DATA MINING BASED APPROACH TO FORECAST DENGUE DISEASE IN GEOGRAPHICAL AREAS BY APPLYING MACHINE LEARNING ALGORITHMS

Rama Krishna K[1]*, Dr. Mohan K G[2], Dr. Mahalakshmi R[3]

Abstract

Dengue is a viral disease caused by the Aedes species mosquitoes to people. It mostly occurs in subtropical and tropical climates. It might have mild and worsened symptoms such as headache, nausea, high fever, body aches, and rashes. Some patients recovered from this disease without any medical and some undergo severe treatment and some may loss their lives. It gets worsen with climatic change and early forecasting is an important factor for the forecasting of Dengue. This work focuses to forecast dengue from four cities in India: Chennai, Hyderabad, Bangalore, and Mumbai. Moreover, we propose machine learning approaches such as Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), and Logistic regression (LR) methods and compared them with state-of-art works on weather-based Dengue disease forecasting. Monthwise dengue patients and related rates in different cities in India are provided by the National Vector Borne Disease Control Program (NVBDCP). Here, we have taken the data' from four cities namely Chennai, Hyderabad, Bangalore, and Mumbai in which the dataset used in forecasting includes the fields such As City, Month, Year, Normalized Difference Vegetation Index (Ndvi) (NDVI NE Northeast, NDVI NW Northwest, NDVI SE Southeast, NDVI SW Southwest), Humidity (%), Air Temperature (Minimum, Maximum and Average), Precipitation (Rainfall in mm), Dew Factor (Dew Point Temp). For analysis, we have collected data from the four cities and conducted it over the JAVA simulator. The proposed work is compared with state-of-art works in terms of Root Mean Square (RMSE), Mean Absolute Error (MAE), and Mean Squared Error (MSE). The proposed work surpasses all the other approaches and helps in accepting the forecasting outcomes.

Keywords: Dengue Forecasting, early forecasting, climatic change, and machine learning approaches.

[1]*Assistant Professor, Computer Science and Engineering, Presidency University, Bangalore,
E-mail:- Ramakrishna@presidencyuniversity.in
[2]Professor, Computer Science and Engineering, Gitam University, Bangalore, E-mail:- mkabadi@gitam.edu
[3]Professor, Computer Science and Engineering, Presidency University, Bangalore,
E-mail:- mahalakshmi@presidencyuniversity.in

***Corresponding Author: -** Rama Krishna K
*Assistant Professor, Computer Science and Engineering, Presidency University, Bangalore,
E-mail:- Ramakrishna@presidencyuniversity.in

1. Introduction:

The virus responsible for dengue fever [1] is the root of dengue illness, which is spread by mosquitoes. Many to thousands of millions of humans contract the illness every year, and it is a worldwide issue that impacts many nations that are tropical. Aedes can be fatal; each year, it is blamed for thousands of fatalities. Tropical and equatorial [2] varieties of flies which transmit dengue typically inhabit regions between 35N and 35S and elevations below 1000 meters. Since the disease-carrying mosquitoes typically thrive in environments with water and are located close to human habitations, people are also able to transmit dengue from one neighborhood to another. It has mainly examined the relationships between various variables and dengue transmission characteristics.

Analysis of variance [3], examinations of correlation, and time conjunction extraction are the data analysis approaches that are primarily taken into account during such inquiries. Data for the studies are frequently gathered in Mexico, a country regarding states that vary greatly in terms of terrain, weather, finances, and ethnicities. The rising popularity of networking [4] apps makes it possible to analyze individual sentiments, beliefs, and ideas on a variety of subjects, including governance, athletics, entertainment, schooling, technology, and the arts. Internet viewers routinely publish content or change their statuses to reflect their current situation, For instance, to assess the impact of environmental variables [5].

