounded as being poised to set a global common ground for inter-discipline collaboration in terms of critical combatants of global health problems.

This paper discusses the model built to forecast dengue in Baguio City, Philippines. As highlighted in the analysis performed in [11], Random Forest, Linear Regression, Support Vector Machines (SVM) and Gradient Boosting models were used to assess the correctness of the projected dengue cases over five barangays, explaining the highest average incidence rate per year of the disease. Several other parameters,

including meteorological and time, were applied as input attributes to the model. Results presented here show that the results of the extended SVM model were the lowest of all considered algorithms. Nevertheless, Linear Regression was most efficient for minimal input space and prediction error as close as possible to the linear regression models. Most notably, the Random Forest model performed significantly better when both the features related to the time of the accident and meteorological data, including rainfall (RF) and relative humidity (RH), were utilized. They, too, were least error-prone, especially during the high transmission season of dengue, supporting the use of meteorological factors in models.

Despite various social vectors, tropical countries urgently need solutions to prevent dengue outbreaks. For this reason, the article's authors denoted as [12], underscore the importance of timely public policy interventions to provide effective vector-control strategies that critically depend on accurate prediction tools. The present work expands from a relatively recent approach to forecast epidemic episodes in large Brazilian cities with yearly climate information. The authors present the methodological improvements as careful data preprocessing and modern computational tools from signal analysis and manifold learning. Therefore, their prediction attains an accuracy score of improvement of 0.72 to 0.80. These findings manifested that manifold learning can be a feasible approach to coarse dengue outbreak predictions in large urban zones when used on climatic data. This approach can dramatically reduce the data needed for outbreak analysis, primarily if the municipality's public health department cannot collect and report various features necessary for more advanced machine learning techniques.

Dengue and Chikungunya are both vector-borne diseases prevalent in tropical regions and display overlapping clinical features that make the diagnosis for a clinician challenging. While conducting a study in [13], the authors suggest applying machine learning classifiers to decide between dengue and Chikungunya in children using blood test outcomes, not symptoms. To develop a dataset, the study involved 447 pediatric patients of Hospital Infantil Napoleón Franco Pareja and obtained three essential values, namely platelet count, white blood cell count, and hematocrit percentage. Stratified Random Sampling was then used to split this dataset into training and testing datasets. The authors used Grid Search with Stratified 5-fold Cross-Validation to analyze and compare the performance of several classification algorithms: Logistic Regression, Support Vector Machines, and CART Decision Trees. The results showed that the L2 Logistic Regression, including second-degree polynomial terms, yielded higher classification accuracy than other examined classifiers with the cross-validated ROC AUC equal to 0.6894. After additional testing on the validation set, this reached an ROC AUC score of 0.8502. Perhaps owing to the many limitations, it is comforting to note that even when model sensitivities were reduced due to small sample size and highly skewed data, ROC AUC scores were promising, affirming the feasibility of this machine learning application in clinical care, particularly in targeting the specific needs of pediatric populations with EVD and distinguishing between co-occurring viral diseases like dengue and Chikungunya in children.

Dengue fever has, in the recent past, been affecting many patients, and its occurrence has been on the rise. The authors of the publication, marked as [14], state that the primary goals of dengue control and outbreaks should be focused on timely diagnosis and prevention actions, especially in the areas where this disease is frequently observed. In other words, for a solid theory by which you can predict the occurrence of dengue, you must first know how to prevent its occurrence. Other dynamics like an increase in temperature, increasing sea surface temperatures, rapid growth in urban areas, and increased rainfall resulting from global warming also play a role in the outbreak. In addition, abrupt human movements and migrations due to increasing speed urbanization and climate change increase the chances of infected *A. aegypti* mosquitoes. The study reveals that high classification accuracy can indeed be gained by deep learning methods in dengue patient detection by pointing out the applicability and explaining that they can do regression without much tuning. The outputs generated by these methods are quickly interpretable and further allow for detection between healthy people and dengue patients based on the factors. These factors improve the stability and accuracy of the dengue disease survivability prediction model and avoid overfitting the model. Besides, the proposed method is relevant to different practical problems, including decision support systems in the healthcare context, individualized health communication, and efficient risk management.

