TEST Engineering & Management

May – June 2020
ISSN: 0193-4120 Page No. 5684-5689

Predictive Analytics of Dengue Disease using Machine Learning Classification Approaches

[1]Dr. Suneetha Keerthipati, *Associate Professor & Head, Department of MCA, SVEC,umasuni.k@gmail.com*
[2]Dr. M. Lavanya, *Associate Professor, Department of CSE, SVEC, mlavanyamca.2003@gmail.com*
[1]Ms. Manasa Kommineni, *Assistant Professor, Department of MCA, SVEC, manasa8486@gmail.com*

Article Info
Volume 83
Page Number: 5684 - 5689
Publication Issue:
May - June 2020

Article History
Article Received: *19 November 2019*
Revised: *27 January 2020*
Accepted: *24 February 2020*
Publication: *17 May 2020*

Abstract

The main objective of proposed system is to develop a system which can use the existing related Information to answer the causes of dengue and make use of this information to forecast the dengue occurrence within a specific region so that the medical experts can predict, control and manage the epidemic of disease at the earliest. Dengue disease is one of the major disease which causes to deaths for the many people with lack of experience. According to the surveys there are 390 million dengue cases reported per year. By collecting the data from various repositories and with the combination of predictor variables such as meteorological data, disease surveillance, health data and socioeconomic data one can able to predict the precision of the model. This disease majorly effect on the Asian and American regions. Dengue disease is caused by the Aedes aegypti and Aedes albopictus mosquito bite, it spread other viruses such as yellow fever, chickengunya. Now Dengue has become the global problem for most of the countries. The best way to treat the disease is to predict the symptoms early and take necessary precautions to prevent the disease. In this paper, we present various machine learning algorithms to detect the dengue at the early stages which will be helpful to avoid from death. This paper focuses on better understanding of geospatial insights, health care monitor and management systems.

Keywords: *Machine learning, Dataset, Support Vector Machine (SVM), Decision Tree.*

1. INTRODUCTION

Dengue is a disease which is caused by mosquito bite found in the low temperature and storage of water in long period areas, transmitted through Aedes Aegypti mosquito and Aedes lbopictus mosquitoes. In rainy season is the one of the major causes for growing mosquitoes and that carry these diseases to breed. According to WHO (World Health Organization) every year across worldwide three ninety million dengue cases are present. According to the health ministry reports it is majorly spread in the month of the October and majority of infections are from Orissa ,West Bengal, Karnataka and Kerala. According to a survey metrics jumped from 3,135 in 2018 to 3,345 cases in 2019.

- Every year the no. of deaths are increasing due to dengue fever which can be identified within 6-7 days of time with the symptoms of getting fever and body pains. This virus enters into blood and gets infected which causes headache, rashes, body pains, tiredness, fever and queasiness. The sever dengue symptoms like High fever with internal as well as external bleeding.

- Dengue disease cannot spread directly from person to person, it is virus infected disease it spread only through Mosquito bite. The main aim of this paper is to predict the dengue disease using machine learning algorithms. It works initially by identifying the symptoms of a patient and with this identification it predicts the seviority of the disease.

TEST Engineering & Management

May – June 2020
ISSN: 0193-4120 Page No. 5684-5689

2. LITERATURE SURVEY

Literature Survey is the process of reading, analysing and summarizing about a specific topic and it is not to solve the problem but gathering the information from the various methods like websites, Articles so on.

- ***https://www.ncbi.nlm.nih.gov/pmc/?term=dengue***

This website provides information about Dengue disease, economic impacts of disease and Different Symptoms caused for the disease. It also provides various methods for detecting the disease and prevention methods. It provides probability of dengue transmission and propagation in a non-endemic temperature area. Disease Burden of Dengue disease in various countries and Estimation of dengue cases using the symptoms.

- ***https://nvbdcp.gov.in/dengue***

This site provides Dengue /DHF situation in India, affected states and number of dengue cases present in every year from 2015 to 2019. The information such as infected persons and death rate of each state in India. Monitoring of dengue cases and generate reports send it to the governments to take necessary actions to minimize the disease rate.

