



Dengue Prediction Using Classification Techniques

B. Sundari^[1], V. Anbarasan^[2]

¹ (Assistant professor, Department of Computer Science, Christ college of arts and science college, Kilachery)

²(Professor, Head department of computer application, Christ college of arts and science, Kilachery)

Abstract:

Data mining is the method of finding knowledge such as patterns, associations, changes, anomalies and significant structures, from huge amount of data stored in database, data warehouses or other information storage place. Dengue is a mosquito borne viral disease which is mainly transmitted by the species of female mosquitoes named “Aedes Aegypti”. This research work has been focused two stages namely classification and association. The main goal of research work is to predict the people who are affected by dengue depending upon categorization of symptoms. The classification is done by J48 classifier and the association is done to find which period has higher occurrence of dengue fever. This study of predicting the relationship between the dengue symptoms will surely help the biotechnologists and bio informaticians to discover antibiotic for dengue.

Keywords — Data mining, prediction, Dengue

I. INTRODUCTION

Dengue fever is a painful, debilitating mosquito-borne disease caused by any one the four dengue virus. It is transmitted by the bite of an Aedes mosquito infected with dengue virus. It can be spread directly from one person to another person. As estimated 390 million dengue infection occur worldwide each year. Most cases occur in tropical and sub-tropical areas of the world. Dengue fever is classified into two categories namely type 1 and 2. The first type is a classical dengue which is referred as dengue fever and the other type is known as dengue hemorrhagic fever. The disease might develop further into dengue hemorrhagic fever (DHF) which results in decrease in the number of blood platelets.

Symptoms of Dengue: Symptoms which usually begin four to six days after infection and last for up to 10 days, may include sudden high fever, severe headaches, Pain behind the eyes, severe joint pain, Muscle pain, Fatigue, Nausea,

Vomiting, Skin rash which appears two to five days after the onset of fever.

II. REVIEW OF LITERATURE

1. [P.Manivannan, Dr.P.Isakki] Provided k-means clustering algorithm for prediction of dengue fever. K-means clustering: K-means is a clustering algorithm used to classify or group the objects based on attributes and that are divided into k-number of group. In this paper, k-means clustering algorithm used to partition the data sets into k-clusters. This work is done on predicting the dengue based on categorization of age group using k-means clustering. This technique is one of the best technique for predicting the dengue patients with serotypes.

2. [Panuwat mekha, khukrit osathanunul et al] Genome sequences: In molecular biology and genetics, the term pertains to the complete set of genetic material in an organism. The genome of eukaryotes is contained in a single, haploid set of chromosomes. Dengue genes are extracted from the

virus pathogen resource. They are divided into four distinct types. In this work we compared the different methods for classification and finally concluded that using codon usage is the best one for dividing the viruses. It can be support of gene expression and molecular classification. It achieves 96.22% of prediction accuracy by using a neural network.

3. [Abdul mahatir najar, Mohammad isa irawan et al]

Dengue hemorrhagic fever: Dengue hemorrhagic fever (DHF) is a disease that is transmitted through mosquito's bites, especially *Aedes aegypti* or *Aedes albopictus*. It is commonly founded in tropical and sub-tropical areas. In this work focused on prediction related to the risk level of DHF outbreaks. Because each region needs to get treatment according to its risk levels. The result shows that extreme learning machine method can predict the risk level of dengue hemorrhagic fever by using 50 hidden neurons.

4. [priyo s sasongko, helmie a wibawa et al]

In this paper focused on to find the best back propagation algorithm to find the detection of dengue and to determine the possibility of DHF. Problems that can be arises from early detection of dengue disease is the data used is YES or No. To solve this problem by using optimization on multi-level perceptron. The result of this work is Levenberg Marquart algorithm shows the best performance in detection of dengue disease early.

