Disease Prediction in Real Time Based on Climate Data

Dr.Nidha Z D Shafee[1], Dr.V.Sundaravel[2], Dr.Isaac Christian Mose[3] , Dr.Mohamed Musthafa[4*] , Panneerselvam periasamy[5], Dr.Sasikala Gunasekaran[6]**

1. Clinical practitioner (Physician), First street, Alameen colony, South ukkadam, Coimbatore. nidhazuwairya@gmail.com
2. Professor, Department of medicine, Nandha Medical College and Hospital, Erode, Tamil Nadu. dr.sundaravel@yahoo.co.in
3. The Dean, Professor, Department of Medicine, Nanda Medical College and Hospital, Erode, Tamilnadu. drisaacmoses@yahoo.co.in
4. Professor & HOD, Department of Pharmacology, Nandha Medical College and Hospital, Erode, Tamil Nadu. musthafaseik@gmail.com
5. Assistant Professor, Department of Physiology, Government Erode Medical College, Perundurai, Erode, Tamilnadu, India. pannphysio@gmail.com
6. Department of Nursing, Government Erode Medical College Hospital, Perundurai, Erode, Tamilnadu, India. saipannsasi@gmail.com

*****Corresponding author:**

Dr.Mohamed Musthafa**,** Professor & HOD, Department of Pharmacology, Nandha Medical College and Hospital, Erode, Tamil Nadu. musthafaseik@gmail.com

Abstract: The weather can have a major impact on people's lives and welfare, even if no one can do anything to control it. disasters brought on by global warming. Furthermore, we are more susceptible to infectious diseases as a result of climate change. Global warming can affect infectious diseases in three main ways: the pathogen itself, the vector or host that spreads it, and the rate at which it spreads. The effects of climate change can be felt in all three of these ways. World Health Organisation (WHO) reports that 30 new diseases have been identified since 1996. Furthermore, arthropod-transmitted diseases including dengue fever, malaria, and chikungunya have made a resurgence and are rapidly expanding into formerly unaffected areas of the world. Factors like as precipitation, humidity, temperature, and others have various purposes in the transmission cycle. There is a chance that certain diseases will become more common if any of these conditions are changed. Early warning systems (EWS) for infectious diseases must account for the consequences of climate change, as evidenced by the rapid global spread of abroviral infections. In prior investigations, only weather records from the past were taken into account. The increasing levels of climate variability and unpredictability, on the other hand, are leading to the decline of traditional systems. In this research, we provide a method for predicting the number of new dengue cases over the next five days based on nothing but the weather forecast. Our findings have global relevance, and they pave the path for future research into specific daily disease outbreak scenarios through the use of real-time meteorological API. The purpose of this is to safeguard humanity from climate-related health hazards.

Keywords: Climate, Weather, Disease, Real-Time, Dengue, Early warning system

I. INTRODUCTION

In 1996, the World Health Organisation recognised a total of thirty newly discovered diseases. In addition to this, diseases that are transmitted by arthropods, such as dengue fever and malaria, have made a comeback and have spread to new locations. There is a correlation between the frequency with which extreme weather events take place and the shifting of weather patterns, which in turn is connected with the frequency with which mosquito-borne illness epidemics take place. There has to be a virus, someone who is vulnerable to the disease, and a carrier of the virus for there to be an infectious disease. In addition, for these components to grow, reproduce, and spread farther, a particular climate is required. As a result, fluctuations in the weather and environment may have an effect on the pathogens, vectors, and transmission contexts of vector-borne illnesses.

The likelihood of disease epidemics, such as the human abroviral disease known as dengue, which is currently rather common, is growing as a consequence of extreme weather events and ongoing changes in the weather. According to the World Health Organisation (WHO), dengue fever affects between 50 and 100 million people annually, with over 500,000 of those affected requiring hospitalisation. Rising temperatures are contributing to an increase in the incidence of dengue disease. Alterations in weather patterns, including precipitation, temperature, and humidity, have the potential to have a significant impact on the transmission of diseases that are carried by vectors. During the gonotrophic stage of the dengue virus life cycle, the temperature of the surrounding environment has an effect on the feeding habits of adult dengue vectors, the degree to which the virus can be imitated, and the length of time it takes for dengue virus larvae to develop. According to one school of thought, dengue vectors cannot procreate in the absence of stagnant puddles of precipitation. Changes in climatic elements such as precipitation, temperature, and wind speed can have a significant impact on the rate and scope of the spread of an infectious disease. These factors can also play a role in the transmission of the disease.

