Dengue Cases Prediction Using Machine Learning Approach

Aima Aziz¹, Azka Aziz²

¹ MPhil Scholar, Computer Science and Information Technology, Khwaja fared University of Engineering and Technology, Pakistan. Email: aimafarhan567@gmail.com
² MPhil Scholar, Computer Science and Information Technology, Virtual University, Pakistan. Email: azkaaziz15@gmail.com

ABSTRACT

Dengue fever, spread by mosquitoes, affects about 3.9 billion people worldwide. Health officials could use indicators of dengue fever outbreaks to start taking preventative measures. Controlling dengue fever may be more straightforward for local authorities if they have timely and accurate disease forecasts. As one of the most rapidly spreading diseases globally, dengue fever is a threat to everyone. Dengue outbreaks can be predicted using machine learning, according to this study. Dengue prediction models could benefit from nature-based algorithms being boosted or used. The only thing that mattered in the prediction and training model was the week of the year, which was the only thing that signified. A standard machine learning algorithm cannot simulate long-term dependencies in time-series data, which is necessary for accurate projections in Dengue fever cases. When it comes to developing risk criteria for severe Dengue, machine learning could be a valuable implement in determining the possible behavior to formulate.

Keywords: Dengue Fever (DF) Machine learning Training dataset Prediction Forecasting Experimental data

1. Introduction

Dengue fever is a mosquito-borne sickness that is spread by the dengue virus. The severity of the sickness varies according to the local climate. There are some symptoms, including a high fever and a skin rash. Bleeding, low blood platelet levels and blood plasma leakage or shock syndrome can occur in a tiny percentage of cases (Xu et al., 2019; Zhang, Su, & Chen, 2021). When a mosquito assaults someone with dengue fever, they've been nibbled by a mosquito carrying the dengue worm. Tens to hundreds of millions of people are struck down each year by the illness, prevalent in tropical states worldwide. Dengue fever can be fatal. Thousands of individuals are killed each year as a result. Mosquitoes transmit Dengue in tropical and subtropical regions. In general, they prefer to stay below 1000 meters above sea level, though they can go as high as 2000 meters (Mathulamuthu, Asirvadam, Dass, & Gill, 2017). For those who live near a body of water, mosquitoes are more likely to carry diseases. People, not just mosquitoes, can spread Dengue from one community to another.

Dengue virus has conformist much consideration (Oguntimilehin, Adetunmbi, & Abiola, 2013). Dengue's range is affected by various factors, including setting, demographics, and socio economics. Regression analysis, correlation studies, and time series analysis are the most common data analysis techniques in this research. It is common for researchers to obtain data from Mexico, which has a variety of states with varying climates, economies, and populations. The authors of (McGough, Clemente, Kutz, & Santillana, 2019) employ an auto-regressive model to examine how climate parameters (such as temperature and precipitation) affect the incidence of Dengue on the Texas-Mexico
border between 1995 and 2005. From 1985 to 2007, researchers utilized multiple linear regression models to investigate the relationship between dengue disease incidence and climate change in 12 warm and humid Mexican states. Using wavelet analysis, researchers in Mexico and Thailand have found temporal and frequency-specific correlations between temperature and dengue fever across many years. Figure 1 displays Symptomatic Dengue infections. Bomfim et al. (2020); (Kondeti et al., 2019) researchers premeditated Singapore’s level of immunity and hyperendemicity from 1980 to 2009.

![Figure 1: Symptomatic Dengue infections](image)

In other words, observing for clusters of temperature and dengue patterns that span across multiple states and even years, but not consecutively. As these patterns expand in both space and time, new clusters will form at the points where they break down (Alfred & Obit, 2021). Implementing a global model could aid in the more detailed modelling of changes over time and place.