Therefore, the study aims to establish predictive indicators and predict the hospital LOS of dengue patients while focusing on the efficient utilization of the hospitals. In the qualitative study performed in [15], authors

reviewed 1,360 medical records of the confirmed dengue cases documented in the Max Group of Hospitals in India for 2012–2017. Two data mining algorithms, logistic regression (LR) with elastic-net regularization and random forest, were used to extract predictive factors and predict LOS. Therefore, these classifiers' efficacy was measured by an area under the curve (AUC), sensitivity, and specificity. It was observed that both classifiers yielded high accuracy; the AUC score aspired by the logistic regression was 0.75, while random forest had an AUC score of 0.72. Of 1,148 patients enrolled in the study, 364 (32%) displayed a simple, prolonged LOS 5 days, with a mean hospitalization time of 4.03 ± 2.44 days. Our study identified the 10–20 years age group as the most affected by dengue, accounting for 21.1% and more affected males and specifically on independent variables, transfused patients, those admitted as emergencies, those who needed assisted ventilation, low hemoglobin levels, high TLC, abnormal hematocrit, and low lymphocyte counts had an increased mean LOS. Based on the present research, it is clear that logistic regression with elastic net was the best model in identifying that an LOS of more than five days had a close, significant relationship with some patient factors. From this approach, one may determine which patients are at a higher risk and direct the time and effort in clinical facilities.

Therefore, the study aims to establish predictive indicators and predict the hospital LOS of dengue patients while focusing on the efficient utilization of the hospitals. In the qualitative study performed in [16], authors reviewed 1,360 medical records of confirmed dengue cases, which were documented in the Max Group of Hospitals in India for 2012–2017. Two data mining algorithms, logistic regression (LR) with elastic-net regularization and random forest, were used to extract predictive factors and predict LOS. Therefore, these classifiers' efficacy was measured by an area under the curve (AUC), sensitivity, and specificity. It was observed that both classifiers yielded high accuracy; the AUC score aspired by the logistic regression was 0.75, while random forest had an AUC score of 0.72. Of 1,148 patients enrolled in the study, 364 (32%) displayed a simple, prolonged LOS ≥ 5 days, with a mean hospitalization time of 4.03 ± 2.44 days. Our study identified the 10–20 years age group as the most affected by dengue, accounting for 21.1% and more affected males and specifically on independent variables, transfused patients, those admitted as emergencies, those who needed assisted ventilation, low hemoglobin levels, high TLC, abnormal hematocrit, and low lymphocyte counts had an increased mean LOS. Based on the present research, it is clear that logistic regression with elastic net was the best model in identifying that an LOS of more than five days had a close, significant relationship with some patient factors. From this approach, one may determine which patients are at a higher risk and direct the time and effort in clinical facilities.

Transmissible diseases like dengue fever remain a global threat and cause more than 400 million incidents annually. In the research conducted and explained in [17], the authors investigate dengue fever's predictive model, paying particular attention to geographic differences. Choosing a proper approach for developing a prediction model for dengue suffers from the fact that geographic variations play a crucial role in the disease's spread. Dengue is naggingly endemic in different regions of the world, and hence, the models used in the eradication have to be region-based. The study provides a model of dengue incidence prediction for four provinces in Thailand while considering regional factors. Thus, there is variability in the forecasted outcomes across the provinces; attaining this objective necessitated using Support Vector Regression (SVR) as the core modeling technique. In addition, the authors presented a new technique for evaluating the regression models that improve on MSE but are inefficient in capturing the spatial behavior of data patterns. A new evaluation method provided about 71% of the prediction of potential capabilities in the Kamphaeng Phet province. This research offers important implications for Thai administrative bodies, particularly in formulating public health policies and identifying and planning responses to dengue incidences.