In the context of global warming, recent pandemics have demonstrated how important it is to develop an early warning system (EWS) for infectious diseases. In previous studies, the sole data source utilised was historical weather records. However, because of the increasing unpredictability and fluctuation of the environment, the conventional methodologies are becoming less useful. Predicting the future state of the atmosphere in terms of its attributes such as humidity, temperature, precipitation, and so on is one of the most difficult problems that scientists are currently seeking to answer. It is also one of the most important topics. In the past, accurate disease forecasting involved making informed estimates about where an epidemic would begin and how elements such as temperature and geography might influence the later spread of the disease.

As a result, models that are based on the circumstances of the weather help in determining when and where human cases are likely to occur. It is possible that the disease burden could be decreased if control efforts and limited preventative interventions were correctly focused using such data.

For this reason, we provide a Real-time Disease Prediction system that, based on an analysis of API keys and the current weather conditions, may provide an estimate of the number of dengue cases that will occur over the course of the following five days. Our group carried out this investigation into disease prediction in real time by utilising sufficient meteorological data and a technique to supervised machine learning known as Random Forest. Our investigation consisted of multiple steps, each of which will be discussed in further detail in the following paragraphs. In section 2, we will explore prior efforts in relation to weather patterns and illnesses.

In Section 3, we provide a brief review of our issue statement, and in Section 4, we detail our research approach. In Sections 5, 6, and 7, the findings and conclusions of the study are broken down and analysed in further detail.

II. RELATED WORKS

We focus on a few of the studies that have been done to establish the connections between weather and sickness. There have been reports of the appearance of vector-borne diseases at high elevations in various different regions of the world, which has caused concerns for the general public's health. The topic of climate change has garnered a lot of interest from scientists in recent years due to the fact that insects and other small species are easily impacted by it. There are two primary classifications that these probes can be placed into: The theoretical studies that establish the parameters of the influence that weather has on human health are complemented by the empirical research that has been conducted on the subject.

In the following, we will give a concise summary of some of the efforts that have been done to attempt to put a number on the consequences that climate change will have on human populations.

There have been a number of studies done on how the changing climate influences the transmission of diseases that are carried by mosquitoes. For example, it has been demonstrated in [1] and [2] that greater temperatures encourage the proliferation and spread of viruses. In Botswana, those in charge of making decisions regarding malaria were assisted by the findings of a forecasting system[3], which improved their attempts to more effectively allocate resources.

It has been discovered by researchers that temperatures ranging from 27 to 30 degrees Celsius encourage the growth of viruses [4].A thermodynamic model that is given in Reference [5] demonstrates that taking into consideration short-term temperature changes is an effective way to improve one's understanding of the kinetics of disease transmission. According to the findings of a plethora of studies that investigated the health repercussions of global warming and altered climate patterns, long-term warming has a tendency to influence the regional spread of various infectious illnesses by producing the chance for more clustered disease outbreaks. This was discovered by researchers who were looking into the health implications of global warming and altered climate patterns.[6][7][8].

Using specialist empirical models, it has also been possible to determine whether or not climate has a role in the dissemination of specific infectious diseases. A number of researchers have investigated the use of time series analysis to determine whether or not it could be useful in disease prediction. According to the Poisson regression analysis of the link between meteorological variables and human plaque occurrences in Reference [9], plaque risk can be predicted by using temperature and time-lagged amounts of late winter precipitation. This is shown by the analysis of the association between meteorological variables and human plaque incidences. For the period between 2000 and 2010, the prediction of dengue incidence in Singapore was carried out with the use of a time series Poisson multivariate regression model. They did an analysis that accounted for seasonality, autoregression, trend, and different lag periods to find the optimal time to produce dengue forecasts [10], using only cumulative rainfall and weekly mean temperature as independent parameters. This allowed them to determine the ideal time to create dengue forecasts. Analysis of time series has been suggested for use with additional vector-borne diseases and geographical areas. Recent research, such as that which is given in References and so on, provides a wealth of knowledge on the influence of weather and climate on many diseases, as well as the considerations that should be addressed when formulating health policy, etc.

