



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume:** 11    **Issue:** XI    **Month of publication:** November 2023

**DOI:** <https://doi.org/10.22214/ijraset.2023.57226>

[www.ijraset.com](http://www.ijraset.com)

Call:  08813907089

E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)

# Dengue Spread Prediction Using Time Series Model

Shushant Shrama<sup>1</sup>, Shikhar Sharma<sup>2</sup>, Deepesh Mishra<sup>3</sup>

Department of Computer Science and Engineering, Galgotias College of Engineering and Technology, India

**Abstract:** *The relentless spread of dengue poses a critical challenge to global public health. In response, this research project introduces an innovative framework that combines time series analysis and vision-based methodologies to forecast the spread of dengue. The motivation stems from the imperative to proactively mitigate disease outbreaks, allocate resources judiciously, and implement effective preventive measures. This project introduces an innovative framework that combines time series analysis and vision-based methodologies to forecast the spread of dengue.*

*Motivated by the imperative to proactively mitigate disease outbreaks, allocate resources judiciously, and implement effective preventive measures, the project employs a meticulous methodology, beginning with an extensive literature review to identify key determinants influencing disease transmission dynamics. A comprehensive dataset is curated, comprising meteorological parameters, historical disease records, and sociolect-demographic data, ensuring data integrity through rigorous cleaning. The core lies in the development and evaluation of time series forecasting models, employing Python language and modules within the language. Integrating insights from both time series analysis and vision-based methodologies, the model advances disease forecasting, visualized through heat-maps for intuitive representation of disease risk areas.*

**Keywords:** *Dengue, Time Series Analysis, Vision-Based Approaches, Disease Forecasting, Spatial temporal Model.*

## I. INTRODUCTION

The relentless march of infectious diseases, particularly dengue, continues to pose formidable challenges to public health worldwide. In an age of heightened global mobility and inter connectivity, the rapid transmission of this disease demands innovative approaches for prediction and mitigation. The very essence of public health lies in the ability to foresee and control disease outbreaks, to allocate resources judiciously, and to implement effective preventive measures. It is within this context that this research project emerges, driven by a compelling motivation to harness the power of time series analysis for the precise forecasting of disease spreading rates.

The motivation underlying this research endeavor is deeply rooted in the imperative to safeguard human lives and well-being. The global landscape of infectious diseases is marked by continual shifts and transformations, driven by factors as diverse as climate change, urbanization, and international travel. These dynamics have ushered in a new era of uncertainty and complexity for public health professionals, necessitating a fundamental shift in disease surveillance and response strategies. The resurgence of once-controlled diseases and the emergence of novel pathogens have underscored the need for proactive, data-driven approaches. Accurate forecasting of disease spreading rates has become not just a scientific endeavor but a moral and societal imperative.

It equips healthcare authorities, epidemiologists, and policymakers with the foresight required to implement timely and effective interventions. By anticipating disease outbreaks, we gain a critical edge in the ongoing battle against dengue, arming ourselves with the knowledge needed to allocate resources efficiently, devise targeted vaccination campaigns, and safeguard vulnerable communities. At the heart of this research lies a meticulous and multidisciplinary methodology, designed to unravel the intricate web of factors influencing disease spread.

The project's inception commences with a rigorous review of existing literature, an exercise aimed at elucidating the multifaceted determinants of disease transmission dynamics. This foundational knowledge forms the bedrock upon which subsequent phases of the project are constructed. Simultaneously, the project embarks on the assembly of a comprehensive and meticulously curated data set. This data set comprises a rich tapestry of meteorological parameters, historical disease occurrence records, and relevant sociolect-demographic data. However, it is the data cleaning process that truly sets the stage for analytical precision. Ensuring data reliability is paramount, as any inaccuracies or inconsistencies may jeopardize the integrity of subsequent analyses. The model is trained to work on the data of Mumbai city in India.