Vector-borne diseases are more than one-fifth of infectious diseases, affecting more than 700,000 lives yearly. As pointed out by the authors in [and the percent descends], the research published in [18], malaria and dengue disease impose a considerable disease burden; nearly 219 million people are affected by malaria annually; over 400,000 people die of malaria annually, the majority of whom are children below five years of age. Dengue, recognized as the most common viral illness, has approximately 96 million clinical manifestation cases and 40,000 death cases per year. As seen with dengue, the number of primary vector-borne diseases has increased for the past decade. In response to this problem, the authors use an RNN-based deep learning model for forecasting future incidences of dengue fever that can aid governments in preventing the disease's spread and help healthcare systems prepare for growing medical needs. Understanding the

absolute effect climate change has had on the trend of vector-borne diseases, the model uses data on disease incidence and climatic factors in Kerala to project future episodes. The research concludes that RNN is better than the other classifiers in the study, mainly estimating dengue cases in the subsequent year, as SVR gives a better estimation of Chikungunya. In this regard, it also focuses on designing a simple graphical user interface (GUI) to improve the end-user application of the developed predictive models.

China experienced an unprecedented dengue outbreak in 2014, reaching the highest number of cases recorded in the past 25 years. In the study referenced as [19], the authors address the significant delays in official case count data, highlighting the challenges in timely tracking the onset and magnitude of local dengue outbreaks. To enhance forecasting capabilities, they developed an ensemble penalized regression algorithm (EPRA) that integrates various penalty methods—LASSO, Ridge, Elastic Net, SCAD, and MCP—with iterative sampling and model averaging techniques. This approach utilizes multiple streams of near-real-time data, including dengue-related Baidu searches, Sina Weibo posts, and climatic conditions, alongside historical dengue incidence records. The study compares the predictive power of the EPRA against alternative penalized regression models that employ single penalties, retrospectively forecasting weekly dengue incidence and identifying outbreak occurrences using various thresholds from 2011 to 2016 in Guangzhou, southern China. The findings demonstrate the effectiveness of the EPRA in providing timely and accurate forecasts, contributing to improved public health responses to dengue outbreaks.

These models offer valuable insight into the fight against dengue fever in endemic countries and the improvement of the management of vector-borne diseases. An accurate prediction of dengue cases is essential in controlling the disease, especially given that little can be done to cure it. In the work previously referred to as [20], the authors propose a deep learning model using random forest (RF) approaches for feature extraction about dengue fever extracted from text data. The proposed methodology's operations include data gathering, initial treatment of input texts, and feature extraction. The compatibility of the proposed RF-based feature extraction is tested on dengue data sets, and the developed technique shows more than 12% absolute improvement over existing feature extraction techniques. In addition, this investigation exhibits the model's effectiveness in improving efficiency in feature extraction with 10% less error rates than those currently in use. In this respect, the findings knowledge enhance the theories that underpin separate types of predictive modeling in the public health domains, including surveillance and response to dengue fever.

In the research being discussed in [21], parasitic, viral, and bacterial diseases transmitted through insect vectors are described as significant concerns to public health, with more than 700000 deaths recorded annually from diseases such as yellow fever and malaria, especially in the tropical and subtropical regions. These diseases happen mostly among a lowly class of people. The authors underline deep learning, a fundamental component of artificial intelligence that allows the development of multilayer networks that mimic human brain neural structures. Deep learning appears when combined with machine learning algorithms and might act effectively in predicting the future of medical trends in data with the help of supervised learning. This work is more inclined toward enhancing the effectiveness of multiclass classification of chikungunya, malaria and dengue vectors in the Indian subregion. Data from 2013 to 2017 is analyzed and optimized, resulting in an artificial neural network algorithm for outbreak risks in Italy. In particular, this work adds a feature to analyzing medical datasets using contrasting data sets, which has yet to be done in this context. The ANN algorithm used in the current research yields a prediction accuracy of about 86%, and thus, the results offer a good foundation for upgrading public health response strategies.

