ZKG INTERNATIONAL ISSN: 2366-1313

TIME SERIES ANALYSIS BASED PREDICTION OF DENGUE SPREAD USING CLIMATE DATA

[1]DR.B.SRINIVASA RAO, [2]B.SHIVA, [3]G.RISHITHA, [4]G.SRIKANTH

[1](Assistant Professor) ,**CSE.** Teegala Krishna Reddy Engineering College Hyderabad

[234]B,tech scholar ,**CSE.** Teegala Krishna Reddy Engineering College Hyderabad

ABSTRACT

Dengue is endemic in all surrounding international locations with the 4 serotypes circulating with in side the location inside a duration of ten years. Countries or territories with the very best quantity of mentioned dengue instances have been Puerto Rico, the Dominican Republic, Martinique, Trinidad and Tobago and French Guiana. The studies employs device gaining knowledge of algorithms to research the temporal relationships among weather variables (inclusive of temperature, precipitation, and humidity) and Dengue instances. Through the improvement and validation of predictive models, the observe seeks to beautify our knowledge of the dynamic interaction among climatic elements and sickness transmission. This proposed gadget is constructed to expect the unfold of dengue fever with weather information the use of the idea of time collection evaluation. In addition, this assignment additionally plays the exploratory information analytics at the dengue dataset over a duration of time. Finally, prediction evaluation additionally completed with using development rendered with the aid of using device gaining knowledge of algorithms.

1 INTRODUCTION

1.1 Overview

Dengue is a probably life-threatening arboviral ailment transmitted with the aid of using woman Aedes mosquitoes, in particular A. aegypti, A. albopictus, and A. vitattus. These vectors are not unusualplace tropical hematophagous ectoparasites. This zoonotic ailment unfold from African or Asian non-human primates 500 to one thousand years ago, however in the final 60 years it has unfold from simply nine

ISSN: 2366-1313

international locations experiencing excessive epidemics to come to be endemic in over a hundred international locations worldwide, even affecting non-tropical or subtropical areas. Moreover, about a hundred million humans every year be afflicted by the symptomatic ailment because of its 4 serotypes. Given the big effect of environmental modifications on ailment transmission, the One Health method is urgently had to put in force the mixing among human, animal, and ecological fitness.

The goal of this paper is to offer an perception into strategies that may be used for destiny predictive fashions primarily based totally at the One Health angle, mainly in admire to Latin America however additionally elsewhere. One Health is a multidisciplinary method that recognizes the synergy among human and animal fitness and their shared environment. This method has come to be an increasing number of vital withinside the twenty first Century with the convergence of the pressures of converting climate, migration of human and animal populations, and the developing human populace that will increase the proximity among flora and fauna and humans. Indeed, the time period One Health changed into simplest coined withinside the early 2000s with the arrival of the zoonotic SARS and H5N1 influenza diseases.

Whilst the One Health angle is broadly visible as essential and an increasing number of used for higher ailment control, epidemiological processes have now no longer saved up with this change. Conventional epidemiological views generally tend to view ailment widely from a humansimplest angle, that specialize in human demographic situations with frequently simplest climatic/environmental elements accommodating the ailment vector fitness. 2 For example, even as environmental and sociological issues frequently take a again seat in One Health, they regularly occupy the centre degree in epidemiology. Factors which includes suggest temperatures and rainfall utilized in predicting dengue, with a completely indistinct attention of ways they have an effect on the mosquito vectors, are an emergent task to be taken into consideration. High rainfall, for instance, is useful to mosquitoes as it offers water-stuffed places for eggs and larvae, even as the mosquitoes are often impervious to moves with the aid of using raindrops that would in any other case kill them. In addition, temperature and rain usually have an effect on many different infectious and tropical diseases.

1.1 Description

Dengue is a human arbovirus sickness transmitted through the girl mosquito of the genus. The studies employs device studying algorithms to investigate the temporal relationships among weather variables (consisting of temperature, precipitation, and humidity) and Dengue cases. Through the improvement and validation of predictive models, the have a look at seeks to decorate our knowledge of the dynamic interaction among climatic elements and sickness transmission. Dengue fever stays a giant public fitness difficulty in lots of regions, with its unfold stimulated through diverse environmental elements, in particular weather variables. This have a look at employs time collection evaluation strategies to forecast the unfold of dengue fever primarily based totally on historic weather facts. By integrating superior statistical strategies with meteorological parameters consisting of temperature, humidity, and rainfall, this studies goals to offer correct predictions of dengue occurrence patterns. The findings of this have a look at preserve capability for informing public fitness techniques and interventions, helping withinside the proactive control and manipulate of dengue outbreaks.