2. Literature Review

As far back as the 1980s, dengue fever has been an issue in Malaysia by (Salim et al., 2021). It’s no surprise that Selangor has the most cases. Dengue fever outbreak predictors could provide health professionals with the knowledge they need to take preventative measures straight immediately. There was a search for the best machine learning model to predict when Dengue outbreaks would occur in Selangor, Malaysia, from 2013 to 2017. Temperature, wind speed, humidity, and rainfall were all considered by each model. The linear kernel SVM (accuracy = 70 percent, sensitivity = 14 percent, specificity = 95 percent, precision = 56 percent) was chosen as the best predictor based on the experiment results (Halim, Octavia, & Handojo, 2019). This test sample had 63.54 percent sensitivity to SVM (linear), which was better than the 14.4 percent sensitivity for imbalanced data (original data). Week-of-the-year (SVM) model had only one significant factor: the week of the year.

Dourjoy, Rafi, Tumpa, and Saifuzzaman (2021) suggested that Dengue, a virus that makes people revolt, is common. More people have been infected with dengue fever than in previous years. Dengue fever cases are also rising, with new cases reported daily (Dourjoy et al., 2021). Dengue fever can strike anytime, anywhere, to any number of people. Dengue is a disease that can be prevented early on if people are told to make the appropriate decisions and exercise caution. If this does happen, it's impossible to predict in advance. Thus, it's difficult to say. There is a purpose to this study: to look at the signs and symptoms of dengue fever and make accomplished estimates about what will happen in the following years. Two different machine learning methods have been used to forecast what would happen: A support vector machine (SVM) and a random forest classifier were employed in this situation. Finally, we analyzed the performance of these two employees and presented them in a confusion matrix.
Appice, Gel, Iliev, Lyubchich, and Malerba (2020), in their research, demonstrated that in tropical nations, a significant public fitness issue is dengue disease, which is transmitted by mosquitoes and, therefore, extra disposed to it occurs (Appice et al., 2020). Comprehensive machine learning techniques analyze temperature and dengue data. A year's worth of temperature data is then used to predict dengue fever's prevalence accurately. It is possible to extract the temporal dynamics from historical data using various techniques. The forecast is created with the nearest neighbor predictor based on trend association. A case study has examined the number of dengue and dengue hemorrhagic fever cases in Mexico's 32 federal states from 1985 to 2010. According to the empirical research, the proposed strategy is a good one, according to the results of the experimental study.

Xu et al. (2020) developed an efficient and effective dengue forecasting model using LSTM recurrent neural networks (Xu et al., 2019). Only monthly dengue cases and climate conditions were considered in the study. It was shown that LSTM models performed better than other models in the past when forecasting DF cases one month into the future. The LSTM model outperformed the other forecasting models used in predicting dengue cases. Additional advantages of using transfer learning (TL) include making the model more scalable in locations where Dengue is less prevalent. For Dengue and other comparable diseases, the data could improve projections.

Salami, Sousa, Martins, and Capinha (2020) displayed that Dengue fever is a global public health concern when it spreads (Salami et al., 2020). Dengue's spread from locations where it is frequent to areas where it isn't common has increased in cases. This is due to the expansion of the worldwide air transportation network. Dengue importation is challenging to predict due to the dynamic nature of the air transportation network, which includes a wide variety of aircraft—trained four machine learning classifier algorithms using a six-year history of dengue imports from 21 European nations. The imports and the centrality of the air transportation network were also measured using connectivity indicators. It determined how accurate the classifiers were using the receiver operating characteristic curve, sensitivity, and specificity. Predictions made by our more precise model were spot-on. It had a maximum sensitivity of 0.88 and an area under the receiver operating characteristic score of 0.94. Dengue rate, population size, and the number of air passengers were all considered in determining which characteristics were most significant. Additionally, measurements of network centrality, which reflect how European countries fit into the air transport network, were used to make the forecasts. A machine learning algorithm performed exceptionally well in estimating the number of persons who contract Dengue. Additionally, working model-agnostic methods to explain how the predictions were made. – Making a reliable early warning system for dengue imports can be done using similar techniques.