In the analysis done in [22], dengue fever (DF) is an acute viral disease of the instantaneous self-limited febrile disease characterized by headaches, muscle and joint pains and rashes. Many DF patients develop a more severe form of the DHF disease, characterized by clinical findings such as plasma leakage from capillaries, pulsed hemoconcentration, ascites and pleural effusion. This can be followed by shock upon the resolution of fever, making the precise prediction of fever defervescence essential to managing such patients. Nevertheless, in the context of this disease, more research is needed to develop models to estimate the time to defervescence in patients with DF. The authors present a noninvasive predictive day of discharge system using artificial neural networks (ANN) based exclusively on clinical fever data and temperatures. The envisaged multilayer feed-forward neural network (MFNN) claims high effectiveness with a prediction

accuracy of about 90% for identifying the day-of-fever resolution in dengue patients. This result has significant ramifications for medical practice and consumer treatment of dengue.

The study proposed in [23] seeks to develop and improve the basic SSA by including Levy Flight as iSSA. This change is intended to increase the algorithm’s ability to perform exploration rather than being stuck into some subset of the local optimum. The extended SSA was used to estimate dengue incidents monthly using data with factors like rainfall, humidity, temperature, and the number of dengue incidents. The performance of the iSSA was evaluated based on RMSE and then compared with that of the SSA and ACO. According to the findings, the iSSA yielded a smaller RMSE; this is an implication that the methodology was accurate in its predictions relative to the traditional linear models compared in this study; additionally, the rate of the convergence of the iSSA was faster and slower which points to its efficiency relative to other traditional models.

Table 1 highlights the key findings of the literature review regarding dengue fever prediction. The studies described a range of techniques for using machine learning methods to predict the incidence of dengue. This work identified that ensemble models and deep learning successfully classified dengue cases; the regional aspects and climactic conditions were influential features and highlighted the prospect of machine learning for enhancing the response to dengue incidents. In sum, across the research, the possibilities of machine learning algorithms are underlined as providing significant benefits in enhancing dengue prevention and combating.

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	The predictive capability of climate data is location-dependent
[10]	Predict dengue incidence rate (DIR) in Brazil	Machine learning ensemble model, 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	Machine learning 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 length of stay (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	Similar findings to [15], logistic regression with elastic net performs best
[17]	Predict dengue incidence in Thailand	Support vector regression	Regional factors influence dengue incidence, and SVR provides accurate predictions.
[18]	Forecast dengue incidence using deep learning	RNN-based model	RNN outperforms other classifiers in dengue case 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 algorithm 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 than traditional methods

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

3. Conclusion

Therefore, 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 prediction means timely forecast has never been more critical. Various examples over the years have shown that machine learning models help analyze big data, providing insights into patterns that the human eye cannot find and providing guaranteed predictions that other statistical models may not provide. These models enable incorporating multiple inputs like climatic data, demographics, and even new data streams like micro-blogging to present a holistic view of optimizing dengue control. The review walks through aspects of both the going and ideal systems. It includes various possible techniques ranging from Random Forests, Support Vector Machines, and a combination of hybrid and ensemble techniques, all offering different enhancements to the actual prediction. Nevertheless, it has become distinct that none of these models or approaches can be used consistently across geographical contexts. Therefore,

