How to Cite:
Chengathir, S. M., Bhuvaneswari, T., Maruthupandi, J., & Naga, P. R. (2022). Prediction of dengue using data mining classification algorithms. *International Journal of Health Sciences*, *6*(S1), 11860–11871. https://doi.org/10.53730/ijhs.v6nS1.7907

Prediction of dengue using data mining classification algorithms

Chengathir Selvi M
Dept. of Computer Science and Engineering, Mepco Schlenk Engineering College, Sivakasi - 626005, Virudhunagar District, Tamil Nadu, India
Corresponding author email: chengathir@gmail.com

Bhuvaneswari T
Dept. of Computer Science and Engineering, Mepco Schlenk Engineering College, Sivakasi - 626005, Virudhunagar District, Tamil Nadu, India

Maruthupandi J
Dept. of Information Technology, Mepco Schlenk Engineering College, Sivakasi - 626005, Virudhunagar District, Tamil Nadu, India

Naga Priyadarsini R
Dept. of Computer Science and Engineering, Mepco Schlenk Engineering College, Sivakasi - 626005, Virudhunagar District, Tamil Nadu, India

***Abstract*---**Dengue is a life-threatening disease prevalent in several developed as well as developing countries like India. This is a virus born disease caused by breeding of Aedes mosquito. Datasets that are available for dengue describe information about the patients suffering with dengue disease. Dengue disease has symptoms like: Fever Temperature, WBC, Platelets, Severe Headache, Vomiting, Metallic Taste, Joint Pain, Appetite, Diarrhea, Hematocrit, Hemoglobin, and how many days suffer in different city. The main objective of this paper is to classify dengue data and assist the users in extracting useful information from data and easily identify a suitable algorithm for accurate predictive model from it. The proposed system is to determine the prediction of dengue disease and their accuracy using classifications of different algorithms to find out the best performance. Data mining is a well-known technique used by health organizations for classification of diseases such as dengue, diabetes and cancer in bioinformatics research. IBM Watson Analytics is used to analyze the influence of different parameters on the given data set. In the proposed approach, R programming is to evaluate data and compare results. In this paper, the dengue data set are classified and a comparative analysis of different data mining classification algorithms such as Support Vector machine, Navies Bayes and Random Forest

Manuscript submitted: 27 March 2022, Manuscript revised: 9 April 2022, Accepted for publication: 18 May 2022

are analyzed. From the analysis, Random Forest classifiers produce high accuracy of 96.32% in dengue classification.

***Keywords*---**Dengue, classification, Random Forest, IBM Watson Analytics.

Introduction

Dengue infection has endangered 2.5 Dengue infection has endangered 2.5 billion populations all around the world. Every year there are 50 million people who suffer from it globally. Pakistan has been victim of this rapidly growing sickness from last few years. Since 2007 in Pakistan, large number of cases was marked especially in Lahore. In 1994 at Karachi Pakistan's first case of dengue was appeared and Dengue's outbreak in 2011, that was more life-threatening than preceding years and 1400 people were affected. Dengue fever is a disease caused by dengue virus and known as break bone fever that is transmitted by Aedes mosquito. Dengue virus causes life threatening dengue hemorrhagic fever whose symptoms include bleeding, low levels of blood platelets, low blood pressure, metallic taste in mouth, headache, muscle joint pain and rashes. Dengue fever occurs in form of cycles and this cycle is present inside the body of an infected person for two week or less than two weeks. It causes abdominal pain, hemorrhage (bleeding), and circulatory collapse and Dengue hemorrhagic fever. Dengue is divided into two types, i.e., type 1 and type 2, according to world health organization. First one is classical dengue called dengue fever and the other is dengue hemorrhagic fever. DHF1, DHF2, DHF3 and DHF4 are further four types of dengue hemorrhagic fever. DHF is revealed by start of fever which continues for 2 to 7 days with number of signs like leakage of plasma, shock and weak pulse. In earliest cases it's hard to differentiate dengue fever from dengue hemorrhagic fever. The main objectives of this paper are:

(1) To extract useful classified accuracy for prediction of dengue diseases.
(2) Comparison of different data mining algorithms on dengue dataset.
(3) Identify the best performance algorithm for prediction of diseases.