333/For the forecasting of dengue disease and forecasting several approaches have been stated by several scientists. To ensure the novelty we propose a novel machine learning approaches to predict the forecasting of dengue disease along with the dengue disease data of four cities and climatic factors. Moreover, we utilized LR to forecast the outbreak of the disease in four cities. Major contributions are summarized below;

- The data are collected from four cities namely Chennai, Bangalore, Hyderabad, and Mumbai. Then the collected data are converted into the database by the data integration and cleaning approach.
- The exploratory analyzes are made and along with the climatic factors the input is forwarded to the machine learning approaches to forecast the forecasting of the disease. Performance evaluations are made.
- Utilizing the LR the disease outbreak is forecasted.

The rest of the work is organized as follows, in section 2 the relevant works are analyzed and the proposed work is elaborated in section 3. The simulation conducted is explained in section 4. Finally, the work is concluded in section 5.

2. Literature survey:

Appice et al. [10] have described a machine learning strategy to examine the relationship between climate and viral information source fluctuations in time in order to accurately forecast dengue based on temperature annually. Using clustering calculations, it can simulate the distribution and time stability of an upward established by the relationships within past observations on weather and an infectious component. Then, the frequency series proximal neighbor generator is strengthened using the aforementioned cluster information. It is used to assess the efficacy of the suggested approach. Moreover, it wants to investigate whether supplementary time series methods of clustering work.

Amin et al. [11] have presented deep learning approaches to make use of communication knowledge for quick and accurate disease diagnosis. The performer is denied access to any individual's private data or medical histories. It determines that a disease called dengue is present exclusively on updates and determines the fact that there is a widespread conversation regarding the illness but no confirmed infection or the extent to which there is a genuine outbreak. When contrasted with the most advanced methods currently available in this sector, the suggested model has the ability to attain a precision of 92%. Hence, it is difficult to access the large dataset.

Latif et al. [12] have evaluated Machine Learning and Deep Learning Methods for the automated selection of relevant data, and for the precise identification of health conditions. The data that is gathered through attributes are used by the healthcare provider as the parameter to facilitate the disease's autonomous evaluation. This gathering was originally carried out by information databases. Incorporating test findings and patient information in addition to this depiction will yield the most effective outcomes. Thus, it is demanded that it will be used for widespread applications, such as healthcare structures.

Gangula et al. [13] have modified the Ensemble Machine Learning technique in mixed approaches to pinpoint traits linked to growth. Dengue fever that has is a serious disease that is spreading across the globe and places an enormous strain on nations with endemic populations; an outstanding durability method must be developed. The reinforcement of prompt examination is of the utmost importance for the purpose to improve managing cases, enabling effective utilization of scarce funds, and recognize individuals. Thus, with the huge amount of data, the process is costly.

Chamola et al. [14] suggested machine learning algorithms in the control of pandemics and disasters. It can be handled by predictive models, and they can also spot anomalies in the information. Least-squares assessment is a crucial method when confronted with exceptional threats. While trying to forecast extremely improbable occurrences including crises as well as epidemics each of us ought to devote particular focus to these types of anomalies as opposed to eliminating themselves. It is efficient and accurate to predict the forecast. Moreover, the immune system is not considered in this model.

Umar et al. [15] have implemented the artificial neural networks- a genetic algorithm- Sequential quadratic method (ANN-GA-SQM) employing ten digits of synapses and the sequential layout of neural systems. It uses random mutation to resolve restricted and unstructured processes by an optimization-based global search method. It uses a toolkit that includes intersection, decision-making, transformation, and generation to produce the most favorable outcomes possible from the framework. It is examined by consistently attaining crossover outcomes and a respectable degree of fidelity. However, it is difficult to solve numerically based methods.

Phakhounthong et al. [16] highlighted a classification and regression tree (CART) was used to build a medical ecosystem for a prevalent sickness, which was then used to determine the significance of all components individually. It is necessary to set connected specific time collections, which usually involves selecting, removing, merging, and changing the record set. This strategy might be useful for guiding access to a plan for health records and facility environment management. Therefore, it is difficult but crucial to quickly distinguish mature individuals from persons at elevated risk.