5. [Buchado omkar, Dalsania preet et al]

In this paper they are stratifying dengue into dengue fever (DF), Dengue hemorrhagic fever and healthy patients. Here we are using spider monkey optimization and also used probabilistic neural network for classification. They were used greedy feed forward selection algorithm for selection of the specific gene. The primary task is to find whether the person is suffering from dengue or not. This proposed model gives 90.91% accuracy.

6. [Norhayati binti Mohd Zainee et al] In this research the proposed work is based on vital signs

containing of blood pressure (BP), heart rate and body temperature are the factors of physical test for dengue fever. The reference being used in the implementation was taken from the Hospital Canselor Tuanku Muhriz (HCTM) in Malaysia. It shows that the trained linear discriminant model using median value gives the highest TPR with the lowest FNR but provide a low accuracy during the verification. It provides low accuracy during the verification.

7. [Nirbhay Mathur et al] This study was used to visualize the dengue incidences on weekly basis in Selangor, Malaysia. This research work is done on by using the geographical information system based on k- means clustering and expectation maximization algorithm. This study finally found that some areas of Selangor are on the high risk of dengue incidences.

8. [Marimuthu, T., et al] exposed new bio-computational model for mining the dengue gene sequences. They proposed a bio-computational model called sequence miner to interpret the relationship among the dengue viruses. The accuracy of the proposal model is 96.74%. The relationship between dengue serotypes are predicted via the proposed tool. It helps to the biotechnologies and drug designers for discovering an effective vaccine for dengue.

9. [M Mufli Muzakki et al] In this paper they proposed prediction of DHF in Bandung Regency using K-Means Clustering as preprocessing method and Support Vector Machine (SVM) algorithm. The data being used in the research from Meteorological, Climatological, and Geophysical Agency in Bandung Regency. Weather data can be used for predicting the DHF disease because there is a relation between weather attribute and DHF disease. K Means Clustering label method shows accuracy more than 86%. The implementation of this research work can be useful for Health Department of Bandung Regency and to increase awareness about DHF disease.

10. [Iwan Inrawan Wiratmadja, Siti Yaumi Salamah et al] In this research work focused on predicting the hospital length of stay from the time of admission, which is used for hospital management. With the use of decision tree an accuracy of 71.57% is achieved. The prototype of dengue patient length of stay prediction system was developed using the resulting decision tree classification rules.

11. [R. Sanjudevi and D. Savitha] proposed method is compared with the decision tree. It is implemented a feature model construction and comparative analysis for improving prediction accuracy of dengue disease in three phases. In first phase, dengue disease data sets are collected from UCI repository. In second phase, feature selection is done by forward and backward step wise regression methods. Third phase, SVM and decision tree classification algorithms are applied on the data set. In the fourth phase, the accuracy determination be calculated using Sensitivity (SE), Specificity (SP), and Area under the curve (AUC). Finally, the evaluation is done based on accuracy values. Thus, outputs shows that proposed classification algorithm SVM gives better accuracy than decision tree algorithm with the help of feature selection.

III. TOOLS AND TECHNIQUES

WEKA (Waikato Environment for Knowledge Analysis) is a popular suite of machine Learning Software written in Java, developed at the University of Waikato, New Zealand. WEKA is free software available under the GNU General Public License. WEKA, formally called Waikato Environment for Knowledge Learning, is a computer program that was developed at the University of Waikato in New Zealand for the purpose of identifying information from raw data gathered agricultural domains. WEKA supports many different standard data mining tasks such as data pre-processing, classification, clustering, regression, visualization and feature selection. The basic premise of the application is to utilize a computer application that can be trained to perform machine learning capabilities and derive useful

information in the form of trends and patterns. WEKA is open source application that is freely available under the GNU general public license agreement. Originally written in C the WEKA application has been completely rewritten in Java is compatible with almost every computing platform. It is user friendly with a graphical interface that allows for quick set up and operation.

WEKA operates on the prediction that the user data is available as a flat file or relation, this means that each data object is described by a fixed number of attributes that usually are of a specific type, normal alpha numeric or numeric values. The WEKA application allows novice users a tool to identify hidden information from database and file systems with simple to use options and visual interfaces.