- ***http://www.indiaenvironmentportal.org.in/dengue***

This site Provide the information about dengue disease such as disease rate in major cities of India and death rate. Given the suggestion to the society to clean and maintain required actions to avoid from the mosquito bites.

3. PROPOSED WORK

3.1 Generation of Dengue Data sets:

Data Collection:

For this proposed work, the data sheets are collected from government hospitals, doctors and medical agencies. The dataset is prepared by considering the following attributes of medical data.

uhid : UMR3421			Ipno : 1975215422
patient name : Mr.John			registered on : 10/12/2019
age/gender : 32/ MALE			collected on : 10/12/2019
doctor : DR.d.Sudarshan			reported on : 10/12/2019
HAEMATOLOGY	**observed values**	**unit**	**Biologicla Recerence Interval**
Platelet Count(EDTA Blood/Automated Blood cell Counter	1.8	lakh/cu.mm	1.5 - 4.5
Total WBC Count(TC)	4890	cells/cu.mm	400-12000
Differential Leucocyte Count			
Neutrophils(EDTA Blood/Automated Blood cell Counter	37.6	%	30-60
Lymphocytes	48.1	%	30-60
Eosinophils	4.5	%	16
Monocytes	9.2	%	28-Jan
Basophils	0.6	%	00-1
INTERPRETATION: Tests done on Automated cell Counter			
		MicroBiology	
Dengue Profile			
Dengue NS1 Antigen	Positive		
Dengue IgG Antibodies	Negative		
Dengue IgM Antibodies	Negative		

3.2 Proposed Work:

The various symptoms, algorithms and a variety of medical parameters are used to detect the dengue disease at the early stages.

3.2.1 Symptoms:

The symptoms are majorly identified during the dengue disease. Severe dengue causes internal and external bleeding. For severe dengue we should provide better treatment to the infected person otherwise which lead to death.

May – June 2020
ISSN: 0193-4120 Page No. 5684-5689

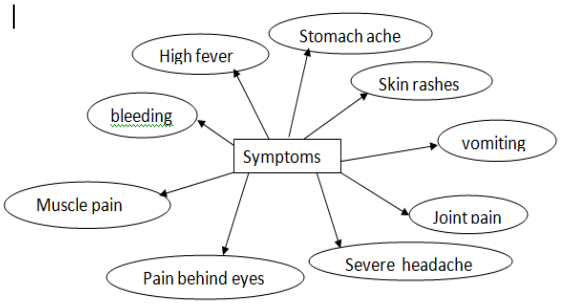

Fig 1: Symptoms for Dengue disease

By observing these symptoms of a person we can predict whether the person is infected with dengue disease or not. Sometimes these symptoms illustrate other diseases.

3.2.2 Medical parameters:

For this proposed work we collected and observed various medical parameters, such as platelet count, RBC and WBC count, Neutrophils, lymposites, Eosinophils, Monocytes, and Basophiles. Dengue tests such as NS1 Antigen, IgG Antibodies, IgM Antibodies. If anyone of the test is positive then it would be treated as dengue disease.

3.2.3 Knowledge flow:

The process of the proposed work is shown in figure 2. The original collected data is divided into two parts such as training dataset and Test dataset. For Training dataset machine learning algorithms were applied and then chosen the best accuracy algorithm, from the model to predict the results.

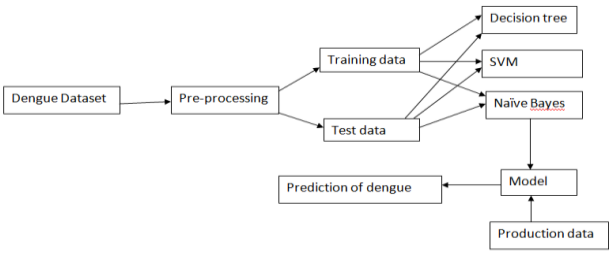

Fig 2: Knowledge flow for Dengue Disease

3.2.4 Algorithms used:

To implement this we used Naïve Baye's algorithm, Decision Tree algorithm and Support Vector Machine (SVM) algorithm for effective prediction of dengue disease.

