March-April 2020
ISSN: 0193-4120 Page No. 7850 - 7853

Analysis of Dengue Fever Identification for the Kanchipuram District Using Machine Learning

Dr. C. Dhaya [1], P. Sakthi Sridevi [2]

[1] Professor of CSE department, Adhiparasakthi Engineering College
[2] II year ME (CSE),APEC

Article Info
Volume 83
Page Number: 7850 - 7853
Publication Issue:
March - April 2020

Article History
***Article Received**: 24 July 2019*
***Revised**: 12 September 2019*
***Accepted**: 15 February 2020*
***Publication**: 09 April 20209*

Abstract

Dengue is named as a break bone fever and also a disease which is found in several developed as well as developing countries like India. The dengue prediction is present with the high risk in the regions of Asia and the America. This disease is caused by mosquito called Aedes aegypti. Dengue is common and it is become a global problem in all over the countries. Here the various machine learning techniques like K-Nearest Neighbor, Support Vector Machines and Decision Tree algorithms are compared for the prediction of dengue disease. MATLAB is used here to compare and evaluate the results of the dengue fever prediction.

***Keywords*:** *Support vector machine, K-Nearest Neighbor, Decision tree.*

I. INTRODUCTION

Dengue has become a common and global problem in all the countries. In this work, to compare various machine learning algorithms i.e. K-Nearest Neighbor, Support Vector Machines and Decision Tree are used to find the accuracy of these algorithms for the prediction of dengue disease. Data mining is the technique and compare the results. The main objective of this work is dengue disease prediction using various machine learning algorithms. First to identify the symptoms of dengue in patients and prediction begins based on this identification. The datasets are used for classification and to predict the accuracy.

II.PROPOSED SYSTEM

This work presents the three different machine learning algorithms i.e. SVM, KNN and DT for prediction or classification of dengue disease based on medical data attributes. Dengue disease dataset is collect from the Melmaruvathur Hospital.

The dataset has some NaN values means it need to NaN value replace by numerical values. This process is done in preprocessing step.

After that, our dataset is split into training and testing data for validation.

Finally training data are trained by different algorithms and testing data are classified based on trained model.

ADAVANTAGES

The prediction rate of accuracy is high.

The computational time also less.

III.IMPLEMENTATION

1. Dataset Preparation

The dengue disease dataset is obtained from the Hospitals. I use 10 medical data attributes such as WBC, Hemoglobin, Platelets, Abdominal Pain, Vomiting, Diarrhoea, Severe headache, Dengue, Metallic Taste, Joint/Muscle Pain for dengue disease prediction.

2. Data Preprocessing

The dataset contains original attributes and NaN values. In programming, we cannot the process the NaN values so these values are transform into

Published by: The Mattingley Publishing Co., Inc.

March-April 2020
ISSN: 0193-4120 Page No. 7850 - 7853

another value i.e. numerical value. NaN values are replaced by the mean value of columns.

3. Data Splitting

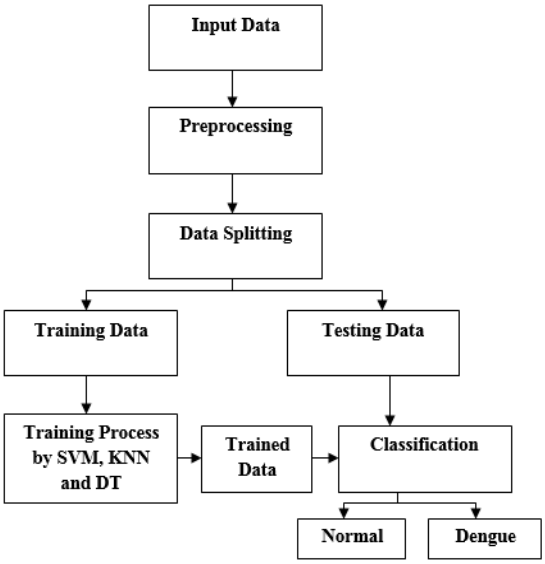
The splitting step is used for creating the training and testing data to analyzing process. In that, our whole dataset is divide into training and testing data; use 80% of data for training and 20% of data for testing.

4. Classification

In classification, training and testing data are valuated based on machine learning models. First, training data was trained by using three different machine learning models such as Support Vector Machine, K-Nearest Neighbors and Decision Tree. After that testing data are validated based on trained data with high classification accuracy rate.