IV. PROPOSED METHODOLOGY

Dengue fever is a virus infection which is transmitted to humans by mosquitoes. It is classified into two types namely type 1 and type 2. The first type is a classical dengue which is referred as dengue fever and the other type is known as dengue hemorrhagic fever. This study is about to predict the occurrence of dengue fever and to predict which one has higher occurrence of fever. To predict the dengue a methodology is proposed with three phases. In first phase is deals with classification and the second phase is deals with association.

Classification is a data mining function that assigns items in a collection to target categories or classes.

Phase1: Classification

Classification technique is used to retrieve important and relevant information about data, and metadata. This data mining method helps to classify data in different classes. J48 classifier is one of the well-known and efficient classifiers when compare with other classifiers. In this study J48 classifier is used to classify the symptoms of dengue to predict the occurrence of dengue fever.

Association rule mining is a procedure which is meant to find frequent patterns, correlations, associations, or causal structures from data sets found in various kinds of databases such as

relational databases, transactional databases, and other forms of data repositories. Association rule mining is suitable for non-numeric, categorical data and requires just a little bit more than simple counting.

Phase2: Association

Association technique helps to find the association between two or more Items. It discovers a hidden pattern in the data set. Apriori algorithm is one of the efficient association techniques. Given a set of records each of which contain some number of items from a given collection, produce dependency rules which will predict the occurrence of an item based on patterns found in the data. In this proposed methodology Apriori algorithm is used to find which period has higher occurrence of dengue fever.

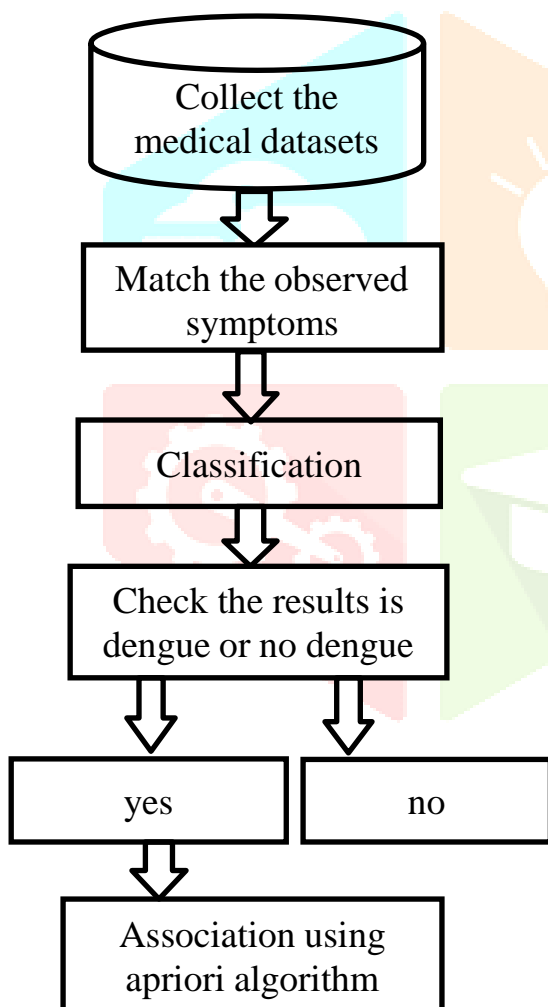


Figure 4.1 Framework of proposed methodology

V.RESULTS AND ANALYSIS

In this research, we calculate various responses from Patients have been collected from hospitals and it gets result of dengue is Dengue or NO Dengue. The collected response gets mined using classification and analysed that is presented below by considering the symptoms: Period of Fever, Fever Temperature, Rashes or Red spots, Pain behind eyes, Headache, Muscle and Joint Pain, Nausea or Vomiting, Low Heart Rate and Fatigue. Results are implemented through observed symptoms and analysis of the dengue dataset. In the analysis, they get the dengue result is varies on nausea and period of fever.