In addition, Support Vector Machine, often known as SVM, has been put through its paces by a select group of researchers for the aim of disease prediction.

III. PROBLEM STATEMENT

This project aims to develop a real-time Early Warning system that can anticipate disease outbreaks (specifically, dengue epidemics) by studying the upcoming weather forecast for the next five days.

Due to the complexity and non-linearity of the underlying data (the weather), traditional approaches are wasteful and fruitless. The proposed method assesses the developed models by altering several sets of parameters in order to make an instantaneous prediction of the disease (here, Dengue). This is accomplished with the help of a five-day weather forecast. Models can be chosen based on their performance as measured by the root mean square error.

The data model and method proposed in this publication outperformed (in terms of computational complexity) and improved upon the findings of previous research in this area. This leads the researchers to conclude that climate change is a primary contributor in recent dengue outbreaks. Because of this, the period of time during which dengue infection is most dangerous may be reduced.

IV. METHODOLOGY

The processes that will be followed in order to carry out the tests that are required by our investigation are going to be described in great detail in this part. Following the presentation of the dataset that was used to construct dengue case predictions based on real-time meteorological conditions, a discussion on the methodologies that were utilised to produce these forecasts took place.

A. Data

Our major goal is to use the current and predicted weather in the Philippines to make an estimate of the number of dengue cases that will occur there. To do this, we used a merged dataset comprised of information from the following places: In the first, we'll see how past weather patterns have correlated with outbreaks of Dengue fever, and in the second, we'll utilise an API key to construct a five-day weather forecast. The first piece of data came from Kaggle, a website that runs competitions for predictive analytics, and the second from Open weather map, a website that provides weather data for any location in the world in exchange for an API key.

Here are the procedures we went through to build a prediction model using this particular collection of data:

B. EDA

Exploratory data analysis is always the first stage in data analysis. In its simplest form, this strategy uses a toolbox of methods to solve a wide range of problems, such as detecting outliers, identifying missing data, maximising insights within a dataset, discovering underlying linkages, etc. These methods—mostly quantitative but also graphical—let the data speak. Visualisations help data analysts convey their knowledge to non-experts. EDA was used to investigate potential correlations between our dataset's characteristics and variables. Scatter plots helped us understand how temperature and humidity affect dengue cases. This helps us understand dengue and weather. A bar graph (identical to Fig. 1) shows the Philippines' monthly dengue case counts for the past two years. This basic, handy stacked bar chart shows the monthly dengue case counts and the yearly sums. The top row provides yearly totals and the bottom row monthly case counts. Except for July, August, and September, 2009 had more dengue fever cases than 2010.This data forces us to review our three-month database. Please note that the monthly dengue cases for 2009-2010 are being used as an example to investigate if these variables are correlated. After gathering context for our dataset, we will move on to data purification. This topic is examined more below.

Figure 1. Dengue fever cases in the Philippines are displayed monthly in a bar graph.

C. Data Cleaning

Screening the data is done so that the information used in the analysis is as "precise" as it can be after the data has been collected. The goal of this procedure is to identify and fix any information that is either missing, wrong, or unnecessary. Using insufficient or wrong data can lead to a number of problems, one of which is the drawing of incorrect or inadequate conclusions. Therefore, data cleansing must be incorporated into every step of the data analysis process.