## II. LITERATURE REVIEW

In 2012, Ruiz Moreno et al.[1] introduced a pivotal model aimed at understanding the transmission dynamics of the Chikungunya virus (CHIKV). This model, which has since become a cornerstone in the study of CHIKV, focuses on predicting the spread of the virus through the *Aedes* mosquito vector.

The research underscores the significance of comprehending the factors influencing CHIKV transmission, particularly in regions with susceptible populations. The model incorporates crucial variables including total population, land area, mosquito population, temperature, and the number of infected individuals. These inputs are drawn from publicly available data sources and carefully calculated or inferred to effectively implement the model. By utilizing the R programming language, researchers are able to generate weekly forecasts of infected individuals, enabling a dynamic and real-time understanding of the virus's propagation. Over time, the Ruiz Moreno model has undergone refinements and adaptations, broadening its applicability beyond its initial scope. Notably, it has been extended to encompass diverse countries and regions, particularly focusing on the Pan-American context. This adaptation is particularly critical in regions where CHIKV poses a significant public health threat due to limited immunity within the population. By offering a robust framework for modeling CHIKV transmission, the Ruiz Moreno model has become an invaluable tool in public health efforts aimed at mitigating the impact of the virus.

Its ability to generate accurate forecasts and adapt to varying contexts has made it an essential resource for researchers and policymakers alike in the fight against CHIKV.

In 2016, J.A.Smith et al[3], provides a comprehensive exploration of the Chikungunya virus (CHIKV) threat within the Pan-American region. The study begins by introducing the DARPA "Forecasting Chikungunya" Challenge, initiated in August 2014, aiming to engage researchers in predicting CHIKV's impact on populations with untested immunities, particularly in the United States. The report emphasizes individual-level prevention strategies and underscores the role of public health officials in promoting prevention.

The core focus of the report lies in extending the applicability of the Ruiz-Moreno model, initially proposed in 2012, to predict CHIKV transmission through the *Aedes* mosquito in any country or region, with a specific emphasis on Pan-American nations. The research methodology involves gathering data on parameters like population, land area, mosquito population, temperature, and infected individuals from publicly available sources.

However, infected vectors like mosquitoes present a more challenging scenario, as they can travel long distances through various means such as wind or passive mechanisms like inside airplanes or sea cargoes. To address this challenge, the PROVNA Project[4] ("Defining Eco-regions and Prototyping on Earth Observation (EO)-based dengueDisease Surveillance System for North Africa") was launched under the initiative of the World Organisation for Animal Health (WOAH). The project aims to assist North African countries in targeting their surveillance efforts towards Rift Valley fever by leveraging remote sensing and earth observation data. Annamaria Conte and Paolo Calistri's team at IZS-Teramo are at the forefront of this effort. They emphasize the importance of identifying ecoregions with similar environmental characteristics in North Africa, including temperature, vegetation, and soil moisture. By utilizing data from NASA and the European programme Copernicus, combined with information on animal populations and occurrences of Rift Valley fever, extracted from sources like the World Animal Health Information System (WAHIS) and FAO Empres-I, the team can identify areas at risk. The team combines spatio-temporal data to develop a prototype capable of predicting locations and timings of potential outbreaks. They consider past conditions of temperature, rainfall, and vegetation to forecast future risks.

A recent review by Sadeghieh et al. [5] found that current approaches to understanding and predicting VBD risk are typically focused on predicting risk in existing endemic zones (88% of VBD models in 1996–2016) rather than forecasting transmission risk in new regions. However, these approaches do not capture the myriad of climate impacts on vectors, hosts and pathogen seasonality which interact to shape patterns of VBD transmission [7].