Insects are now responsible for transmitting more cases of dengue fever than any other pathogen by (Davi et al., 2019). It is well established that laboratory and clinical methods used to determine dengue phenotypes are inherently flawed. Based on human genome data, demonstrate how to apply machine learning to predict the severity of dengue illness. Methods: 322 single nucleotide polymorphisms in innate immunity were tested on 102 dengue patients and 102 healthy individuals who did not have the virus (SNPs). Dengue and severe dengue patients are classified using a support vector machine method based on the best possible set of loci. After that, patients are divided into two groups based on the output of an artificial neural network (ANN). 13 immunological SNPs were used to train the ANN, which had a median accuracy of more than 86% and sensitivity and specificity of more than 98% and 51%, respectively, in dominant or recessive models. This approach can be used to identify persons who have a high chance of having a severe dengue-like illness, even if they have not been infected. Other Mendelian-based and hereditarily prejudiced disorders can be deliberate using this approach.

When a disease extent rapidly, it is referred to as a widespread discussed by (Dhaka & Singh, 2020). There is a great deal of tension in society due to an outbreak (Dhaka & Singh, 2020). Medical monitoring could benefit from a system that warns individuals when an epidemic occurs. Insects are commonly responsible for the propagation of epidemic
diseases. These "vectors" contain pathogens. These pathogens then propagate the disorders. Alerting systems may determine how the weather and other conditions influence the proliferation and spread of these vectors. High-tech sensors and satellites have made it possible to anticipate the weather. Support Vector Regression and Multiple Linear Regression and the Random Forest Regression and Decision Tree Regression methods discussed above are employed in this system. Dengue data from 2013 to 2017 was incorporated into the system. Using data from 2013 to 2016 could anticipate what will happen in 2017. The model has also been trained using data from 2013 through 2015. Chikungunya forecasts for 2017 have been generated. As a final step, a comparison of the four algorithms utilized for both disorders has been made.

Macedo Hair, Fonseca Nobre, and Brasil (2019) demonstrated that it makes no difference that the WHO's revised clinical classification for dengue fever, proposed in 2009, is more sensitive. To adequately communicate the warning symptoms and progression of dengue illnesses, there is still a need for a better method of communication. Researchers have utilized traditional statistical approaches to examine risk factors in dengue patients. As a result, individuals can often not appreciate the full scope of their medical histories.

Dorigatti et al. (2018) discovered that persons who had previously been exposed to Dengue had a different level of vaccine-induced protection from virologically proven Dengue throughout active surveillance (0–25 months) [9]. Understanding how efficacy changes by serostatus and age in endemic places is critical to understand why vaccinated people have a higher risk of being hospitalized over time and to develop safe and effective vaccination strategies for dengue patients.

Kannimuthu et al. (2019) presented that Dengue is a deadly disease transmitted by feminine mosquitoes. It carries a high risk of injury or death. Dengue fever, commonly referred to as breakbone fever, is a contagious illness that can be extremely dangerous if not treated promptly. Aedes mosquitoes carry and transmit four distinct viruses that cause Zika. It’s critical to develop a reliable strategy for forecasting dengue fever. Without needing to be explicitly programmed, AI is a technology that allows robots to learn and grow from their own experiences. Artificial Intelligence (AI) is a sub-category of machine learning. This study examines the effectiveness of various models to predict dengue sickness.

According to the World Health Organization, more than 2.5 billion people are in danger from mosquito-borne diseases by (Pham et al., 2018). This year's increase in dengue outbreaks has wreaked havoc on the people of Malaysia. As a result, detecting a dengue outbreak and acting swiftly to prevent its damage and loss is exceptionally vital. This study investigates whether Kuala Lumpur’s dengue fever cases may be predicted using machine learning and deep learning (DF). Our investigation discovered that a wide range of factors could induce dengue outbreaks. We also used the EVI (improved vegetation index), humidity, and wind velocity as inputs for our prediction engines. From 2002 to 2012, Kuala Lumpur residents reported DF incidences daily. There were daily incidences of DF from 2002 to 2012. Collected data on these aspects and cleaned it up from 2002 to 2012. The GA RNN model was the most effective of the three. It had an MAE of 10.95 when forecasting how many DFs generated.