Chatterjee et al. [17] have implemented a modified bag of features method used a lot in tasks where the classification of images is required. The vast amount of distinct classifiers from a picture combined with a massive amount of key point sites makes categorizing images challenging. As a result, because these factors are so important in the process stage, it seems imperative to reduce the variety of elements and integrate more tools. The procedure for making choices can be accelerated and improved in a quick, precise, and productive manner. The computational expense is higher as a result.

3. Proposed Machine Learning based Forecasting model of dengue disease

The proposed machine learning approaches for the forecasting model of dengue disease are elaborated in figure 1. At first, the data are collected from the four cities such as Chennai, Bangalore, Hyderabad, and Mumbai which includes weather as well as the dengue disease. Then the database is generated from the collected data using the data cleaning and integration approach as shown in the figure. Then the exploratory analysis is effectuated and the climatic parameters are fed along with the exploratory analysis data into the machine learning approaches like SVM, DT, RF, and LR. Followed by performance evaluations are made using the parameters such as RMSE, MAE, and MSE. Finally, the outbreak of dengue disease is forecasted using the LR approach.

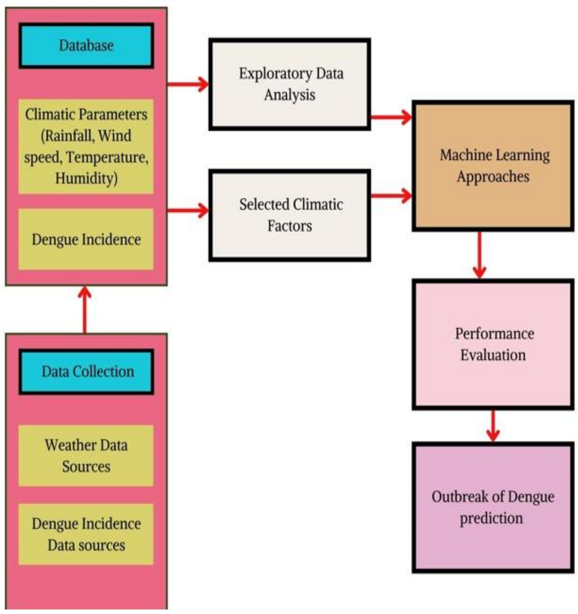

Fig 1: Proposed dengue disease forecasting approach

3.1 Data collection site

We have conducted this study in India which has the world's most biodiverse ecozones such as highlands, plains, rivers, islands, tropical regions, mountains, and grasslands. According to 2023's global climate index India ranked 8^{th} position.

In India, dengue surveillance is a passive one and dengue cases are diagnosed by public health care professionals from various levels. The surveillance systems are conducted by the state and union government known as National Vector Borne Disease Control Program (NVBDCP) [18]. Periodic reviews of are made and recorded on a daily basis of dengue incidence.

3.2 Database

Dengue is one of the climatic-based diseases in India. Monthwise dengue patients and related rates in different cities in India are provided by the NVBDCP. Here we have taken the data' from four cities namely Chennai, Hyderabad, Bangalore, and Mumbai. Climatic conditions are responsible for the fast spread of the disease. Different meteorogical factors are considered from the weather data which consists of Month, Year, Normalized Difference Vegetation Index (Ndvi) (NDVI NE Northeast, NDVI NW Northwest, NDVI SE Southeast, and NDVI SW Southwest), Humidity (%), Air Temperature (Minimum, Maximum and Average), Precipitation (Rainfall in mm), Dew Factor (Dew Point Temp). Here we have taken the data' from four cities namely Chennai, Hyderabad, Bangalore, and Mumbai from 16^{th} October 2018 to 1^{st} October 2022. The details are elaborated in table 1.