In order to anticipate the establishment of Vector-Borne Diseases (VBDs) in marginal environments beyond their current ecological niche, innovative approaches are needed to predict the intricate and interconnected processes that govern these dynamics. Mathematical models offer a versatile method for predicting disease risk in marginal temperate environments. By directly integrating fundamental biological mechanisms, these models enable forecasts of the potential impact of climate change on novel combinations of environmental conditions, extending beyond those observed in endemic zones [8,9]. In doing so, we can investigate the likely impacts of predicted scenarios, such as the expected increase in the length of vector biting seasons with increasing temperatures.

### III. MYTHOLOGY

The subsequent sub-section, comprising Data Sources and Collection, Data Cleaning and Integration, Exploratory Data Analysis, Feature Engineering, Model Execution, and Model Evaluation, provides a detailed elucidation of diverse data sources and the process of collecting them. It delves into teleprocessing techniques implemented, including data imputation for handling missing values, integration of climate and health data, feature selection strategies, and outlier detection methods. Exploratory data analysis is conducted to unveil the correlations between climate parameters and dengue incidences, employing various visualization techniques such as heat maps and feature plots. The section further outlines the application of different time series and regression machine learning models, accompanied by insights into the chosen model evaluation metrics.

#### A. Data Sources and Collection:

Obtaining comprehensive and reliable data is foundational to our project. We will source our data from the Indian Meteorological Department, focusing specifically on the state of Maharashtra. To enhance granularity and precision, the state will be divided into five distinct regions. This regional breakdown is crucial for a more nuanced analysis, considering the diverse climatic conditions and geographical variations within Maharashtra. The Indian Meteorological Department is a reputable source, ensuring the data's credibility and relevance to our study. This step lays the groundwork for meaningful insights into the interplay between weather conditions and the occurrence of dengue disease. The dataset comprises mainly two sets of data: Climate data and Health data.

##### 1) Climate Data

Maharashtra exhibits diverse climatic regions such as Konkan, Khandesh, Desh, Vidarbha, and Marathwada. To account for the varying intensity of the disease and climatic conditions, Mumbai is selected as the sole city for this study. Monthly climate data spanning a decade from 2009 to 2019 is obtained from the Indian Meteorological Department (IMD). The parameters under consideration include Monthly Mean Maximum Temperature (MMAX) (°C), Monthly Mean Minimum Temperature (MIN) (°C), Total Monthly Rainfall (TMRF) (mm), Relative Humidity (RH) (%), and Mean Wind Speed (MWS) (km/h). Table 2 outlines the specific location of Mumbai within Maharashtra state, along with relevant population and climatic conditions.

Table 1. City Wise Population and Weather Condition.

Division Name	City	Population	Weather conditions
Konkan	Mumbai	125	Tropical wet, and dry climate
Konkan	Thane	18.9	Tropical monsoon climate
Konkan	Ratnagiri	3.27	Tropical
Pune	Pune	31.2	Hot semi-arid climate
Pune	Solapur	9.51	Dry, arid, and semi-arid climate
Pune	Satara	3.26	Tropical pleasant climate
Nashik	Nashik	14.86	Mid tropical climate
Nagpur	Nagpur	24.1	Tropical Savanna climate
Amaravati	Amaravati	6.47	Tropical wet, and dry climate
Aurangabad	Parbhani	3.07	Tropical hot. and wet climate

##### 2) Health Data

Monthly data on dengue disease incidence is gathered from the National Vector Borne Disease Control Program (NVBDCP) for Mumbai city in Maharashtra, as indicated in the climate data section. This information spans a decade, covering the years 2009 to 2019. The collected data, available in excel format, exhibits inconsistencies and missing values. Subsequently, the climate and disease incidence data for all nine designated cities are consolidated into a CSV file, and necessary preprocessing steps are undertaken. Figure 1 illustrates a map of Maharashtra state with cities categorized by region, with Mumbai specifically chosen for the study.