local models subsuming such regional factors are critical for increasing the prediction accuracy and efficiency. In addition, better methods of collecting data, enhancing quality, and increasing the use of standard measures will help enhance these models. Several research questions persist, especially regarding data access and how to effectively merge multiple data files. Such issues need to be solved through cooperation between governments, healthcare-related organizations, and tech developers with the help of constant data analysis and practical, sustainable models. More research should also be done to increase the model's interpretability so that it can be implemented easily by the public health systems. As a whole, the use of machine learning in dengue forecasting marks a significant leap in the application of big data analytics in public health. This shift from reactive to proactive disease management underscores the transformative power of technology in public health. The combined impact of such advancements promises a substantial reduction in the global burden of dengue through early interventions and targeted prevention strategies. As technology advances, 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, Oct. 2022, doi: 10.3390/TROPICALMED7100322.
- [2] A. Abdulrahman, .. Z. M., A. M. Zaki, F. H. H.Rizk, M. M. Eid, and E.-S. M. EL EL-Kenawy, "Exploring Optimization Algorithms: A Review of Methods and Applications," *Journal of Artificial Intelligence and Metaheuristics*, vol. 7, no. 2, pp. 08–17, 2024, doi: 10.54216/JAIM.070201.
- [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, Jun. 2022, doi: 10.1371/JOURNAL.PNTD.0010509.
- [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, Apr. 2022, doi: 10.1371/JOURNAL.PGPH.0000188.
- [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, Nov. 2021, doi: 10.3389/FPUBH.2021.798034/BIBTEX.
- [6] A. L. Buczak, B. Baugher, L. J. Moniz, T. Bagley, S. M. Babin, and E. Guven, "Ensemble method for dengue prediction," *PLoS One*, vol. 13, no. 1, p. e0189988, Jan. 2018, doi: 10.1371/JOURNAL.PONE.0189988.
- [7] M. A. Majeed, H. Z. M. Shafri, Z. Zulkafli, and A. Wayayok, "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, Mar. 2023, doi: 10.3390/IJERPH20054130.
- [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, Jul. 2024, doi: 10.1080/10095020.2022.2144770.
- [9] S. T. da Silva et al., "When climate variables improve the dengue forecasting: a machine learning approach," *The European Physical Journal Special Topics* 2024, pp. 1–15, Jun. 2024, doi: 10.1140/EPJS/S11734-024-01201-7.
- [10] A. Sebastianelli et al., "A reproducible ensemble machine learning approach to forecast dengue outbreaks," *Scientific Reports* 2024 14:1, vol. 14, no. 1, pp. 1–17, Feb. 2024, doi: 10.1038/s41598-024-52796-9.

- [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 LNNS, pp. 392–400, 2023, doi: 10.1007/978-3-031-17601-2_38/FIGURES/3.
- [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, Apr. 2022, doi: 10.1016/J.ESWA.2021.116324.
- [13] K. Laoprasopwattana, L. Kaewjungwad, R. Jarumanokul, and A. Geater, "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, May 2012, doi: 10.1097/INF.0B013E31824BB06D.
- [14] T. J. Schaefer, P. K. Panda, and R. W. Wolford, "Dengue Fever," *BMJ Best Practice*, pp. 5–6, Mar. 2024, Accessed Oct. 22, 2024. [Online]. Available: <https://www.ncbi.nlm.nih.gov/books/NBK430732/>
- [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, doi: 10.1007/978-3-319-59427-9_2.
- [16] S. A. Kularatne and C. Dalugama, "Dengue infection: Global importance, immunopathology and management," *Clinical Medicine*, vol. 22, no. 1, p. 9, Jan. 2022, doi: 10.7861/CLINMED.2021-0791.
- [17] N. Sumanasinghe, A. R. Mikler, J. Muthukudage, C. Tiwari, and R. Quiroz, "Data driven prediction of dengue incidence in Thailand," *Advances in Intelligent Systems and Computing*, vol. 566, pp. 95–107, 2018, doi: 10.1007/978-3-319-60663-7_9/FIGURES/3.
- [18] N. M. Rajendran, M. Karthikeyan, B. Karthik Raja, K. Pragadishwaran, E. A. Gopalakrishnan, and V. Sowmya, "Communicable Disease Prediction Using Machine Learning and Deep Learning Algorithms," *Lecture Notes in Networks and Systems*, vol. 757 LNNS, pp. 979–992, 2023, doi: 10.1007/978-981-99-5166-6_66/FIGURES/13.
- [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, Jan. 2019, doi: 10.1016/J.SCITOTENV.2018.08.044.
- [20] J. Daniel et al., "Recurrent Neural Networks for Feature Extraction from Dengue Fever," 2022, doi: 10.1155/2022/5669580.
- [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, doi: 10.1007/978-3-030-66519-7_9/FIGURES/2.
- [22] F. Ibrahim, M. N. Taib, W. A. B. W. Abas, C. C. Guan, and S. Sulaiman, "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, Sep. 2005, doi: 10.1016/J.CMPB.2005.04.002.
- [23] Z. Mustafa et al., "Predictive Analysis of Dengue Outbreak Based on an Improved Salp Swarm Algorithm," doi: 10.2478/cait-2020-0053.