Table 1: The meteorological parameters to be evaluated for the selected city

<table>
<tr><th>Meteorological Parameters</th><th>Description</th><th>Unit</th></tr>
<tr><td>Minimum Absolute monthly temperature</td><td rowspan="5">These values were evaluated with the three specialized thermometers on an average of four times per day. So for each day itself, we could measure minimum, maximum and average values and use them for the monthwise calculation.</td><td>Celcius</td></tr>
<tr><td>Maximum Absolute monthly temperature</td><td>Celcius</td></tr>
<tr><td>Average temperature months (minimum)</td><td>Celcius</td></tr>
<tr><td>Average temperature monthwise (maximum)</td><td>Celcius</td></tr>
<tr><td>Average temperature monthwise</td><td>Celcius</td></tr>
<tr><td>Relative average humidity monthwise</td><td>Based on the WMO standards the humidity values are measured and measured four times a day. Based on the values the average values are evaluated monthwise</td><td>%</td></tr>
<tr><td>Relative average humidity monthwise (minimum)</td><td>The minimum average humidity value monthwise could be evaluated from the measured values</td><td>%</td></tr>
<tr><td>Rainfall Monthwise</td><td>For the measurement of rainfall the WMO's permitted meter is placed in various places and measured four times per day. The summation of four values provides the rainfall of each day and calculated the monthwise rainfall value</td><td>mm</td></tr>
<tr><td>Rainy days monthwise</td><td>It is measured on the basis of days with more than 0mm of rain and calculated on the rainy days of the month</td><td>Days</td></tr>
<tr><td>Maximum rainfall in a day monthwise</td><td>Choose from the measured values</td><td>mm</td></tr>
<tr><td colspan="3">WMO-World-Meteorological Organization</td></tr>
</table>

The collected data related to climate and dengue are integrated to generate the database of the work as shown in figure 1. The collected data are inconsistent with the diverse health and weather nature. From the data collected from the Chennai, Bangalore, Hyderabad, and Mumbai the irrelevant data are ignored while performing the integration. While some values are missing when considering the climate factors. To ignore the missing values the data cleaning is effectuated and we have adopted the data imputation [19] approach. the missing values are replaced with previous values collected in different years. The month imputation mean value can be evaluated as,

$$V = \left(\sum D_j = S_{ij}\right) / D \quad (1)$$

$$V = \left(S_{ij1} + S_{ij2} + S_{ij3} + ... + S_{ijD}\right) / D \quad (2)$$

The missing attribute values (V) could be evaluated with the sum average value S_{ij} for the i^{th} month of the j^{th} year. The available data (both dengue and climatic) for the j^{th} year is denoted as D.

3.3 Exploratory Data Analysis

This is effectuated to analyze and find the mandatory patterns and graphical analysis. The correlation among the dengue and climate factors of cities Chennai, Bangalore, Hyderabad, and Mumbai are effectuated and filters the significant details. The correlation is determined with the assist of Pearson correlation and generated the heat maps for four cities. Pearson correlation [20] is nothing used to analyze the correlation among two parameters and is suitable for the methods which utilized the covariance and numeric values. it can be defined as,

$$P = \frac{\sum_{i=1}^{n}(mi - m)(ni - n)}{\sqrt{\left[\sum_{i=1}^{n}(mi - \hat{m})^2\right]\left[\sum_{i=1}^{n}(ni - \hat{n})^2\right]}} \quad (3)$$

The correlation results determine that the climatic parameters impact the dengue incidence differently and the dengue increases with the decreasing temperature.

3.4 Machine Learning approaches for dengue disease forecasting

This study presents the ML algorithm, which is a combination of a Support Vector Machine, Random Forest, Decision Tree, and Logistic regression methods to forecast dengue disease.

3.4.1 Support Vector Machine (SVM):

The linear function form of hypothetical space is Support Vector Machine (SVM). In the input space, the forecasting of the dengue disease and the best hyper-plane is determined. The maximum point determination and margins measurements are performed with Hyperplane [21]. SVM is the closest pattern with the distance among the hyperplane is margin [22]. The number of patient data is $Y_j \in \Re^d$ where $j = 1,2,...,M$. The below equation notifies the class label with each data and the number of patient data *M*.

$$Y_j \in \{-1,+1\} \quad (4)$$

In *d*-dimensional search space, the hyper-plane separates and it is defined as,

$$W \cdot Y_j + B = 0 \quad (5)$$

The plane position and normal plane are *B* and *W*. The negative class and positive class classify the data Y_j as explained in equations (6) and (7).

$$W \cdot Y_j + B < -1 \quad (6)$$

$$W \cdot Y_j + B \geq 1 \quad (7)$$

The distance among the closet pattern and hyperplane is maximized by obtaining optimal margins.