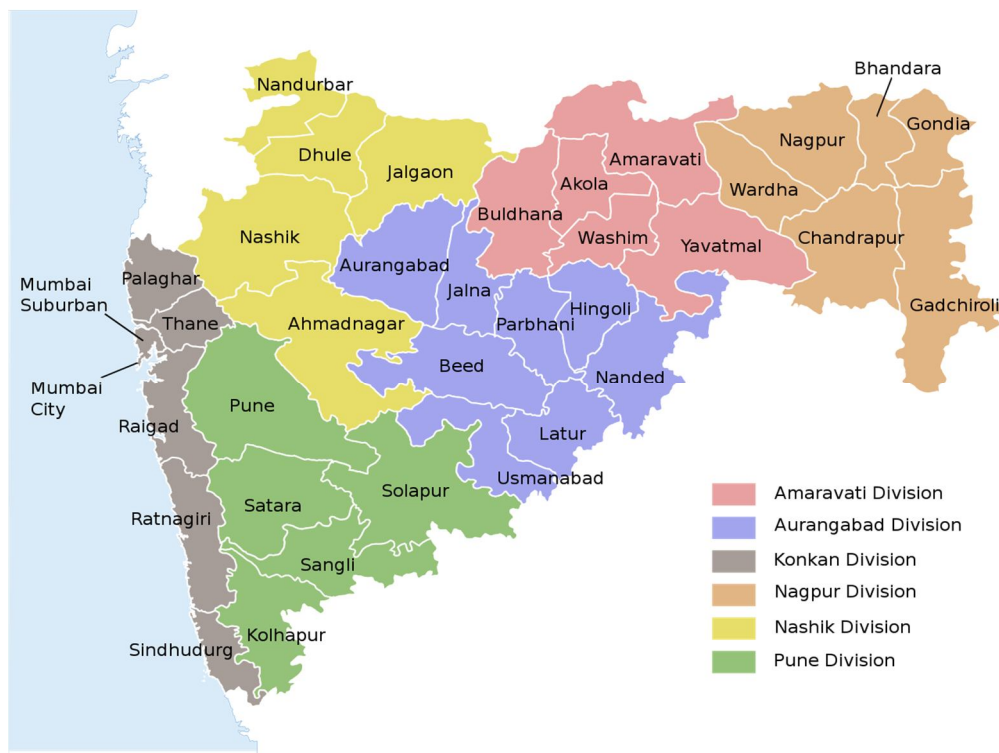


Fig 1. Map of Maharashtra with Region Wise Distinction.

### B. Data Cleaning and Integration

Once data is collected, a critical step is the cleaning and integration process. This involves addressing any missing values, outliers, or inconsistencies in the datasets. Ensuring data cleanliness is paramount to the accuracy of subsequent analyses. The integration phase involves merging different datasets if needed, creating a unified and coherent datasets. This meticulous process guarantees the reliability and quality of the datasets, providing a solid foundation for robust analysis and modeling.

### C. Exploratory Data Analysis

Exploratory Data Analysis (EDA) is the phase where we delve into the datasets to unravel its inherent patterns and characteristics. Through visualization and statistical techniques, we aim to understand the distribution of variables, identify trends, and recognize any outliers. EDA is not only a precursor to modeling but also a crucial step in gaining a deeper comprehension of the data. This step is pivotal for making informed decisions about feature selection, model choice, and the overall strategy for our project.

### D. Feature Engineering

Feature engineering involves creating new features or modifying existing ones to enhance the dataset's predictive power. This step is essential for optimizing the performance of machine learning models. In the context of our project, it may involve deriving additional meteorological indicators or refining existing variables to better capture the nuances of weather conditions and their impact on vector-borne diseases. Well-crafted features contribute significantly to the models' ability to generalize and make accurate predictions.

### E. Model Execution

With a refined datasets, we proceed to model execution. The model which will be implemented: Time Series.