In this paper, dengue dataset collected from Primary Health center is used for classification method. The major steps include collection of the dengue dataset, classification of dengue data set using different data mining classification algorithms and then comparison of results of various classification algorithms.

Related Work

Kamran Shaukat, et al. (2015) [1] proposed a Dengue fever prediction: a data mining problem. The main objective of this paper is to classify the dengue fever dataset to evaluate and compare their performance using WEKA. They have used methods like Naïve bayes, J48, SMO, REP. The datasets were collected from District Headquarter Hospital (DHQ) JHELUM. The rate of reorganization is based on the evaluation metrics like TP rate, FP rate, Recall, Precision, F- measure. This paper concludes that the Naïve bayes, J48 are the top performance classifier techniques has achieved an accuracy of 92% and 88% respectively. In future the

system can be extended to implement the other different kinds of algorithm and can use other tools for prediction.

M Bhavani, et al. (2017) [2] proposed a Data mining approach for precise diagnosis of Dengue fever. The main objective of this paper is to calculate the performance of various classification techniques and compare their performance. They have used methods like REP Tree, J48, SMO and Random Forest. They collected the Dataset of dengue that gives the information about the patients suffering with the dengue disease. At the beginning stages it is difficult to differentiate the dengue fever and dengue hemorrhagic fever. Several data mining techniques are used for the prediction of dengue fever. From this paper it is found as SMO techniques have 84% of accuracy. This paper concludes that these techniques were evaluated and their performance were compared. In future the system can be extended to support the different algorithms.

YuhanisYusof et al. (2011) [3] proposed a Dengue Outbreak Prediction: A Least Squares Support Vector Machines Approach prediction model that incorporates Least Squares Support Vector Machines (LS-SVM) in predicting future dengue outbreak. They have used methods like Decimal Point Normalization, Dengue fever, Least Squares Support Vector Machines, Support Vector Machines. In this paper they have taken Data sets used in the undertaken study includes data on dengue cases and rainfall level collected in five districts in Selangor. The implementation phase of this paper includes collecting raw data, normalization Pretreatment, Training and testing with LS-SVM model, comparison of prediction accuracy and MSE between proposed model and NNM. This paper concludes that the prediction of the dataset shows that the LS-SVM prediction model outperformed the Neural Network model in terms of prediction accuracy and computational time. In future the system can be extended to support actual prediction scenario, there are several numbers of factors having their influence on dengue outbreak, such as humidity, temperature and cloudiness.

M.Krishna Satya Varma et al. (2015) [4] proposed a Dengue data analysis using decision tree model to create predictive models, using decision tree algorithm. They have used methods like supervised classification model, unsupervised clustering model. Unsupervised clustering model were used for identifying the significant characteristic of insolvent customers and supervised classification model for insolvency prediction. The dataset contains the details of male, female belonging to different villages. The study shows that out of 344 dengue cases 48.93% are tribal areas. Dengue cases are reported mainly in Tribal and hill areas. This paper concludes that decision tree algorithm helped in discovering rules for data mining applications were applied and data were analyzed using decision tree model. In future the system can be extended to support the prediction in different areas.

S.Freeda Jebamalar et al. (2017) [5] proposed a survey on prediction of Dengue fever using data mining techniquesto make a survey on application of data mining techniques. They have used methods like Naive bayes, J48 classifier and clustering algorithm like Kmediods, Dbscan, k-means etc., to accurately predict dengue disease. The author of this paper took the dataset from Dharmapuri health center and used Infomatica tool for the analysis and to detect the affected

patients earlier. He then applied the data mining techniques on the collected dataset and he had found that the NMTool generated Decision Tree gives 100% accuracy. This paper concludes that there is a reduced time for prediction of disease. In future the system can be extended to support to improve the accuracy by incorporating some more features and latest datamining techniques.