The implications of correct dengue unfold predictions primarily based totally on weather facts are giant for public fitness government and policymakers. Furthermore, the insights received from this have a look at may also make a contribution to the improvement of early caution structures for dengue fever, improving preparedness and reaction skills at local and countrywide levels.

1.2 Objective

This research demonstrates the potential of time series analysis techniques for predicting dengue spread using climate data. By leveraging historical patterns and correlations between meteorological variables and dengue incidence,

accurate forecasts can be generated to support proactive public health interventions. Continued refinement and validation of predictive models will be essential for advancing our understanding of dengue transmission dynamics and improving outbreak management strategies in endemic regions. The results of the time series analysis will include forecasts of dengue incidence based on climate data, along with measures of model performance and accuracy. Visualization techniques such as time series plots, seasonal decomposition

ISSN: 2366-1313

charts, and correlation matrices will be employed to illustrate the relationships between climate variables and dengue transmission dynamics. Validation of the predictive models will be conducted using out-of-sample testing and crossvalidation techniques.

2 .LITERATURE SURVEY

2.1 Review of literature

Majeed, M.A.; Shafri, H.Z.M.; [1] dengue fever cases in Malaysia using machine learning techniques. A dataset consisting of weekly dengue cases at the state level in Malaysia from 2010 to 2016 was obtained from the Malaysia Open Data website and includes variables such as climate, geography, and demographics. Six different long short-term memory (LSTM) models were developed and compared for dengue prediction in Malaysia: LSTM, stacked LSTM (S-LSTM), LSTM with temporal attention (TA-LSTM), S-LSTM with temporal attention (STA-LSTM), LSTM with spatial attention (SA-LSTM), and S-LSTM with spatial attention (SSA-LSTM). Cabrera, M.; Leake, J.;

[2] epidemiological prediction of dengue fever using the One Health perspective, including an analysis of how Machine Learning techniques have been applied to it and focuses on the risk factors for dengue in Latin America to put the broader environmental considerations into a detailed understanding of the small-scale processes as they affect disease incidence. Determining that many factors can act as predictors for dengue outbreaks, a large-scale comparison of different predictors over larger geographic areas than those currently studied is lacking to determine which predictors are the most effective. Dey, Samrat Kumar, et al.

[3] develop a machine learning model that can use relevant information about the factors that cause Dengue outbreaks within a geographic region. To predict dengue cases in 11 different districts of Bangladesh, we created a DengueBD dataset and employed two machine learning algorithms, Multiple Linear Regression (MLR) and Support Vector Regression (SVR).

This research also explores the correlation among environmental factors like temperature, rainfall, and humidity with the rise and decline trend of Dengue cases in different cities of Bangladesh. The entire dataset was divided into an 80:20 ratio, with 80 percent used for training and 20% used for testing. 5 The research findings imply

ISSN: 2366-1313

that, for both the MLR with 67% accuracy along with Mean Absolute Error (MAE) of 4.57 and SVR models with 75% accuracy along with Mean Absolute Error (MAE) of 4.95, the number of dengue cases reduces throughout the winter season in the country and increases mainly during the rainy season in the next ten months, from August 2021 to May 2022. Kakarla, S.G., Kondeti, P.K., et al. [4] applied vector auto regression, generalized boosted models, support vector regression, and long short-term memory (LSTM) to predict the dengue prevalence in Kerala state of the Indian subcontinent. Consider the number of dengue cases as the target variable and weather variables viz., relative humidity, soil moisture, mean temperature, precipitation, and NINO3.4 as independent variables. Various analytical models have been applied on both datasets and predicted the dengue cases. Among all the models, the LSTM model was outperformed with superior prediction capability (RMSE: 0.345 and R2:0.86) than the other models. Roster, Kirstin, et al. [5] developed a model for predicting monthly dengue cases in Brazilian cities 1 month ahead, using data from 2007–2019. We compared different machine learning algorithms and feature selection methods using epidemiologic and meteorological variables. They found that different models worked best in different cities, and a random forests model trained on monthly dengue cases performed best overall. It produced lower errors than a seasonal naive baseline model, gradient boosting regression, a feed-forward neural network, or support vector regression.