A. Naïve Baye's Algorithm:

It's a supervised learning and statistical method used for classification. This algorithm requires a minimal training data set as an input to identify the parameters required for classification task. To estimate the posterior probability of each class we used Naive Bayesian equation to predict the class having the highest posterior probability as an outcome.

i) Confusion Matrix: The following is the confusion matrix plotted on Actual vs Predicted data used to predict the dengue disease based on its positive or negative values which will help us in doing correct predictions of dengue disease.

	Positive prediction	**Negative prediction**
Positive Actual	True +ve	False -ve
Negative Actual	False +ve	True –ve

- **True Positive (TP) :**It correctly predicts as positive means that they are having dengue disease.
- **True Negative (TN) :**It correctly predicts as negative means that they don't have dengue disease.
- **False Positive (FP):** It incorrectly predicts as positive means that they have dengue disease.(type 1 error).
- **False Negative (FN) :** It incorrectly predicts as negative means that they don't have dengue disease (type 2 error).

The below figure 3 shows the results of confusion matrix by applying Naïve Bayesian algorithm:

TEST Engineering & Management

May – June 2020
ISSN: 0193-4120 Page No. 5684-5689

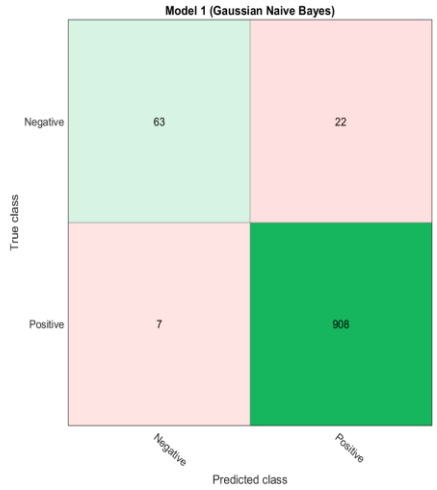

Fig 3: Confusion matrix (Naïve Bayesian algorithm)

ii) Parallel Coordinator:

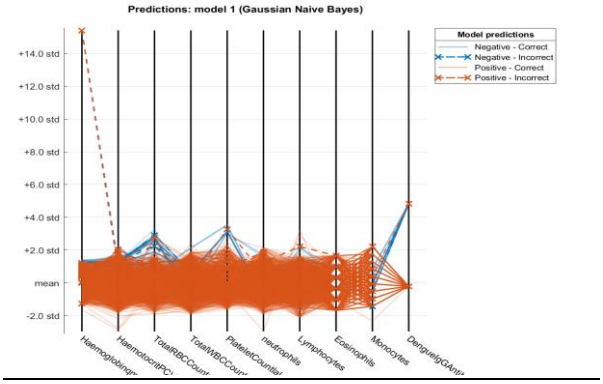

Fig 4: Parallel Coordinator (Naïve Bayesian algorithm)

iii) ROC Curve:

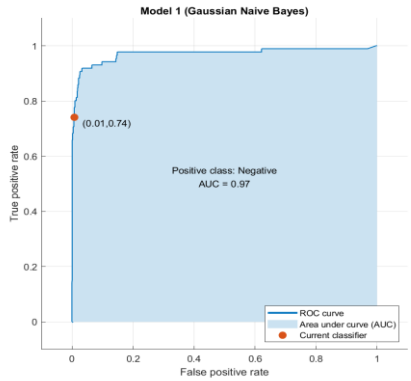

Fig 5: ROC Curve (Naïve Bayesian algorithm)

iv) Scatter plot:

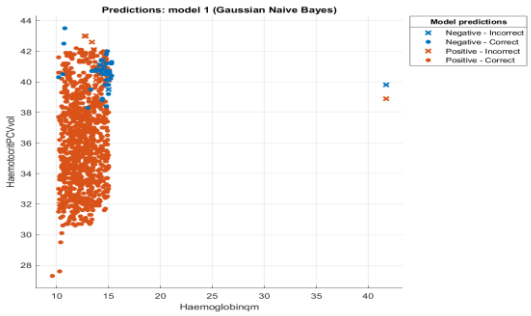

Fig 6: Scatter plot (Naïve Bayesian algorithm)