Sandeep K.Sood et al. (2017) [6] proposed a Wearable IoT sensor based health care system for identifying and controlling chikungunya virus to generate alert to the government and healthcare agencies to control the outbreak of CHV in infected regions. They have used methods like wearable IoT sensor, fog computing, mobile computing, cloud computing, fuzzy-c means, social network analysis. The identifying and controlling the spread of CHV was a slow approach. Cloud computing enhances the quality of remote health care services. Fog computing provides benefits like low latency, location awareness, high mobility. The fuzzy-c means diagnose the possibly infected user and generate alert to users from fog layer. Social Network Analysis represent the state of CHV. This paper concludes that alerts are generated to the government organization to prevent the outbreak of these virus. In future the system can be extended to achieve bandwidth efficiency, minimum efficiency time and minimum delay.

Gunasekaran, Manogaran et al. (2017) [7] proposed a new architecture of IoT and big data ecosystem for secured smart health care monitoring and alerting system to store and process scalable sensor data for health care application. They have used methods like Meta Fog redirection-Apache pig and HBase for collecting and storing data, Grouping and choosing-for secure integration of fog computing with cloud. Wearable device will generate a large amount of structured and unstructured data that is complex to process, analyze, store data to make decision and to provide data security so they have used Bigdata. The “Map reduce” is used as a prediction model. The “performance parameters” like throughput, sensitivity and accuracy were used to prove the efficiency. The fog computing acts as a distributed computing infrastructure that services network edge and remote data center. In this paper they have used Big data as background (volume, variety and velocity), cloud computing, fog computing. They have used MF-R with GC architecture for data collection and transfer phase. This paper concludes with the usage of HBase for cloud storage. In future the system can be extended to use stochastic Gradient descent algorithm with logistic regression for developing the prediction model.

Prabal verma et al. (2017) [8] proposed a Cloud-centric Iot based disease diagnosis healthcare framework to develop a m-healthcare disease diagnosing frame work to predict the potential disease with its level of severity. This application provided a platform to millions of people to get regular health updates for a healthier lifestyle. They have used methods like user-oriented, UCI dataset. Key terminologies are defined to generate user-oriented health measurements by exploring the concept of computational sciences. Student perspective health data is generated using UCI dataset and medical sensors to predict the student with different disease severity. M-Health Monitoring system consists of three phases, 1. Data collection, 2. Decision making 3. Alert generation. The main motive of proposed health diagnosis system is to generate student diagnosis result (SDR) based on the health measurements collected by medical IoT devices. The

diagnosis process can be made more effective and reliable. This paper concludes that the patterns, scale, and frequencies-based diagnosis results play a significant role in identifying a person with potential disease type. In future the system can be extended as the theoretical foundation for providing better healthcare services in the digital world.

Jorge Gomez et al. (2016) [9] proposed a Patient monitoring system based on IoT monitoring the health and workout routine recommendation to patient with chronic disease such as diabetes, heart and pressure problems. They have used methods like context model, architecture performance. They used M-health and E-health to improve, help and assist health. The context model makes use of person, time, device, type of disease and location. The architecture performance makes use of server, detector context, reasoning engine, client. The related work focused on monitoring, diagnosis, treatment of patient remotely to improve the quality of life of patients. Case study dealt with validating the efficiency of the system using applications. This paper concluded that the system developed was used to monitor the patients with chronic disease that improved the quality of life of patients. The context model proved to be efficient when making interfaces. In future the system can be extended with having mobile applications.

Emmenual Reddy et al. (2015) [10] proposed a mobile application for dengue fever monitoring and tracking via GPS: case study for Fiji to design a Dengue monitoring and tracking system based on GPS. They have used methods like Dengue virus, Dengue monitoring, Dengue tracking, GPS, Mobile application. Initially they had used methods to provide an alternative and more convenient method for checking symptoms on dengue patients. GPS technology is also used to track the location of the dengue infected patients. The application can be improved to ensure real-time tracking of dengue outbreaks and compatibility of the application with iOS and windows mobile platform. The dengue fever monitoring and tracking application are to provide the general public with a dengue symptoms checker, receive feedback of the papaya leaf remedy, self-report of Dengue case and to store the location of the user via GPS. The data which is stored in the server. This paper conclude that the data can be retrieved from the database and displayed on the Dengue tracker website and can be used by the relevant authorities to clear and fumigate affected areas. In future the system can be extended to involve the implementation of the application for large scale use of thousands of users.