Insects are the most prevalent vectors of transmission for this disease. Dengue Fever (DF), carried by mosquitoes, poses a threat to more than a third of the world's population. If a vaccine or cure isn't available, accurate forecasting of dengue transmission will aid in public health efforts to keep it under control. DF extent was predicted using four dissimilar models: a linear one three nonlinear ones (support vector and random forest), all of which were scrutinized in this study by (Siddiq, Shukla, & Pradhan, 2021). For the prediction models, the DF cases in Jeddah, Saudi Arabia, were used both the temperature and humidity, which are critical in accurately forecasting confirmed cases, were taken into account. Support Vector Classification (SVC) performed the best among the models evaluated. Compared to linear regression's 52% accuracy, followed closely by random forest's 55% accuracy and then decision trees' 57% accuracy, this model was 76% correct.
3. Assembling and Categorizing Patient Data

Gathering information and determining patient status to collect data revolved around the medical records of each patient. A categorical variable named "age group" was utilized in the exploratory analysis to understand better how the variables were distributed among children and adults. "Age group" was the name given to this variable. Age and days were normalized before creating the similarity matrix to reduce their variability. Complete blood count analyses and imaging data to illustrate how much fluid is present in the cavities were derived from at least two separate tests of hematocrit and platelet changes, respectively (Nan et al., 2018). These factors were utilized to create some warning flags. The clinical classification was based on a WHO recommendation. Trained professionals did it.

Dengue without warning signals, Dengue with cautionary signs, and severe Dengue were all classified. For more than an hour, a person might vomit more than five times a day or more than three times an hour. They also have a pleural effusion and ascites, which can be seen by the vesicular murmur or the diminished thoracic-vocal trill. Additionally, aberrant imaging findings may reveal abdominal distention or dullness decubitus. Two lab findings needed to be checked: thrombocytopenia (platelet count, 50,000/mm3) and hematocrit change of 20%, either raised or dropped by 20% throughout the convalescent period. The following criteria defined a severe case of Dengue. With respiratory difficulty, shock or fluid accumulation might result from plasma leakage. Classified shock as having two or more of the hypo perfusion-related symptoms, such as low blood pressure (less than 20 mmHg) or hypotension age as part of the picture (Akhtar, Pasha, & Rehman, 2020; Natali, Babrak, & Miho, 2021). The diagnosis and clinical criteria for these signs and indications should be discussed elsewhere.

4. Methodology

Located in the middle was Cluster 2. Dengue fever and severe warning signs were the most common symptoms, but it also had characteristics in common with the low-risk clusters (myalgia, arthralgia, and discomfort behind the eyes). Vomiting, mucosal draining, higher WBC count, and ALT levels, and lower platelets were reported in senior individuals with Dengue during the critical period. In Cluster 2, you'd find people of all ages, from infants to seniors. This cluster had a wide range of warning indicators, which may have contributed. Figure 2 demonstrate the proposed approach for dengue fever.

![Figure 2: Proposed approach for dengue fever](image)

Dengue clinical classification is based on 30 variables in the patient’s clinical profile up to seven days after becoming ill. The average age was 31 years old, and 49.1 per cent of the population was female. 31% of those polled were children, while 69% were adults. In 35 percent of the patients, dengue serotypes were detected in Figure 3.
Only 93 patients were detached because they did have not good laboratory data. They had extra illnesses, 55 patients went additional than eight days after the indications in progress, and 76 patients couldn’t be standard because they had an clinical demonstration to be considered as Dengue showed in Figure 4. There were 654 confirmed cases of Dengue as a result of the outbreak.