Data collection

The input dengue data have been collected from urban Ho Chi Minh City, Vietnam. Initially the data size is 1009 records and 11 attributes. The data has following set of attributes.

Attributes	Possible values
Period of Fever	Three Days, Four Days, Five Days
Fever Temperature	100°C, 102°C, 104°C
Rashes or Red spots	Yes or No
Pain behind Eyes	Yes or No
Headache	Yes or No
Muscle and Joint Pain	Yes or No
Nausea or Vomiting	Yes or No
Low Heart Rate	Yes or No
Fatigue	Yes or No
Result	Dengue or No Dengue

Table 5.1: Symptoms of Dengue Virus

Comparative analysis of LMT, J48, NBTree, and Hoeffding

The experiments have been conducted and different decision tree algorithms are applied on dataset in WEKA framework. The results of Decision tree classifiers LMT, J48, NBTree, and Hoeffding are compared. In our experiment,

parameters such as error rate, execution time, and accuracy are evaluated and compared.

Algorithm	Correctly classified instances	Instances Incorrectly classified
J48	91.66%	8.33%
Naïve Bayes	89.66%	10.33%
Hoeffding	79.16%	20.83%
LMT	51.48%	48.51%

Table 5.2: Performance of Classifiers

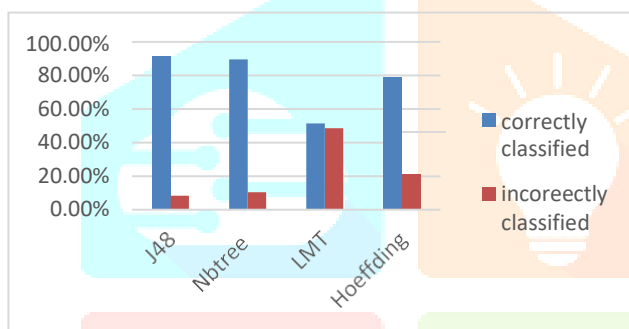


Figure 5.1: Comparison on Accuracy of Classifier

Performance of classifiers

Algorithm	Time Taken to build model (in seconds)
J48	0.01
Naïve Bayes	0.03
Hoeffding	1.02
LMT	0.04

Table 5.3: Comparison on Accuracy of Classifier

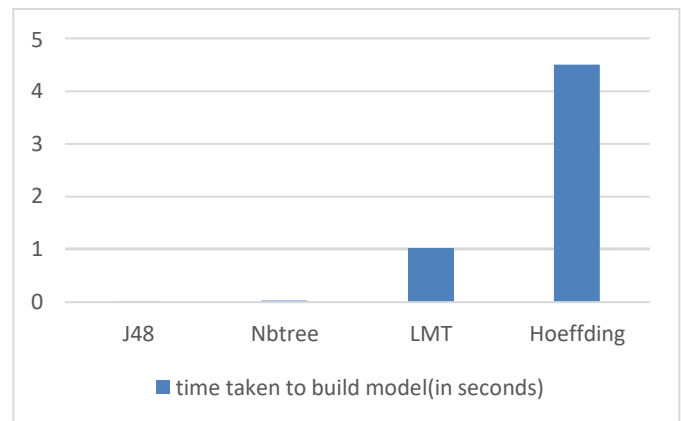


Figure 5.2: Performance of different decision tree classifiers

VI.CONCLUSION

The proposed research has used data mining technique and WEKA tool to predict the results of dengue fever. It has evaluated, the number of attributes using classification technique Naïve Bayes, Hoeffding, LMT and J48. From the statistics which is given by the various techniques, the accuracy has been acquired as follows. Accuracy of Naïve Bayes is 89.66%, Hoeffding is 79.16%, LMT is 51.48%, and J48 is 91.66%.