Zankhana Mehul Kalarthi (2016) [11] proposed a Review paper on smart healthcare system using IoT provide a quality healthcare to everyone, error free and smooth communication to patients. They have used methods like IoT, E-health, Arduino uno, cloud computing. The function of Arduino uno is to transfer data to cloud with WiFi module, data are stored into MySql database that manages data and provides accessibility. The cloud computing handles authentication, privacy, security, data management. Smart health performed the implementation of different biometric sensor that captured human body parameters. The system flow works as data acquisition, cloud system, real time health portal. The hardware components used were heart beat sensor, temperature sensor, WIFI shield, micro controller. The paper concluded by providing quality healthcare to everyone using remote monitoring system. In

future the system can be extended by providing more sensor and settings to android application and notify the patient by email if data are abnormal.

Alexandre Santos et al. (2014) [12] proposed a IoT and smart object for m-health monitoring and control to provide security, mobile communication i.e., one can remotely take care of patient, establish an ubiquitous ambient assisted living for mobile health application. They have used methods like IoT, RFID M-Health, Ambient assistant living. IoT and RFID enables new context that combine physical and virtual existences. The use of smart phone with Internet access provides strong security concern such as Authentication, Confidentiality, Integrity. The "IoT for M-health" focuses on automatic and easily serialized identifier that has become the main driver for an IoT with the aid of global and generic identification system such the Electronic Product Code [EPC]. The "Security and M-health" deals with security requirements to prevent unauthorized access that attempt to private information and it is important to keep an updated log. For implementation they have used Object name service prototype, Security protocol for M-Health, RFID for indoor location awareness and guidance. They have concluded a Simple and secure IoT architecture that is aimed at establishing a generic and ubiquitous ambient assisted living frame work to be used by mobile health application. In future the system can be extended to incorporate health facilities. This enables testing in real scenario, Encrypted communication and other global security levels.

Shivam Gupta et al. (2017) [13] proposed a IoT based patient health monitoring system to develop a system for home use by patients that are not in critical conditions that needed to be timely monitored by doctor or family. They have used methods like embedded system, AVR micro controller, IoT and patient health monitoring sensor. The fixed monitoring system can be used only when the patient is lying on the bed, these systems are huge and only available in hospital. They have used heart beat monitoring sensor, ECG sensor, WiFi module and LCD 16x2.The wireless sensor network is a wireless network that consist of structurally distributed device that uses sensor to monitor physical and environmental condition. This paper concluded with the development of microcontroller-based system for wireless heart beat and temperature monitoring. In future the system can be extended to have mobile application that sends data to the doctor and immediate remedy related to the healthcare will be taken.

Samir V.Zanjal et al. (2016) [14] proposed a Medicine remainder and monitoring system for secure health using IoT to do medicine prescription through web for secure health using IoT. They have used methods like home health care, dementia, remote monitoring medicine remainder, RFID authentication. The "Home health care" focuses on communication, image sensing and human computer interaction techniques that are used for diagnosis. The paper review technology of home health care has been replaced with web. The "low-cost medical sensing" provides active and real appointment of patients, hospitals, care takers and doctors that uses secure data transmission from source to destination. For "data transmission" they used messaging standards and communication protocol. The system functions as generating alarm according to the schedule and situation and record the data with the help of sensor and that helps in remote monitoring. This paper concludes that the usage of medicine remainder and

monitoring system that provides further description based on the recorded data. In future the system can be extended to provide security using encryption and decryption in RFID.

K.Natarajan et al. (2016) [15] proposed a Smart health care system using IoT [15] to enable efficient machine to machine communication for health care system. They have used methods like IoT, machine to machine communication, pulse oximeter, Entity Oriented Resource. Initially they have used IoT to connect various smart objects through internet, this has spurred the increase of real time data. Using this system we can monitor glucose level, blood pressure, electro cardiogram, body temperature through raspberry pi. The network architecture of this system focuses on three layers they are smart medicine service layer, medical resource layer, sensor data collection layer. The Entity Oriented Resource (EOR) contains <Required person, address, authentication>. Raspberry pi collect and store the medical data. Collected data are transferred to the user through the application. This paper concludes the use of IoT technology in health care to access the wide range of data source using machine to machine communication. In future the system can be extended to use big data accumulated by IoT devices.