3.4.2 Random Forest:

The versatile and simple ML algorithm is Random Forest (RF). Without hyperparameter tuning, a good outcome will be produced the majority of the time. One of the widely used algorithms is RF due to its diversity and simplicity [23]. The tree-structured base classifier with the hierarchical collection is RF. There are many insignificant characteristics in the dataset. Even a few main characteristics are useful for classifier models. The RF method assigns the much more crucial and critical attribute of dengue disease forecasting based on a simple predefined possibility. Breiman created the RF method besides tracing a random selection of function subspaces to extract data subgroups and constructing multiple decision trees. The below points describe the feature and the training data are *MF* and *TD* associated with the RF algorithm.

- Initially, replace and predetermine the probability sample $TD_1, TD_2, ..., TD_k$.
- The decision model tree is constructed for each data TD_k.
- From the available features, the *m*-dimension subspace randomly samples the training health data.
- Based on the features, all possible probabilities are calculated.
- An optimal best data division is produced with the leaf nodes of RF.
- When it attains the saturation criterion, the process is stopped.

3.4.3 Decision Tree:

The features and instances of dengue disease forecasting data are easily divided via decision tree (DT). The decision-making and outcome estimations are determined with a decision tree. The decision tree accuracy is improved and the best parameters and final decision-making criteria are identified with the preliminary decision tree model [24]. Its particular objective was to determine and validate major criteria trying to influence decisions and decision-making pathways using healthcare information DT modeling, which is a quantitative and qualitative research design that incorporates both quantitative as well as qualitative techniques.

3.4.5 Logistic regression:

The relationships among the predictor and target variables of dengue disease data forecasting are determined using the logistic regression model. The below expression illustrates the logistic regression model [25].

$$X = G_\theta(Z) = \theta^t Z \qquad (8)$$

The probability (P) of binary 0 (negative class) and 1(positive class) values are forecasted. The number of input samples X with the vector values Z and Eigenvector G_θ.

$$P(X = 1/Z) = G_\theta(Z) = \frac{1}{1+\exp(-\theta^t Z)} = \lambda(\theta^t Z) \quad (9)$$

$$P(X = 0/Z) = 1 - P\left(X = \frac{1}{Z}\right) = 1 - G_\theta(Z) \quad (10)$$

The value of $\theta^t Y$ is set as the range of 0 to 1 and λ is the sigmoid function.

3.5 Performance metrics

For the prediction we utilized RMSE, MAE, and MSE and MAE [26] can be determined as, the difference between the actual and forecasted dengue rate over the test period. And it determines the mean value of absolute errors among the forecasted and actual real values as shown below,

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|z_i - \hat{z}_i\right| \qquad (11)$$

The forecasted value is $\hat{z}_i$ and the actual value is z_i. The error of MAE is responsible to the magnitude of its proportion. In contrast to this, the RMSE can be evaluated with the square root of the difference between the forecasted and actual values as shown below,

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(z_i - \hat{z}_i)^2} \qquad (12)$$

The average difference between the forecasted and actual value is denoted as MSE and can be expressed as,

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left|z_i - \hat{z}_i\right|^2 \qquad (13)$$

Lower values of RMSE, MSE and MAE indicate the better forecasting model.