IV.SYSTEM ARCHITECTURE

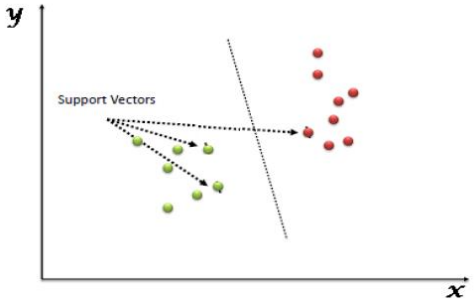
The system must provide all the essential data processing and it may also do some of those tasks identified during the work analysis.



V.ALGORITHMS

A. Support Vector Machine

One of the most popular supervised machine learning models is support vector machine which is used for classification and prediction. SVM finding the hyper-plane in the feature space that making variation between the labels or classes for classification.



Support Vector Machine

B. K-Nearest Neighbor

Hodges et al. in 1951 presented a pattern classification based on nonparametric model which is called as K-Nearest Neighbor rule. KNN is the one of the basic and simple but very intelligent classification algorithm. This algorithm did not create any assumptions for data and normally KNN used for classification process when no knowledge about the data distribution. The working of KNN is finding the k nearest data for the test data in the training set of data. K nearest data finding in training set is based on Euclidean Distance.

C. Decision Tree

Decision Tree is one of the supervised learning algorithms. Mostly classification problems are solved by using decision tree. It easily performs with continuous and categorical attributes. Based on significant predictors, the population is dividing into two or more similar set in DT. The first step of DT is calculating of entropy for each and every attribute. Next, based on the variables/ predictors the dataset is splited with high information gain or less entropy. Above two steps are followed to remaining attributes.

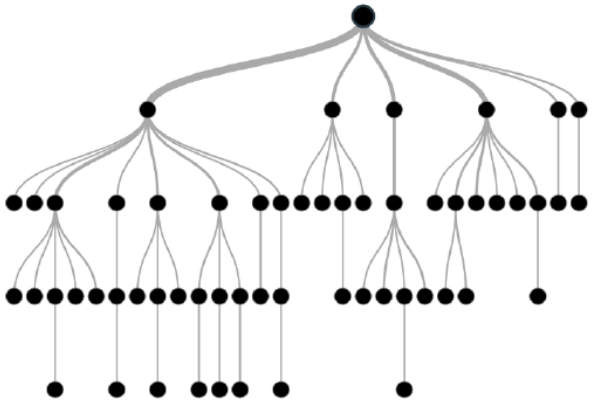
$$Entropy\ (E) = \sum_{k=1}^{l} - q_k q_k \qquad (1)$$

Published by: The Mattingley Publishing Co., Inc.

March-April 2020
ISSN: 0193-4120 Page No. 7850 - 7853

where l is refers to response variable modules count, q_k is the ratio of the count of the kth class procedures to a whole count of models.

Gain (E,G)=Entropy (E)- ∑_(v1ϵ Values(G)) |E_v1 |/E Entropy (E_v1) (2)



VI.SNAPSHOTS

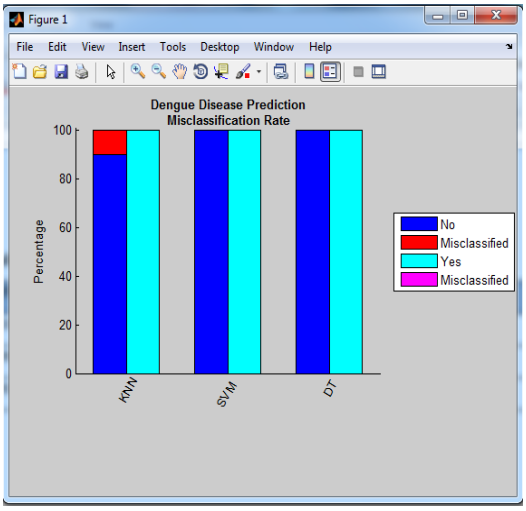
In this part, the classified result from various prediction models have been shown. Different parameters are used to make comparison with different models. The parameters such as Accuracy, Precision and Recall.