Chengathir Selvi Murugesan et al. (2017) proposed a wireless based health care monitoring system for prediction of patients' status through various parameters such as body temperature, pulse and body movement. They proposed a real time monitoring of patients' physiological conditions through wireless sensors. The healthcare professionals can easily monitor and access patients' records easily at any time through the web app developed. It mainly helped for personalization of treatment and management

Proposed Methodology

In this proposed work, the prediction of dengue using data mining classification algorithms are analyzed. The major phases of this proposed work are as follows 1. Collection of data 2. Data cleaning 3. Data classification 4. Best algorithm for data prediction. The following figure shows the proposed system for Dengue prediction and Classification.

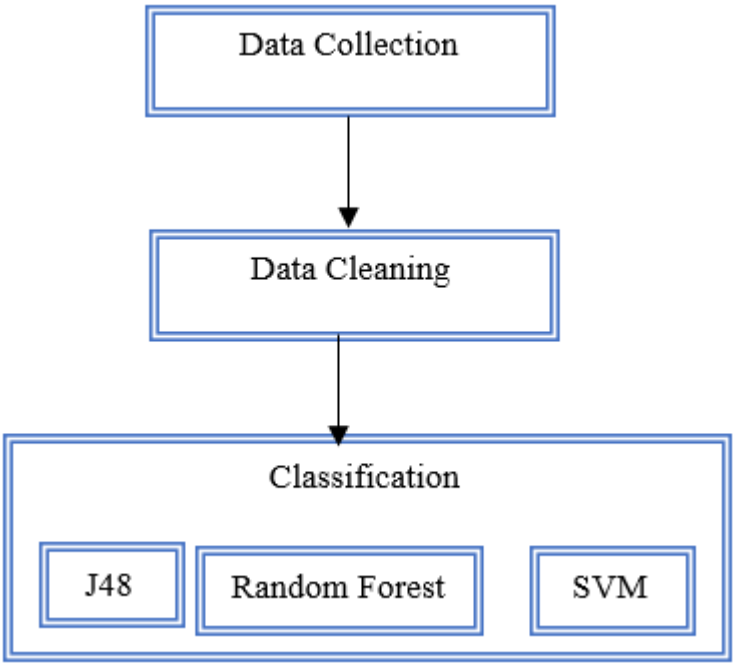

Figure 1. Dengue Prediction and Classification

The prediction of dengue using data mining algorithm are discussed as follows:

Data Collection:
In this Data Collection, Dengue data are collected from primary health care centre. The data that are collected are real time data. The data set consists of 400 records with 9 attributes (Name, Age, Gender, Fever, Myalgia, Flu, Fatigue, Platelet count, Results).

Data cleaning
The data collected from real world are not clean so data pre-processing is an important process. Data cleaning removes all the unwanted or empty records

Classification techniques
The data mining classification algorithms (Random Forest, SVM, Naive bayes, J48) are used to find out the accuracy of dengue disease.

Find out best algorithms
The accuracy values obtained from different algorithms are compared to find out which algorithm gives us more accuracy. The goal is to determine which algorithm works well in predicting the dengue disease.

Experimental Results and Discussion

The following figure 2 shows the Dengue Data set collected from Primary Health center is shown. The different parameters such as EPID, fever, Bleeding, Myalgia, Flu, Fatigue and Results are listed for 400 patients.

dengue

Filter

	EPID.	Fever	Bleeding	Myalgia	Flu	Fatigue	Results
1	1	Yes	No	Yes	No	Yes	Positive
2	2	Yes	No	Yes	Yes	Yes	Positive
3	3	Yes	No	Yes	No	Yes	Positive
4	4	Yes	No	No	No	Yes	Positive
5	5	Yes	No	No	No	No	Negative
6	6	Yes	No	No	No	Yes	Negative
7	7	Yes	No	Yes	No	Yes	Positive
8	8	Yes	No	No	No	No	Negative
9	9	Yes	Yes	Yes	No	No	Negative
10	10	Yes	No	Yes	No	No	Positive
11	11	Yes	Yes	No	No	Yes	Positive
12	12	Yes	No	Yes	No	Yes	Positive
13	13	Yes	Yes	Yes	No	No	Positive
14	14	Yes	No	Yes	No	Yes	Negative
15	15	Yes	No	No	No	Yes	Positive

Figure 2. Dengue Data set

Figure 3 Platelet Count VS Age

Figure 4 Platelet Count VS Age and Bleeding

Using IBM Watson Analytics, different analysis is carried out to find the influence of different parameters. The figure 3 to 7 shows different analysis such as the platelet counts VS Age of the collected Dengue Dataset, Platelet Count VS Age and Bleeding, Platelet Count VS Age filtered by flu, Platelet Count VS Age filtered by Fever and Platelet Count VS Age filtered by Fever respectively. .