3.6 Outbreak Detection

The month in which the dengue disease is outbreak in those cities are forecasted with the aforementioned machine learning approach [27]. the outbreak is estimate with the standard deviation and the exceeding the incidence rate. This outbreak detection is evalauted with the metrics such as Accuracy, Precision, Sensitivity, and Specificity. they are detemined as follows,

$$Accuracy = \frac{Exact\ \Pr ediction}{Total\ \Pr ediction} \qquad (14)$$

$$\Pr ecision = \frac{Exact\ \Pr ediction}{Total\ \ predicted\ \ months\ that\ having\ outbreaks} \quad (15)$$

$$Sensitivity = \frac{Exact\ \Pr ediction}{Total\ number\ of\ outbreak\ months} \quad (16)$$

$$Specificit\,y = \frac{Exact\ \Pr ediction\ \ without\ \ outbreak}{Total\ number\ of\ \ months\ \ without\ \ outbreak} \quad (17)$$

4.Result and Discussion:

The experimental work are handled using JAVA simulator. Table 2 explains the number of cases for dengue forecasted in various Machine Learning Algorithms. Total number of cases is 4600. Among this, the dengue outbreak yes and no cases are tabulated regarding to machine learning algorithm. The actual number of predictive cases is 3356.

Table 2: Dengue disease forecasting results

Various machine learning algorithms	Number of cases	Cases of dengue outbreak (Yes)	Cases of dengue outbreak (No)
Support Vector Machine (SVM)	1500	1128	372
Random Forest (RF)	1000	800	200
Decision Tree (DT)	1300	928	372
Logistic regression (LR)	800	500	300

The number of dengue cases based on the period of October-2018 to October-2022 as shown in Figure 2. The dengue cases based on the different time period is calculated with respect to four cities namely Hyderabad, Chennai, Mumbai and Bangalore. By this plot, we reveal that more number of dengue cases in the time period of October 2018 and April 2029 respectively.

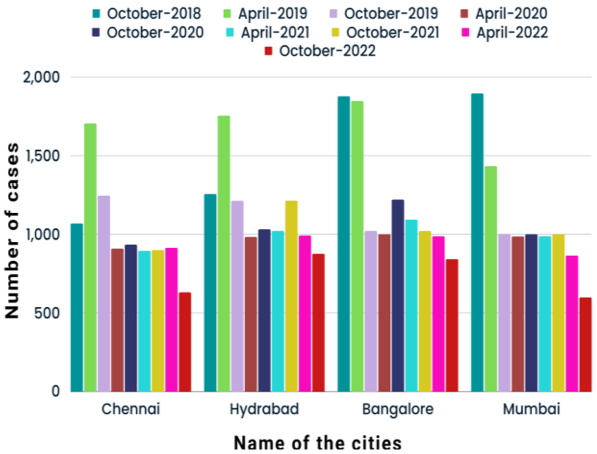

Fig 2: Number of dengue cases based on the period of October-2018 to October-2022

Table 3 describes the dengue forecasting overall results based on different factors. We have collected a data from four cities namely Chennai, Hyderabad, Bangalore, and Mumbai in 16th October 2018 to 1st October 2022. The minimum, maximum and average air temperatures are 20.21°C, 24.01°C and 28.95 °C with the Dew point temperature is19.80 °C. Precipitation or rainfall is 19.71 mm. the proposed method.

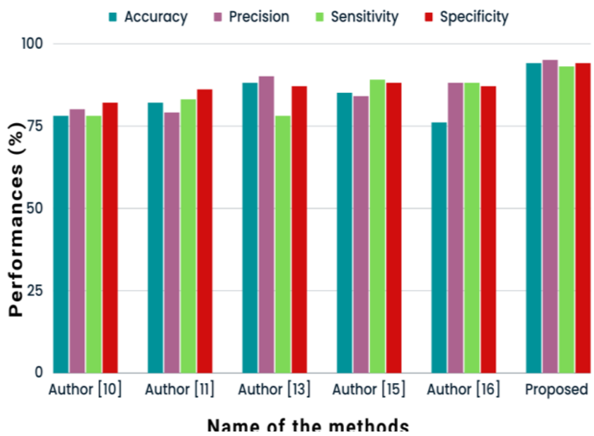

Fig 7: Performance evaluation of accuracy, precision, sensitivity and specificity

The study was conducted over the period from 16th October 2018 to 1st October 2022 in four cities and the statistical analysis based on the dengue fever cases and remote sensitivity data are displayed in table 3. The minimum, maximum, mean values are evaluated based on the data collected from the selected period of time. The NDVI values of all the cities are presented in table 4.