There were 523 participants in this study, and their serotypes are listed. Nausea and sickness were reported by 54.2% of children and 78% of progenies, the latter being more shared in broods. The patients also present with Petechia, clinical fluid accumulation (pleural effusion/ascites) in figure 5, liver enlargement of more than 2 cm, and a history of abdominal pain (22.6 per cent). More adults (87%), more retro-orbital eye pain (73.1%), and extra arthralgia (76.8%) were affected by myalgia than posterities, but the number of cases was minor in grown-ups. According to a table, these symptoms and indicators appear in persons of all ages.
The stochastic SOM approach employed the dissimilarity matrix generated by the unsupervised random forest. In the end, the best grid was selected because of its high energy density and the high number of topographical and quantization mistakes. Figure 6 represents the quantization error cluster to ensure that the grid had minimum topographic (0.05) and quantization (0.43). It was chosen. The grid's lowest topographic inaccuracy preeminent represents the network's data structure; hence it established the most consideration.

According to (Kondeti et al., 2019), the conditions for the age shift are set one by one and establish the change from appearance. Alfred and Obit (2021); (Halim et al., 2019) Capable to do this by using an age-segregated seroprevalence dataset to estimate the force of dengue infection. Figure 7 displays the selected particular cluster.
The hospitalization rate was high in this specific cluster, as shown in Figure 8. It was necessary to separate this group from the others based on warning signals such as fluid accumulation in the body, pain in the abdomen area, leukopenia, lethargy, liver enlargement, restlessness, and a rise in hematocrit.

Thrombocytopenia (reduction in platelets) was also critical. Also, several symptoms were more prevalent in children under 18 than adults who had them. According to Wakimoto et al., abdominal pain, bleeding, tiredness, and liver enlargement were all connected to severe Dengue in children (Siddiq et al., 2021). Figure 9 demonstrate the total
cases scatter plot analysis findings confirm the necessity of monitoring these things in dengue-infected children.

Figure 9: Total cases scatter plot analysis

More children were with warning symptoms in this study and fewer older persons. More research is required to fully understand why people of this age have such disparate medical characteristics. Figure 10 represents the Scatter plot analysis and forecasting dengue fever cross-sectional design, and the main drawback was the utilization of data that had already been gathered. Clinical symptoms may not have been thoroughly documented, especially in less severe cases, despite patients’ actuality tracked throughout time. As a result of the decreased frequency of the variables strap positive test, tapered pulse pressure, reduced awareness, and AST/ALT levels, it’s vital to focus on these four variables.
This study's wide variety of clinical profiles (ambulatory/hospitalized patients, adults/children) was beneficial; it helped us better understand how the complete spectrum of Dengue clinical manifestations may be classified. Arbovirus clustering in a city like Rio de Janeiro is an issue that needs to be addressed. Figure 11 shows a yearly analysis of the dengue fever scatter plot. It was possible to see how Zika and Chikungunya looked in humans before spreading to the United States.
Figure 11: Yearly analysis of dengue fever scatter plot

More than 60 percent of the patients in the high-risk group were under the age of 10 in our study, indicating that in dengue-infested countries like Brazil, the first thing to look at is his age. Most of the patients in the high-risk group were under the age of forty and exhibiting signs or symptoms of shock.

5. Discussion and Conclusion

Dengue has a broad spectrum of symptoms, and clinicians have been unable to forecast how the disease will progress because of the variety of instances and the severity levels that overlap. Their clinical profile can identify people with Dengue, and computer tools can identify trends. Age is the most critical factor to consider, according to our research. In children, keep an eye out for symptoms like stomach pain or tenderness, clinical fluid buildup, mucosal bleeding, lethargy and restlessness, liver enlargement, and increased hemoglobin concentration. Doctors and pediatricians may be able to forecast how severe a dengue illness will be in the future if further research is done. According to these studies, age should be the first factor to consider in areas where dengue fever is prevalent. In children, warning indicators are fundamental. In the future, other research may provide light on the whole range of dengue symptoms by examining these results over time. According to this study, dengue outbreaks can be predicted using machine learning. A dengue forecasting model could benefit from boosting or utilizing nature-based algorithms in the future.
References