It is not uncommon for a given observation to have "missing data," which is a data value that is empty when it is saved for a particular variable, while dealing with problems that are grounded in the actual world. Data cleansing is an essential part of the data exploration and data preparation process since the existence of missing values in a dataset can have a significant effect on the data insights and the performance of our prediction model. Because they cause so much distress, they must be carefully monitored and controlled if we are to eliminate prejudice and develop trustworthy models. How to deal with circumstances where data is missing is an important question. Some ways to deal with them are to erase the data, to impute it, to use a prediction model, or something similar. Since our dataset also had missing values, these techniques were used to complete it. When there were blanks in the minimum and maximum temperature columns, for instance, we filled them in with the corresponding values from the previous three days.

The amount of snowfall over the previous three days was used to make up for the missing data point in the snowfall column. Similar efforts were made to address any data gaps that might have existed.

Our second method of analysis consisted of using feature engineering. Using this technique, extra information was culled from the supplied dataset. The first part of the procedure alters the existing variables, while the second stage introduces brand new ones. The first technique is used to modify an existing variable's size, distribution, or relationship, while the second is employed to produce novel variables out of the existing ones. This necessitated the creation of new variables such as L3_temp (for the average temperature over the past three days), L3_rain (for the average rainfall over the past three days), etc. It is expected that these newly generated variables will have an effect on the dengue fever plagued region in addition to the existing climate elements. As a result, the strategy aided in illuminating a connection between two variables that was obscure before to its exposure.

Following these procedures, we were able to extract useful insights from the raw dataset we had. Now that we understand the setting, we can talk about applying machine learning algorithms to our dataset.

V. RESULTS

To see how well our reference model predicted dengue cases across climates, we examined data generated using the Random forest technique. This was accomplished with the help of the method's data.We then use this score to determine the following model metrics:-Hold off on merging text and images until after you've finished designing and formatting the remainder of the content. Hard tabs are not allowed; only returns should be used at the end of paragraphs. You're really welcome. Page numbers should not appear in the text of your document. Don't worry about manually numbering the sections; the template will take care of that for you.

Variable Importance Plot

This illustration highlights the fundamental characteristics that are intrinsically linked to the variable of interest and account for the great majority of the variance in the dependent variable. This graph is provided for your consideration in Figure 2.Our data shows that the month variable, sometimes called the factor variable, has the most influence on our dengue prediction analysis. This is to be expected, but it must be stressed that air pressure and rainfall are the two most significant meteorological parameters that affect the analysis. Instead of predicting whether the number of dengue cases would rise or fall, these metrics recommend which factors should be prioritised when doing so. They do not suggest either an increase or decrease in dengue cases.