Table 1: Confusion Matrix

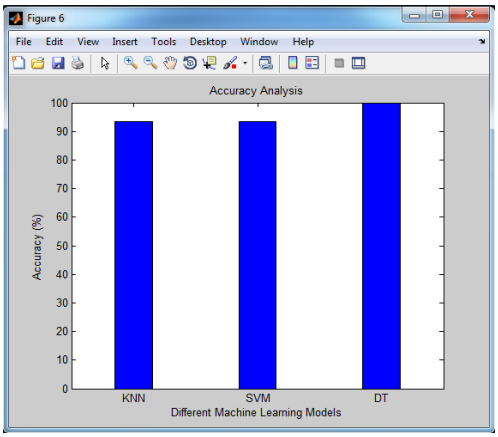
Actual/Predicted	Positive	Negative
Positive	TP	FN
Negative	FP	TN

Table 2: Quantitative Evaluation with different models

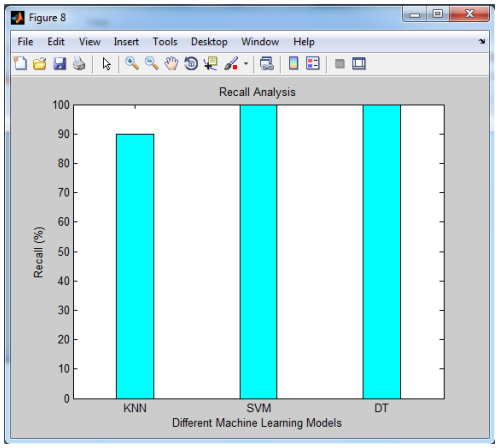
Method	KNN	SVM	DT
Accuracy	0.933	0.933	1
Precision	1	1	1
Recall	0.9	1	1



i. Misclassification Rate



ii.Accuracy Analysis with different models



iii.Recall Analysis with different models

VII.CONCLUSION AND FUTURE WORK

In this work the different machine learning algorithms i.e. Support Vector Machine, K-Nearest Neighbor and Decision Tree are used for prediction

Published by: The Mattingley Publishing Co., Inc.

of dengue disease. Prediction of dengue disease followed the step as data preparation in that data was collected from hospitals, data preprocessing for replacing of NaN values, data splitting in that whole dataset split into training and testing, finally different models based training data are trained and testing data are validated by the trained data in classification step. At last, above different machine learning models are compared in terms of finding quantitative evaluation metrics such as accuracy, precision and recall. Future work is towards the function of various evolutionary algorithms.

REFERENCES

[1] Wikipediahttp://www.bbc.com/news/world-asia-india-37415781
[2] Wikipediascribol.com/science/medicine/the-horrible-effects-of-denguefever/
[3] Wikipediahttps://www.ncbi.nlm.nih.gov/pubmed/21292885
[4] Thitiprayoonwongse DA, Suriyaphol PR, Soonthornphisaj NU," Data mining of Dengue Infection using Decision Tree", Entropy, 2012.
[5] Nandini V, Sriranjitha R , Yazhini TP, "Dengue detection and prediction system using data mining with frequency analysis",Computer Science & Information Technology (CS & IT), 2016.
[6] Padet Siriyasatien, Atchara Phumee, Phatsavee Ongruk, Katechan Jampac haisri,Kraisak Kesorn, "Analysis of significant factors for dengue fever incidence prediction", BMC Bioinformatics ,2016.
[7] Ta-Chien Chan, Tsuey-Hwa Hu, Jing-Shiang Hwang,"Daily forecast of dengue fever incidents for urban villages in a city", International Journal of Health Geographics,2015.
[8] N K Kameswara Rao, Dr. G P Saradhi Varma,Dr.M.Nagabhushana Rao, "Classification Rules Using Decision Tree for Dengue Disease", International Journal of Research Computer and Communication Technology, vol 3, Issue3,2014.
[9] Wikipedia-www.webmd.com/a-to-z-guides/dengue-fever-reference.
[10] Anshul Goyal, Rajni Mehta, "Performance Comparison of Naïve Bayes and J48 Classification Algorithms", IJAER, Vol. 7, No. 11, 2012.
[11] N. Peter, "Enhancing random forest implementation in WEKA", in: Machine Learning Conference, 2005.
[12] Breiman, L.: Random Forests. ML Journal 45(1), 5–32, 2001.
[13] Juan Wang; Qiren Yang; Dasen Ren, "An Intrusion Detection Algorithm Based on Decision Tree Technology, & quot; Information Processing, APCIP 2009. Asia-Pacific Conference, vol.2, no.,pp.333, 335,2009.
[14] David S. K., Saeb A. T., Al Rubeaan K, "Comparative Analysis of Data Mining Tools and Classification Techniques using WEKA in Medical Bioinformatics", Computer Engineering and Intelligent Systems, 4(13):28- 38,2013.
[15] Durairaj M, Ranjani V, "Data mining applications in healthcare sector a study", Int. J. Sci. Technol. Res. IJSTR, 2(10), 2013.