- 1) *Time Series*: Time series analysis is a statistical method used to analyze data points collected, recorded, or measured over a sequence of time intervals. It is particularly relevant for data sets where the order of observations matters, and each data point is associated with a specific time stamp. Time series analysis aims to uncover patterns, trends, and underlying structures within the data to make predictions or derive insights into its behavior over time. The Time Series model is employed to capture temporal dependencies inherent in the spread of vector-borne diseases.

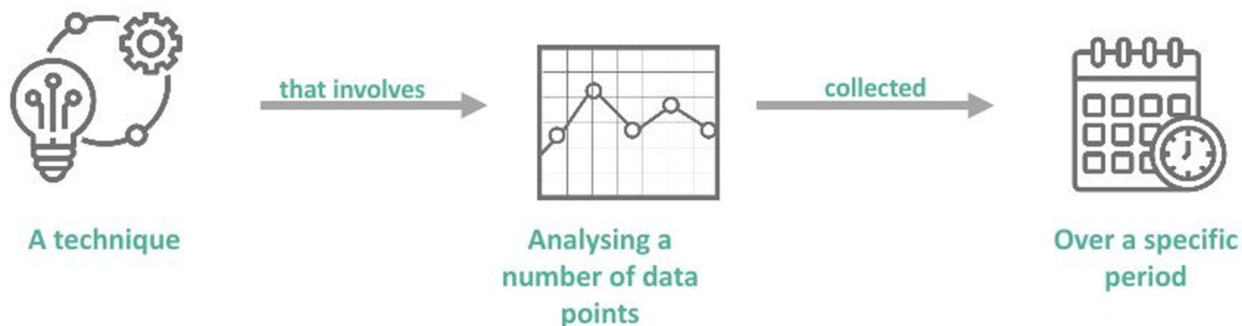


Fig 2. Time Series

This dual-model approach allows us to harness the strengths of both techniques and increase the robustness of our predictions.

**F. Model Evaluation:**

Evaluating the models is a critical step to understand their effectiveness. We employ well-established metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared to quantitatively assess each model's performance. These metrics provide a comprehensive view of how well the models align with the observed data. By comparing the results, we gain insights into which model offers superior forecasting accuracy, guiding us in the selection of the most effective predictive tool.

**IV. COMPARISON AND RESULT**

Based on the analysis of exploratory data, it was noted that each climate variable has distinct effects on dengue incidences in Mumbai. The average temperature range in Maharashtra state spans from 26 to 43°C. As illustrated in Figures 3, histograms generated post Pearson’s correlation reveal that mean maximum temperature (MMAX) demonstrates a negative correlation with dengue incidences in Mumbai, suggesting an increase in cases as the maximum temperature decreases. Mean minimum temperature (MMIN) exhibits a weak to moderate positive correlation with dengue incidences in Mumbai, except for Nagpur. Relative Humidity (RH), identified as the primary climate factor, displays a strong positive correlation with dengue incidences in Mumbai. Similarly, total monthly rainfall (TMRF) shows a moderate positive correlation with dengue incidences in Mumbai, indicating an increase in cases with rising humidity or rainfall. The peak incidences occur between June and September in Mumbai when the average rainfall ranges from 150 to 350 mm. Mean Wind Speed (MWS), considered a less significant climate factor, demonstrates a weak negative correlation with dengue incidences in Mumbai.



Fig 3. Correlation between climate factors and Dengue cases in Mumbai City

After training all regression and time series forecasting models, their performance is assessed using three evaluation metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R Square Error (R2).