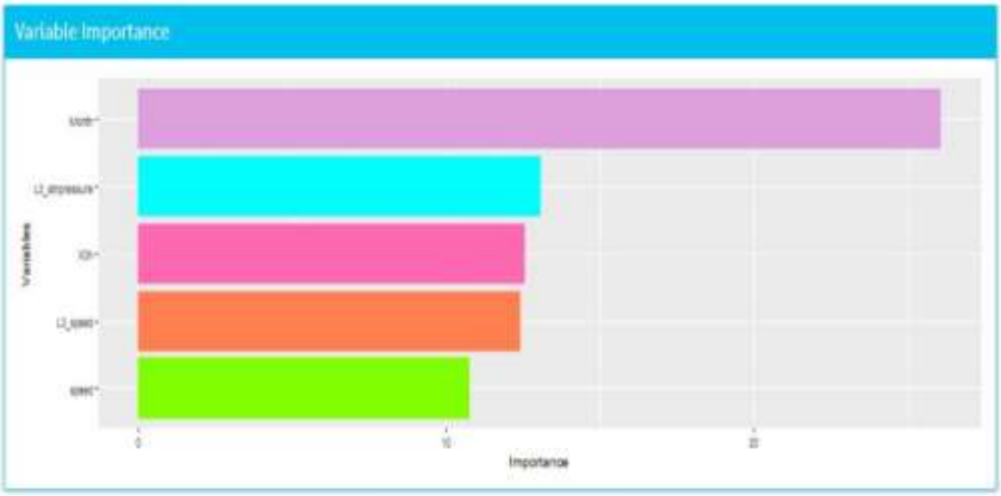

Figure 2: Plot of Variable Significance

Our approach to forecasting dengue cases in the Philippines using weather information from the prior two years is outlined above. Following the same procedure, we will estimate the number of occurrences in light of the weather forecast for the next five days. From today forward, we'll be talking about the upcoming five-day weather forecast for the Philippines, starting with today itself. Remember that the weather data is continuously updated in real time. If we were to repeat this study now, it would automatically factor in the weather forecast for the next five days. If a similar analysis were performed after a month had passed, the results would be different due to the meteorological conditions that existed during the five days this model was being run. As a result, we are forecasting dengue events on a Google map using the aforementioned algorithm and current weather conditions for the next five days. For the next five days, the forecasts will hold true.

Both the present weather in a city overlaid on a map and a forecast of dengue cases over the following five days are displayed in this real-world prototype. The prototype dashboard for the real-time model is shown in Figure 3.

Figure 3: Live Demo model

Today (the day on which we most recently ran this model) there will likely be four new cases of dengue fever, with a range of five to eight over the following four days, as shown in the accompanying chart. Due to the ever-changing nature of a real-time model, the data we collect is constantly up-to-date.

VI. CONCLUSION

Climate change has sped up the spread of infectious diseases around the world. Viruses, hosts, and disease vectors can all be influenced by the weather, which in turn affects how diseases spread. To safeguard society against the health risks associated with climate change, we developed an early Warning system (EWS) that correlates infectious diseases with weather events. This initiative opens up a direct line of communication between experts in the field, researchers in the academy, and policymakers concerned with the current and future impacts of climate change on human populations. As a result, it improves resilience and mitigates risks to health in real time. This research has implications beyond just the areas and diseases that this study focused on. Global in scope, this study utilises a real-time weather API to examine comprehensive daily disease outbreak scenarios and pinpoint the optimal time to implement control measures.

REFERENCES

1. K. Nguyen and Y. -A. Liou, "An Approach for Risk Maps of Vector-Borne Infectious Diseases: Ecological and Adaptive Capacity Indicators," *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, Valencia, Spain, 2018, pp. 1164-1167, doi: 10.1109/IGARSS.2018.8519191.
2. Lu Liang and Peng Gong, "Climate change and human infectious diseases: A synthesis of research findings from global and spatio-temporal perspectives", *Environment International*, vol. 103, pp. 99-108, 2017.
3. Xiaoxu Wu, Yongmei Lu, Sen Zhou, Lifan Chen and Bing Xu, "Impact of climate change on human infectious diseases: Empirical evidence and human adaptation", *Environment International*, vol. 86, pp. 14-23, 2016.
4. James N. Mills, Kenneth L. Gage and Ali S. Khan, "Potential Influence of Climate Change on Vector-Borne and Zoonotic Diseases: A Review and Proposed Research Plan", *Environmental Health Perspectives*, vol. 118, no. 11, pp. 1507-1514, 2010.
5. Costello, M. Abbas, A. Allen, S. Ball, S. Bell, R. Bellamy, et al., "Managing the health effects of climate change", *Lancet*, vol. 373, pp. 1693-1733, 2009.
6. Man De, H. Van, H.H. Berg, E.J. Leenen, J.F. Schijven, F.M. Schets, et al., "Quantitative assessment of infection risk from exposure to waterborne pathogens in urban floodwater", *Water Res.*, vol. 1, no. 48, pp. 90-99, 2014.
7. Christoph Aubrecht and Dilek Ozceylan, "Identification of heat risk patterns in the U.S. National Capital Region by integrating heat stress and related vulnerability", *Environment International*, vol. 56, pp. 65-77, 2013.
8. "Assessing human vulnerability due to environmental change: concepts issues methods and case studies", *Nairobi Kenya: United Nations Environment Programme*, pp. 57, 2002.
9. J. Takahashi, "Fuzzy Database Query Languages and Their Relational Completeness Theorem", *IEEE Transactions on Knowledge and Data Engineering*, vol. 5, no. 1, pp. 122-125, Feb. 2003.
10. R. R. Colwell, "Global Climate and Infectious Disease: The Cholera Paradigm", *Science*, vol. 274, December 2006.