B. Support Vector Machine (SVM):

SVM is used for both regression and classification of data in a flexible manner. SVM can be applied when the data has exactly two classes. Hyper plane will be generated by SVM iteratively which lead to minimize the error. SVM divides the dataset into classes for finding the maximum marginal hyper plane.

i) Confusion Matrix:

Fig 7: Confusion matrix (Linear SVM)

ii) Parallel coordinator:

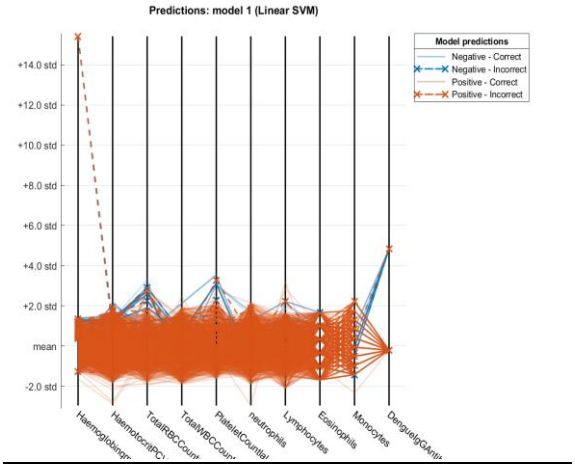

Fig 8: Parallel Coordinator (Linear SVM)

iii) Roc curve:

TEST Engineering & Management

May – June 2020
ISSN: 0193-4120 Page No. 5684-5689

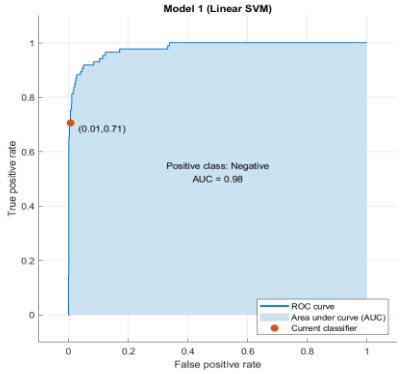

Fig 9: Roc curve (Linear SVM)

iv) Scatter plot:

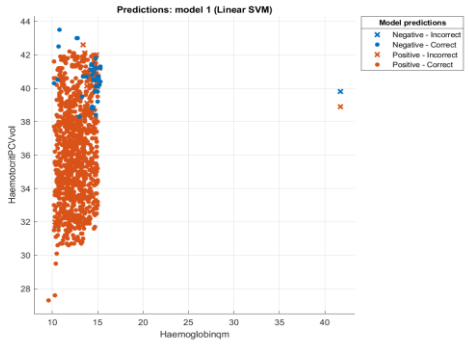

Fig 10. Scatter plot (Linear SVM)

C. Decision Tree:

To predict the responses of data decision trees are used which helps us in predicting the results by observing from the root (beginning) node down to a leaf node, where leaf node consist the responses. Classification trees give responses that are nominal such as 'true' of 'false'.

i) Confusion matrix:

Fig 11: Confusion matrix (Decision Tree)

ii) Parallel coordinator:

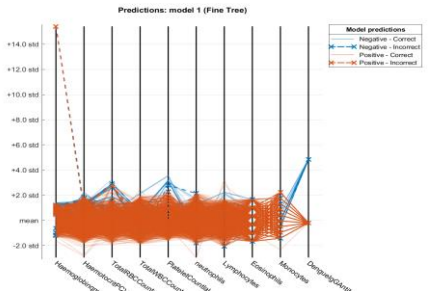

Fig 12: Parallel coordinator (Decision Tree)

iii) ROC curve:

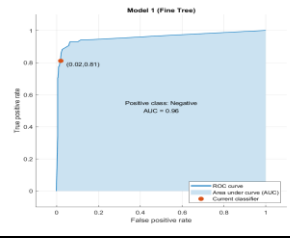

Fig 13: ROC curve (Decision Tree)

iv) Scatter plot:

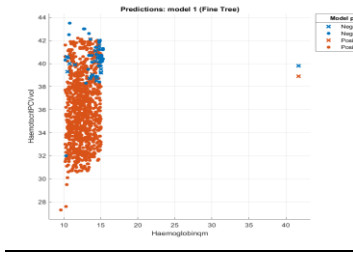

Fig 14: Scatter Plot (Decision Tree)

4. PERFORMANCE OF ALGORITHMS

Figure 15 shows clearly that the Naïve Bayes algorithm is the best with 97.2% accuracy results.