5. Conclusion:

This research work presented a data mining based approach for dengue disease forecasting in geographical areas such as Chennai, Hyderabad, Bangalore, and Mumbai in India by applying machine learning algorithms like Random Forest, Support Vector Machine, Decision Tree, and Logistic regression methods. The four cites of India with respect to as humidity, temperature, and rainfall are collected as inputs. JAVA simulator conducts the simulation works. The proposed method is compared with the authors [10], [11], [13], [15], and [16]. Table 2 explains the number of cases for dengue forecasted in various Machine Learning Algorithms. The actual number of predictive cases is 3356 and the total number of cases is 4600. Among this, the dengue outbreak yes and no cases are calculated regarding to machine learning algorithm. Compared to the existing approaches, the proposed method achieves below 0.9% of MAE and MSE outcomes and also the RMSE outcome is lower. Here we have taken the data' from four cities namely Chennai, Hyderabad, Bangalore, and Mumbai from 16th October 2018 to 1st October 2022. The minimum, maximum and average air temperatures are 20.21°C, 24.01°C and 28.95 °C with the Dew point temperature is19.80 °C. Precipitation or rainfall is 19.71 mm. But the, accuracy, precision, sensitivity and specificity are above 90% for dengue disease prediction outcome is superior to the previous methods.

References:

1. Davi, C., Pastor, A., Oliveira, T., de Lima Neto, F.B., Braga-Neto, U., Bigham, A.W., Bamshad, M., Marques, E.T. and Acioli-Santos, B., 2019. Severe dengue prognosis using human genome data and machine learning. IEEE Transactions on Biomedical Engineering, 66(10), pp.2861-2868.
2. Wirasatriya, A., Sugianto, D.N., Helmi, M., Setiawan, R.Y. and Koch, M., 2019. Distinct characteristics of SST variabilities in the Sulawesi Sea and the northern part of the Maluku Sea during the Southeast Monsoon. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(6), pp.1763-1770.