3. SYSTEM DESIGN

3.1 SYSTEM AECHITECTURE

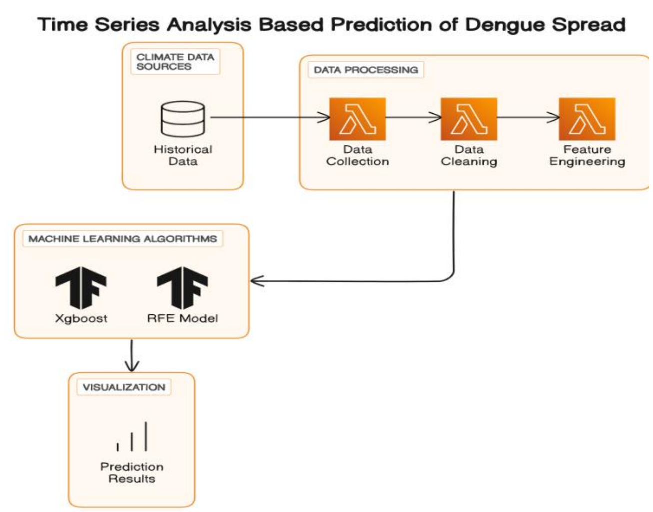

Fig:3.1 System architecture

3.1.1 Data Flow Diagram

The DFD is a easy graphical formalism that may be used to symbolize a device in phrases of enter facts to the device, numerous processing performed in this facts, and the output facts is generated through this device. It is used to version the device

ISSN: 2366-1313

additives. These additives are the device procedure, the facts utilized by the procedure, an outside entity that interacts with the device and the statistics flows withinside the device. 11 The visualizations are important for gaining insights into the dataset's structure, relationships among variables, and the distribution of key attributes, that may tell similarly facts evaluation and modeling decisions.This visualization creates a be counted number plot to reveal the distribution of a specific variable called 'alpha' withinside the dataset. It facilitates you apprehend the frequency or be counted number of every class inside the 'alpha' variable, presenting insights into the magnificence distribution.

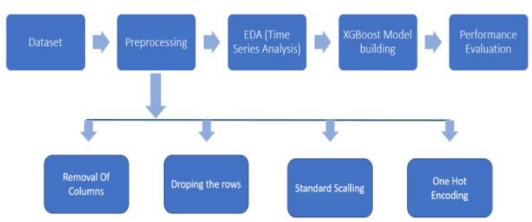

Fig:3.1.1 Data Flow Diagram

Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way. So, for this, we use data pre-processing task.

A real-world data generally contains noises, missing values, and maybe in an unusable format which cannot be directly used for machine learning models. Data pre-processing is required tasks for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model. 12 In machine learning data pre-processing, we divide our dataset into a training set and test set. This is one of the crucial steps of data pre-processing as by doing this, we can enhance the performance of our machine learning model. Suppose if we have given training to our machine learning model by a dataset and we test it by a completely different dataset. Then, it will create difficulties for our model to understand the correlations between the models.

ARIMA stands for Auto Regressive Integrated Moving Average, and it captures the patterns, trends, and seasonality of the data using a combination of past values, differences, and errors. XGBoost is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "XGBoost is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the XGBoost takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of over fitting.

4. OUTPUT SCREENS

	city	year	weekofyear	week_start_date	ndvi_ne	ndvi_nw	ndvi_se	ndvi_sw	precipitation_amt_mm	reanalysis_air_temp_k	reanalysis_avg_temp_k
0	sj	1990	18	1990-04-30	0.122600	0.103725	0.198483	0.177617	12.42	297.572857	297.742857
1	sj	1990	19	1990-05-07	0.169900	0.142175	0.162357	0.155486	22.82	298.211429	298.442857
2	sj	1990	20	1990-05-14	0.032250	0.172967	0.157200	0.170843	34.54	298.781429	298.878571
3	sj	1990	21	1990-05-21	0.128633	0.245067	0.227557	0.235886	15.36	298.987143	299.228571
4	sj	1990	22	1990-05-28	0.196200	0.262200	0.251200	0.247340	7.52	299.518571	299.664286
...	...	...	...	...	...	...	...	...	...	...	...
1451	iq	2010	21	2010-05-28	0.342750	0.318900	0.256343	0.292514	55.30	299.334286	300.771429
1452	iq	2010	22	2010-06-04	0.160157	0.160371	0.136043	0.225657	86.47	298.330000	299.392857
1453	iq	2010	23	2010-06-11	0.247057	0.146057	0.250357	0.233714	58.94	296.598571	297.592857
1454	iq	2010	24	2010-06-18	0.333914	0.245771	0.278886	0.325486	59.67	296.345714	297.521429
1455	iq	2010	25	2010-06-25	0.298186	0.232971	0.274214	0.315757	63.22	298.097143	299.835714