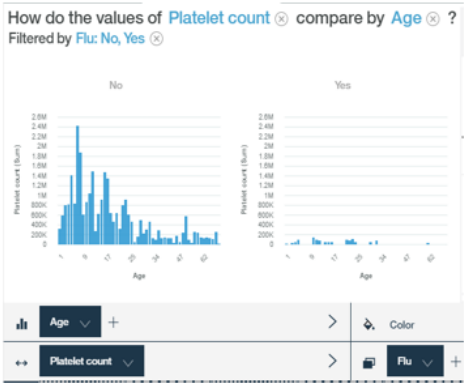

Figure 5 Platelet Count VS Age filtered by flu

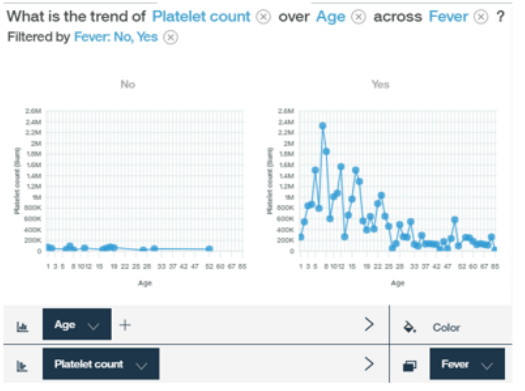

Figure 6 Platelet Count VS Age filtered by Fever

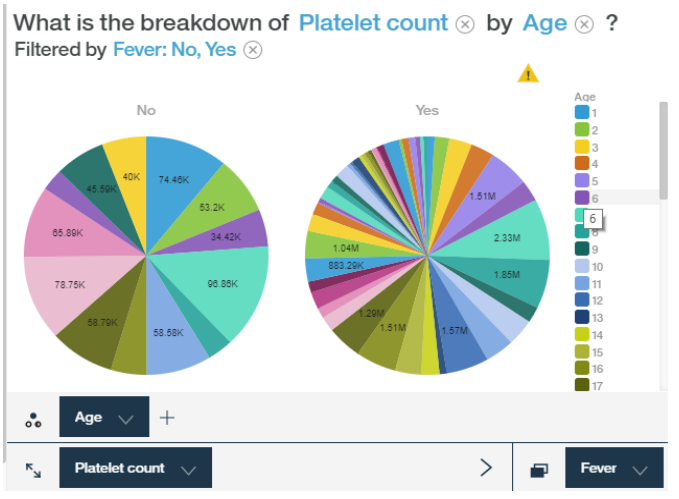

Figure 7 Platelet Count VS Age filtered by Fever

The following Figure 8 shows prediction of different classification algorithms such as SVM, J48 and Random Forest. The metrics used for evaluation are Precision, Recall, F1 Score, Error Rate and Accuracy. From the classified output, Random Forest classifier produced 96.32% classification accuracy.