Table 1. Performance matrix comparison table for Mumbai

Time Series Forecasting	RMSE	MAE	R <sup>2</sup>
Holt's Forecasting	2.22	2.09	-0.39
AR	1.95	1.41	-0.07
MA	1.93	1.29	-0.05
ARIMA	2.19	1.29	-0.35
SARIMA	1.98	1.92	-0.11
Facebook Prophet	3.14	2.1	0.38

The RMSE is a measure of the overall accuracy of a forecasting model. It is calculated by taking the square root of the mean of the squared errors between the actual and predicted values. The MAE is a measure of the average error of a forecasting model. It is calculated by taking the mean of the absolute errors between the actual and predicted values. The R2 is a measure of the goodness of fit of a forecasting model. It is calculated by taking the square of the correlation coefficient between the actual and predicted values. Based on the table, the MA algorithm has the lowest RMSE and MAE, indicating that it is the most accurate forecasting algorithm for the given data set. The ARIMA algorithm has the highest R2, indicating that it has the best fit to the data set. However, the ARIMA algorithm also has the highest RMSE and MAE, indicating that it is not the most accurate forecasting algorithm. Overall, the MA algorithm is the best forecasting algorithm for the given data set, as it has the lowest RMSE and MAE.

## V. CONCLUSION

In this paper the model proposed is driven by the synergy between data-driven machine learning and innovative problem-solving, marks a significant stride in the realm of disease surveillance and prediction. this project represents a comprehensive and insightful exploration into the intricate relationship between meteorological factors and the prevalence of dengue disease in Mumbai. By adopting a multifaceted approach, encompassing data collection, cleaning, exploratory analysis, and the implementation of sophisticated models, we have endeavored to provide a holistic understanding of the dynamics at play. Our findings promise to contribute significantly to the field of disease prediction and public health planning. The synergistic integration of meteorological data and advanced modeling techniques not only enriches our understanding of vector-borne diseases but also equips policymakers with a valuable tool for preemptive and targeted interventions. This research lays the groundwork for a more resilient and proactive public health strategy, fostering the well-being of communities in Mumbai and serving as a template for similar endeavors globally.

## VI. ACKNOWLEDGEMENT

We express our heartfelt gratitude to the researchers, mentors, and medical professionals whose expertise and guidance shaped this project. Special thanks to our project guide Prof. Arti Tiwari and project coordinator Manish Kumar Sharma and Sanjay Kakhil whose participation provided invaluable insights. Our families' unwavering support fueled our determination. Each contribution, whether big or small, played a vital role in our endeavour. This accomplishment serves as evidence of the collaborative spirit, for which we express our sincere gratitude.

## REFERENCES

- [1] Ruiz-Moreno D, Vargas IS, Olson KE, Harrington LC (2012) Modeling Dynamic Introduction of Chikungunya Virus in the United States. *PLoS Negl Trop Dis* 6(11): e1918. <https://doi.org/10.1371/journal.pntd.0001918>.
- [2] Saini, S. S., Singh, S. K., & Mishra, V. K. (2022). A review on prediction of seasonal diseases based on climate change using big data. This paper reviews the use of big data and climate change to predict seasonal diseases.
- [3] World Organization for Animal Health, (2022), Early-warning systems: modeling the spread of denguediseases. <https://www.woah.org/en/article/early-warning-systems-modeling-the-spread-of-vector-borne-diseases/>
- [4] Sadeghieh T, Waddell LA, Ng V, Hall A, Sargeant J. 2020. A scoping review of importation and predictive models related to denguediseases, pathogens, reservoirs, or vectors (1999–2016). *PLoS ONE* 15, e0227678.
- [5] Kearney M, Porter W. 2009. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecol. Lett.* 12, 334-350.
- [6] Paz S. 2015. Climate change impacts on West Nile virus transmission in a global context. *Phil. Trans. R. Soc. B* 370, 1-11.
- [7] Estrada-Peña A, Ayllón N, Fuente J de la. 2012. Impact of climate trends on tick- borne pathogen transmission. *Front. Physiol.* 3, 1-12.
- [8] Tjaden NB, Caminade C, Beierkuhnlein C, Thomas SM. 2018. Mosquito-borne diseases: advances in modelling climate-change impacts.
- [9] Rocklöv J, Dubrow R. 2020. Climate change: an enduring challenge for denguedisease prevention and control. *Nat. Immunol.* 21, 479-483.



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)