1456 rows × 25 columns

reanalysis_tdtr_k	station_avg_temp_c	station_diur_temp_rng_c	station_max_temp_c	station_min_temp_c	station_precip_mm	total_cases
2.628571	25.442857	6.900000	29.4	20.0	16.0	4
2.371429	26.714286	6.371429	31.7	22.2	8.6	5
2.300000	26.714286	6.485714	32.2	22.8	41.4	4
2.428571	27.471429	6.771429	33.3	23.3	4.0	3
3.014286	28.942857	9.371429	35.0	23.9	5.8	6
...	...	...	...	...	...	...
9.800000	28.633333	11.933333	35.4	22.4	27.0	5
7.471429	27.433333	10.500000	34.7	21.7	36.6	8
7.500000	24.400000	6.900000	32.2	19.2	7.4	1
7.871429	25.433333	8.733333	31.2	21.0	16.0	1
11.014286	27.475000	9.900000	33.7	22.2	20.4	4

Figure 4.1: Sample dataset used for Dengue spread prediction

• Figure 4.1 represents a portion of the dataset that was used for predicting the spread of Dengue fever. It includes various features (columns) and corresponding target values (total cases) for a specific period.

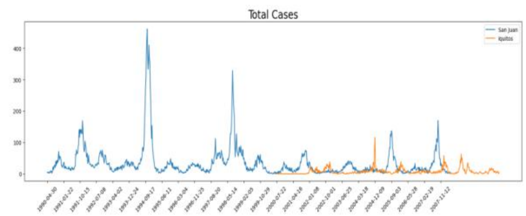

Figure 4.2: line plot to visualize the total number of Dengue fever cases over time for two different cities

• Figure 4.2 is a line plot that displays the total number of Dengue fever cases over time for two different cities: San Juan and Iquitos. The x-axis represents time (likely in weeks or months), while the y-axis represents the total number of cases. There are two lines, one for each city, showing how the cases change over time

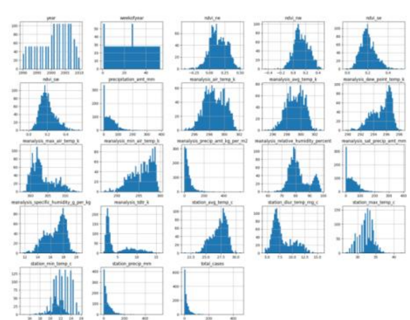

Figure 4.3: histogram for each numerical column in the Data Frame

• Figure 4.3 consists of multiple histograms, each representing the distribution of values for a numerical column in the DataFrame. It provides insights into the frequency of different values within each column.

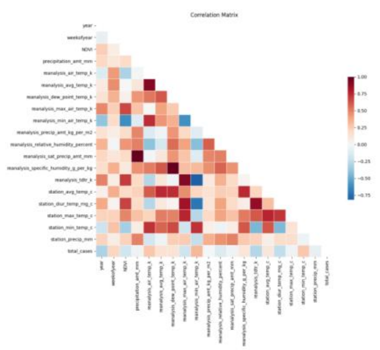

Figure 4.4: Heatmap of correlation of columns in a dataset used for dengue spread prediction.

• Figure 4.4 is a heatmap that visualizes the correlation between different columns (features) in the dataset used for predicting Dengue spread. Each cell in the heatmap represents the correlation coefficient between two columns. A warmer color (closer to red) indicates a stronger positive correlation, while a cooler color (closer to blue) indicates a stronger negative correlation.

	city	year	weekofyear	NDVI	week_start_date	precipitation_amt_mm	reanalysis_air_temp_k	reanalysis_avg_temp_k	reanalysis_max_air_temp_k
0	1	1990.0	18.0	0.150606	1990-04-30	12.42	297.572857	297.742857	299.8
1	1	1990.0	19.0	0.157479	1990-05-07	22.82	298.211429	298.442857	300.9
2	1	1990.0	20.0	0.133315	1990-05-14	34.54	298.781429	298.878571	300.5
3	1	1990.0	21.0	0.209286	1990-05-21	15.36	298.987143	299.228571	301.4
4	1	1990.0	22.0	0.239235	1990-05-28	7.52	299.518571	299.664286	301.9
...	...	...	...	...	...	...	...	...	...
1451	0	2010.0	21.0	0.302627	2010-05-28	55.30	299.334286	300.771429	309.7
1452	0	2010.0	22.0	0.170557	2010-06-04	86.47	298.330000	299.392857	308.5
1453	0	2010.0	23.0	0.219296	2010-06-11	58.94	296.598571	297.592857	305.5
1454	0	2010.0	24.0	0.296014	2010-06-18	59.67	296.345714	297.521429	306.1
1455	0	2010.0	25.0	0.280282	2010-06-25	63.22	298.097143	299.835714	307.8