3. Yang, X., Kumagai, K., Sun, G., Ishibashi, K., Trung, N.V. and Van Kinh, N., 2019, March. Dengue fever screening using vital signs by contactless microwave radar and machine learning. In 2019 IEEE sensors applications symposium (SAS) (pp. 1-6). IEEE.
4. Sivaprasad, A., Beevi, N.S. and Manojkumar, T.K., 2020. Dengue and early warning systems: a review based on social network analysis. Procedia Computer Science, 171, pp.253-262.
5. Jang, C., Park, J.K., Lee, H.J., Yun, G.H. and Yook, J.G., 2020. Non-invasive fluidic glucose detection based on dual microwave complementary split ring resonators with a switching circuit for environmental effect elimination. IEEE Sensors Journal, 20(15), pp.8520-8527.
6. Colorado, H.A., Mendoza, D.E. and Valencia, F.L., 2021. A combined strategy of additive manufacturing to support multidisciplinary education in arts, biology, and engineering. Journal of Science Education and Technology, 30, pp.58-73.
7. Edo-Osagie, O., De La Iglesia, B., Lake, I. and Edeghere, O., 2020. A scoping review of the use of Twitter for public health research. Computers in biology and medicine, 122, p.103770.
8. Edo-Osagie, O., De La Iglesia, B., Lake, I. and Edeghere, O., 2020. A scoping review of the use of Twitter for public health research. Computers in biology and medicine, 122, p.103770.
9. Siriyasatien, P., Chadsuthi, S., Jampachaisri, K. and Kesorn, K., 2018. Dengue epidemics prediction: A survey of the state-of-the-art based on data science processes. IEEE Access, 6, pp.53757-53795.
10. Appice, A., Gel, Y.R., Iliev, I., Lyubchich, V. and Malerba, D., 2020. A multi-stage machine learning approach to predict dengue incidence: a case study in Mexico. Ieee Access, 8, pp.52713-52725.
11. Amin, S., Uddin, M.I., Hassan, S., Khan, A., Nasser, N., Alharbi, A. and Alyami, H., 2020. Recurrent neural networks with TF-IDF embedding technique for detection and classification in tweets of dengue disease. Ieee Access, 8, pp.131522-131533.
12. Latif, J., Xiao, C., Tu, S., Rehman, S.U., Imran, A. and Bilal, A., 2020. Implementation and use of disease diagnosis systems for electronic medical records based on machine learning: A complete review. IEEE Access, 8, pp.150489-150513.
13. Gangula, R., Thirupathi, L., Parupati, R., Sreeveda, K. and Gattoju, S., 2021. Ensemble machine learning based prediction of dengue disease with performance and accuracy elevation patterns. Materials Today: Proceedings.
14. Chamola, V., Hassija, V., Gupta, S., Goyal, A., Guizani, M. and Sikdar, B., 2020. Disaster and pandemic management using machine learning: a survey. IEEE Internet of Things Journal, 8(21), pp.16047-16071.
15. Umar, M., Sabir, Z., Raja, M.A.Z. and Sánchez, Y.G., 2020. A stochastic numerical computing heuristic of SIR nonlinear model based on dengue fever. Results in Physics, 19, p.103585.
16. Phakhounthong, K., Chaovalit, P., Jittamala, P., Blacksell, S.D., Carter, M.J., Turner, P., Chheng, K., Sona, S., Kumar, V., Day, N.P. and White, L.J., 2018. Predicting the severity of dengue fever in children on admission based on clinical features and laboratory indicators: application of classification tree analysis. BMC pediatrics, 18, pp.1-9.
17. Chatterjee, S., Dey, N., Shi, F., Ashour, A.S., Fong, S.J. and Sen, S., 2018. Clinical application of modified bag-of-features coupled with hybrid neural-based classifier in dengue fever classification using gene expression data. Medical & biological engineering & computing, 56, pp.709-720.
18. Sharma, R. and Dutta, A.K., 2011. Malaria and national vector borne disease control programme. The Indian Journal of Pediatrics, 78, pp.1527-1535.
19. Zhang, Z., 2016. Missing data imputation: focusing on single imputation. Annals of translational medicine, 4(1).
20. Cohen, I., Huang, Y., Chen, J., Benesty, J., Benesty, J., Chen, J., Huang, Y. and Cohen, I., 2009. Pearson correlation coefficient. Noise reduction in speech processing, pp.1-4.
21. Battineni, G., Chintalapudi, N. and Amenta, F., 2019. Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (SVM). Informatics in Medicine Unlocked, 16, p.100200.
22. Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P. and Homayouni, S., 2020. Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review. IEEE Journal of Selected Topics in Applied Earth

Observations and Remote Sensing, 13, pp.6308-6325.

23. Sekulić, A., Kilibarda, M., Heuvelink, G.B., Nikolić, M. and Bajat, B., 2020. Random forest spatial interpolation. Remote Sensing, 12(10), p.1687.
24. Park, E.Y., Yi, M., Kim, H.S. and Kim, H., 2021. A decision tree model for breast reconstruction of women with breast cancer: a mixed method approach. International Journal of Environmental Research and Public Health, 18(7), p.3579.
25. Zhu, C., Idemudia, C.U. and Feng, W., 2019. Improved logistic regression model for diabetes prediction by integrating PCA and K-means techniques. Informatics in Medicine Unlocked, 17, p.100179.
26. Chai, T. and Draxler, R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. Geoscientific model development, 7(3), pp.1247-1250.
27. Nguyen, V.H., Tuyet-Hanh, T.T., Mulhall, J., Minh, H.V., Duong, T.Q., Chien, N.V., Nhung, N.T.T., Lan, V.H., Minh, H.B., Cuong, D. and Bich, N.N., 2022. Deep learning models for forecasting dengue fever based on climate data in Vietnam. PLoS Neglected Tropical Diseases , 16(6), p.e0010509.