Inference:

By applying three supervised machine learning algorithms such a Decision tree, Support Vector Machine (SVM) and Naïve Baye's algorithms it is observed that the Naïve Bayes algorithm predicts the correct results with accuracy of 97.2% results.

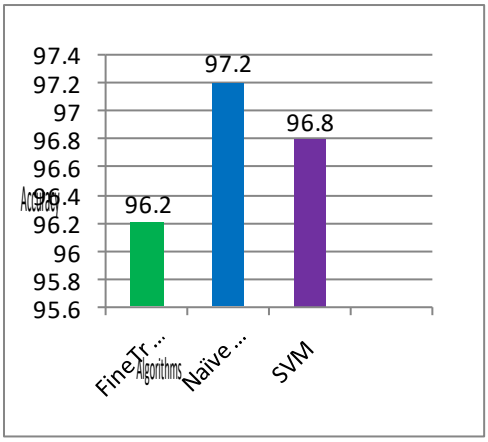

Fig 15: Performance of Supervised learning algorithms

5. CONCLUSION

The main objective of this proposed work is to detect the dengue disease at early stages using various machine learning algorithms like Decision tree, Support Vector Machine (SVM) and Naïve Baye's algorithms. The Accuracy of the various classification algorithms was compared and identified the Naive Baye's is the better performance algorithm. It is observed that the Naïve Baye's

May – June 2020
ISSN: 0193-4120 Page No. 5684-5689

algorithm predicts the correct results with accuracy of 97.2% results.

REFERENCES

[1] Dr. K. Suneetha and Mr. M. Sreekanth "Smart Home Monitoring and Automation Energy Efficient System Using IoT Devices", Emerging Research in Data Engineering Systems and Computer Communications pp 627-637, Springer Nature Singapore Pte Ltd. 2020.

[2] N.Rajathi, S.Kanagaraj, R.Brahmanambika and K.Manjubarkavi,, Kumaraguru College of Technology, Coimbatore-49. "Early Detection of Dengue Using Machine Learning Algorithms". International Journal of Pure and Applied Mathematics Volume 118 No. 18 2018, 3881-3887 ISSN: 1311-8080.

[3]Ms. K. Suneetha and Dr. M. Usha Rani "A Survey on Improving the Efficiency of PrefixSpan Sequential Pattern Mining Algorithm", International Journal of Conceptions on Computing and Information Technology (IJCCIT), ISSN: 2345 – 9808, Volume 2, issue 3, pp. 1-5, March 2014.

[4]Ms. K. Suneetha and Dr. M. Usha Rani "Performance Analysis of Web page Recommendation Algorithm Based on Weighted Sequential Patterns and Markov Model" , International Journal of Computer Science Issues (IJCSI), ISSN: 1694-0784(Print), ISSN: 1694-0814(online), Volume 10, issue 1,No. 3, pp. 250-257, Jan 2013.

[5]Ms. Suneetha. K and Dr. M. Usha Rani " A Survey on Web Recommender Systems, Publications of Problems and Application in Engineering Research-Paper(IJPaper)", ISSN: 2230-8547, e-ISSN:2230-8555,Volume 4, Special issue 01, pp. 46-51, 2013.

[6]T.Sajana, M.Navya, YVSSV.Gayathri2, N.Reshma "Classification of Dengue using Machine Learning Techniques" International Journal of Engineering & Technology.

[7]"DENGUE Disease Prediction using WEKA Data Mining Tool" Kashish Ara Shakil, Shadma Anis And Mansaf Alam Department of Computer Science, Jamia Millia Islamia New Delhi, India shakilkashish.

.