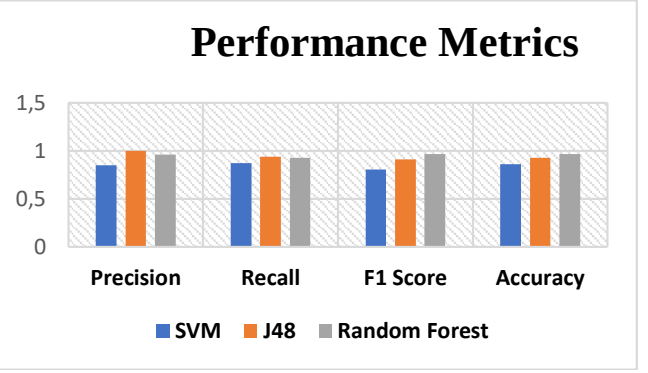

Figure 8. Performance Metric

Conclusion

The spread of dengue is a major problem in India. Most of the data mining algorithms implemented previously for the prediction of dengue requires some modification in the tool used to properly predict the spread of the dengue. The properties to predict the spread of dengue were tested using WEKA tool. This tool is fully automated and does not require any interaction once it is started. This tool is quite simple; the response time taken may depend on the data available in the data set. The proposed system for the prediction of dengue using data mining classification algorithm uses RStudio for processing the data in the dengue data set takes minimal time to process the data and Random Forest classifier produced accuracy of 96.32%.

References

1. Kamran Shaukat et.al., (2017), “Dengue Fever Prediction: A Data Mining Problem”, Journal of Data Mining in Genomics & Proteomics, Volume 6, Issue 3, Pg.no 1000181.
2. M.Bhavani, S.Vinod kumar, (2017), “A Data Mining Approach for precise diagnosis of Dengue fever”, International Journal of Latest Trends in Engineering and Technology, Vol.(7), Issue(4), Pg.no 352-359.
3. Yuhanis Yusof, Zuriani Mustaffa, (2011), " Dengue Outbreak Prediction: A Least Squares Support Vector Machines Approach", International Journal of Computer Theory and Engineering, Vol. 3, Pg No. 4.
4. M Krishna Satya Varma, (2015), “Dengue Data Analysis using decision tree model”, International Conference on Emerging Trends in Science Technology Engineering and Management.
5. S.Freeda Jebamalar, R. Varatharajan, Daphne Lopez , et al., (2017), “A Survey on Prediction of Dengue Fever Using Data Mining Techniques”, International Journal of Science, Engineering and Management (IJSEM), Vol 2, Issue 12, Pg.no-2456 -1304.
6. Sandeep K. Sood, Isha Mahajan, (2017), “Wearable IoT sensor-based healthcare system for identifying and controlling chikungunya virus”, IEEE transaction on Computers in Industry.

7. Gunasekaran Manogaran, et al., (2017), “A new architecture of Internet of Things and big data ecosystem for secured smart healthcare monitoring and alerting system future Generation Computer Systems”, IEEE transaction on Future Generation Computer Systems.
8. Prabal Verma, Sandeep K. Sood, (2017), “Cloud-centric IoT based disease diagnosis healthcare framework”, IEEE transaction on Parallel and distributed computing, The 7th International Conference on Ambient Systems, Networks and Technologies.
9. Jorge Gómeza, Byron Oviedob, Emilio Zhumab, (2017), “Patient Monitoring System Based on Internet of Things”, IEEE transaction on Procedia Computer Science, Pg.no 90 – 97.
10. Emmental Reddy et al., (2015), “Mobile application for dengue fever monitoring and tracking system based on GPS”, International Research Journal of Engineering and Technology (IRJET), Volume: 03, Issue: 04.
11. Zankhana Mehul Kalarthi, (2016), “A Review paper on smart health care using Internet of things”, International Journal of Research in Engineering and Technology, Volume: 05, Issue: 03.
12. Alexandre Santosa, et al., (2014), “Internet of Things and Smart Objects for M-Health Monitoring and Control”, IEEE transaction on Procedia Technology.
13. Shivam Gupta, et al., (2017) “IOT based Patient Health Monitoring System”, International Research Journal of Engineering and Technology (IRJET), Volume: 04, Issue: 03.
14. Samir V.Zanjala, Girish. R. Talmaleb, (2015), “Medicine Reminder and Monitoring System for Secure Health Using IOT”, IEEE transaction on Procedia Computer Science, Pg.no 471 – 476.
15. K. Natarajan et al., (2016), “Smart Health Care System Using Internet of Things”, Journal of Network Communications and Emerging Technologies (JNCET), Volume 6, Issue 3.
16. C. S. M., T. D. Rajeeve, A. J. P. Antony and P. T., "Wireless Sensor based Healthcare Monitoring System using Cloud," *2017 International Conference on Inventive Systems and Control (ICISC)*, 2017, pp. 1-6, doi: 10.1109/ICISC.2017.8068710.