1456 rows × 21 columns

Figure 4.5: data frame after preprocessing used for dengue spread

• Figure 4.5 displays a portion of the dataset after it has undergone preprocessing steps. Preprocessing may include tasks like handling missing values, feature engineering, and encoding categorical variables. It represents the cleaned and transformed data ready for modelling.

	city	weekofyear	NDVI	precipitation_amt_mm	reanalysis_air_temp_k	reanalysis_avg_temp_k	reanalysis_max_air_temp_k	reanalysis_min_air_temp_k
936	0	26.0	0.228307	25.41	296.740000	298.450000	307.3	293.1
937	0	27.0	0.256012	60.61	296.634286	298.428571	306.6	291.1
938	0	28.0	0.170504	55.52	296.415714	297.392857	304.5	292.6
939	0	29.0	0.206918	5.60	295.357143	296.228571	303.6	288.6
940	0	30.0	0.316546	62.76	296.432857	297.635714	307.0	291.5
...	...	...	...	...	...	...	...	...
706	1	48.0	0.084369	16.20	298.814286	298.907143	301.0	297.2
707	1	49.0	0.073851	0.00	299.107143	299.242857	300.7	297.4
708	1	50.0	-0.007986	182.81	299.174286	299.185714	301.5	297.5
709	1	51.0	0.099544	0.00	298.555714	298.607143	300.5	296.9
710	1	52.0	0.071783	1.96	299.044286	299.178571	300.8	297.2

1049 rows × 18 columns

Figure 4.6: data frame of features column of a dataset after preprocessing

• Figure 4.6 specifically focuses on the features column(s) of the dataset after preprocessing. It may display the values, statistics, or distribution of the features that will be used for predicting Dengue spread.

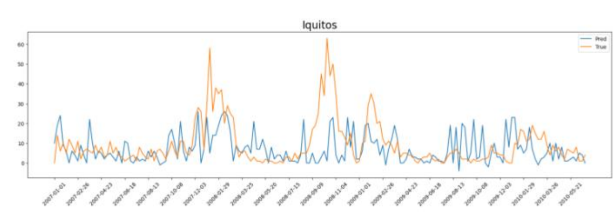

Figure 4.7: line plot to compare the predicted and actual cases of Dengue fever for the city of Iquitos (iq).

ISSN: 2366-1313

• Figure 4.7 is a line plot that compares the predicted cases (likely generated by a machine learning model) with the actual cases of Dengue fever for the city of Iquitos. The x-axis represents time, while the y-axis represents the number of cases. The plot helps assess how well the model's predictions align with the actual data.

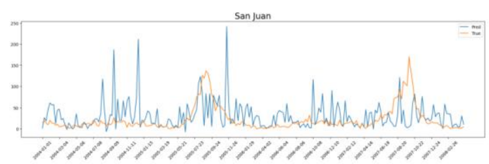

Figure 4.8: line plot to compare the predicted and actual cases of Dengue fever for the city of San Juan (S j).

• Figure 4.8 Similar to Figure 8.7, this figure compares the predicted cases with the actual cases of Dengue fever, but for the city of San Juan.

	city	year	weekofyear	total_cases
0	sj	2008	18	3
1	sj	2008	19	34
2	sj	2008	20	23
3	sj	2008	21	36
4	sj	2008	22	16
5	sj	2008	23	21
6	sj	2008	24	4
7	sj	2008	25	38
8	sj	2008	26	3
9	sj	2008	27	4

Figure 4.9: Prediction results using XG Boost classifier

• Figure 4.9 displays the overall prediction results obtained using an XGBoost classifier. It may include metrics such as Mean Absolute Error (MAE) or other evaluation measures to assess the performance of the model. These figures collectively provide a comprehensive view of the data, its preprocessing, and the results of the Dengue spread prediction model

5. CONCLUSION

This research demonstrates the potential of time series analysis techniques for predicting dengue spread using climate data. By leveraging historical patterns and correlations between meteorological variables and dengue incidence, accurate forecasts can be generated to support proactive public health interventions. Continued refinement and validation of predictive models will be essential for advancing our understanding of dengue transmission dynamics and improving outbreak management strategies in endemic regions. The results of the time series analysis will include forecasts of dengue incidence based on climate data, along with measures of model performance and accuracy. Visualization techniques such as time series plots, seasonal decomposition charts, and correlation matrices will be employed to illustrate the relationships between climate variables and dengue transmission dynamics. Validation of the predictive models will be conducted using out-of-sample testing and cross-validation techniques.

6. FUTURE ENHANCEMENT

The future scope of this research holds several exciting possibilities. Firstly, the incorporation of more advanced machine learning algorithms and deep learning models could enhance the accuracy of Bitcoin price predictions. Additionally, incorporating sentiment analysis of news, social media, and market sentiment could provide valuable insights into market dynamics and help improve forecasting accuracy. Furthermore, expanding the analysis to consider other cryptocurrencies and their interrelationships with Bitcoin could provide a more comprehensive view of the cryptocurrency market. Another avenue for future exploration is the development of real-time prediction models that adapt to changing market conditions, as cryptocurrency markets are highly influenced by breaking news and events. Moreover, the integration of blockchain analytics and on-chain data into the prediction process could offer unique insights into Bitcoin's price movements. Lastly, this research can extend its focus to explore the broader economic and regulatory implications of Bitcoin price predictions. As Bitcoin continues to gain prominence in global financial markets, accurate forecasting becomes not only a financial asset but also a tool for policymakers and investors. In summary, the future scope of this study lies in the refinement of prediction models, the incorporation of additional data sources, and the exploration of the broader implications of Bitcoin price forecasts in the context of the global economy and financial markets.

7. REFERENCES

[1] Majeed, M.A.; Shafri, H.Z.M.; Zulkafli, Z.; Wayayok, A. A Deep Learning Approach for Dengue Fever Prediction in Malaysia Using LSTM with Spatial Attention. Int. J. Environ. Res. Public Health 2023, 20, 4130.

[2] Cabrera, M.; Leake, J.; Naranjo-Torres, J.; Valero, N.; Cabrera, J.C.; RodríguezMorales, A.J. Dengue Prediction in Latin America Using Machine Learning and the One Health Perspective: A Literature Review. Trop. Med. Infect. Dis. 2022, 7, 322.

ISSN: 2366-1313

[3] Dey, Samrat Kumar, et al. "Prediction of dengue incidents using hospitalized patients, metrological and socio-economic data in Bangladesh: A machine learning approach." PLoS One 17.7 (2022): e0270933.

[4] Kakarla, S.G., Kondeti, P.K., Vavilala, H.P. et al. Weather integrated multiple machine learning models for prediction of dengue prevalence in India. Int J Biometeorol 67, 285– 297 (2023).

[5] Roster, Kirstin, Colm Connaughton, and Francisco A. Rodrigues. "Machine-Learning– Based Forecasting of Dengue Fever in Brazilian Cities Using Epidemiologic and Meteorological Variables." American Journal of Epidemiology 191.10 (2022): 1803-1812.

[6] Sarder, Faysal, Sorefa Akter, and Sharmin Akter. "Predicting Dengue Outbreak from Climate Data Using Machine Learning Algorithms." 2022 IEEE International Conference on Data Science and Information System (ICDSIS). IEEE, 2022. 45

[7] Ochida, N., Mangeas, M., Dupont-Rouzeyrol, M. et al. Modeling present and future climate risk of dengue outbreak, a case study in New Caledonia. Environ Health 21, 20 (2022).

[8] Anuranjan, M. B., et al. "Machine Learning Techniques for Predicting Dengue Outbreak." Innovations in Information and Communication Technologies: Proceedings of ICIICT 2022. Singapore: Springer Nature Singapore, 2022. 45-56.

[9] Gupta, G.; Khan, S.; Guleria, V.; Almjally, A.; Alabduallah, B.I.; Siddiqui, T.; Albahlal, B.M.; Alajlan, S.A.; AL-subaie, M. DDPM: A Dengue Disease Prediction and Diagnosis Model Using Sentiment Analysis and Machine Learning Algorithms. Diagnostics 2023, 13, 1093.

[10] Rocha, F.P., Giesbrecht, M. Machine learning algorithms for dengue risk assessment: a case study for São Luís do Maranhão. Comp. Appl. Math. 41, 393 (2022).