From analysis and the accuracy acquired, it has been proved that, the Naïve Bayes tree and J48 is a simple classifier technique to make a decision tree and that results with high accuracy. As an next level of analysis based on the time element, if we consider Naïve Bayes tree takes more than 50 times of execution time as the time taken by J48.

By considering these two factors like accuracy and time in parallel, even though Naïve Bayes tree gives more accuracy than J48, that is not that much wage difference, where as if we consider time lime element Naïve Bayes tree and J48, the difference was very huge and not Tolerable. So, from this point the analysis results that J48 is a simple classifier technique to make decision tree for dengue prediction. It is concluded that the result of dengue fever status is Dengue or No Dengue and from the observed symptoms it is analysed that dengue fever mostly varies on fever period fever followed by nausea and vomiting also find that July to December is the higher occurrence period of dengue fever.

VII. REFERENCES

- 1.P. Manivannan, Dr.P.Isakki, Dengue Fever Prediction Using K-Means Clustering Algorithm, IEEE INTERNATIONAL CONFERENCE ON INTELLIGENT TECHNIQUES IN CONTROL, OPTIMIZATION AND SIGNAL PROCESSING, 978-1-5090-4778-9/17/©2017
- 2.Norhayati binti Mohd Zainee , Kalaivani Chellappan , A Preliminary Dengue Fever Prediction Model Based On Vital Signs And Blood Profile , IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES) 2016.
- 3.NirbhayMathur,VijanthS.Asirvadam,Visualization of Dengue Incidences using Expectation Maximization (EM) Algorithm , IEEE 978-1-5090-0845-2/16/© 2016.
- 4.T.Marimuthu,V. Balamurugan, A NOVEL BIO-COMPUTATIONAL MODEL FOR MINING THE DENGUE GENE SEQUENCES , International Journal of Computer Engineering & Technology (IJCET) Volume 6, Issue 10, Oct 2015, pp. 17-33, Article ID: IJCET_06_10_003
5. Ramandeep Kaur , Gaurav gupta ,demographic analysis of dengue fever using data mining, International journal of advanced research in computer science, Volume 8, No. 7, July – August 2017.
6. M Mufli Muzakki, Fhira Nhita , ‘The Spreading Prediction of Dengue Hemorrhagic Fever (DHF) In Bandung Regency Using K-Means Clustering and Support Vector Machine Algorithm’6th International Conference on Information and Communication Technology (ICoICT) 2018
- 7.A.Shameem Fathima, D.Manimeglai, ‘Predictive Analysis for the Arbovirus-Dengue using SVM Classification’,International Journal of Engineering and Technology Volume 2 No. 3, March, 2012 ISSN: 2049-3444 © 2012 – IJET Publications UK. All rights reserved. 521
8. Wahjoe Tjatur Sesulihatien , Yasushi Kiyoki , ‘Ecological Context-Dependent Analysis and Prediction using MMM : A Case of Dengue Fever Disease’ , International Electronics Symposium (IES) 2015.
9. Shermon S. Mathulamuthu, Vijanth S. Asirvadaml, “Predicting Dengue Incidences Using Cluster Based Regression on Climate Data”, 6th IEEE International Conference on Control System, Computing and Engineering, 25-27 November 2016.
10. Panuwat Mekha, Khukrit Osathanunkul,” Gene Classification of Dengue Virus Type Based on Codon Usage”, IEEE-978-1-5090-4420 2016.
11. DiniRahmawati and Yo-Ping Huang,” Using C-support Vector Classification to Forecast Dengue Fever Epidemics in Taiwan”, International Conference on System Science and Engineering (ICSSE), July 7-9, 2016.
12. Shermon S. Mathulamuthu, Vijanth S. Asirvadam, Sarat C.Dass , Balvinder S. Gill,” Predicting Dengue Cases by Aggregation of Climate Variable Using Manifold Learning” IEEE International Conference on Signal and Image Processing Applications (IEEE ICSIPA), 2017.
13. R. Sanjudevi, D. Savitha,” DENGUE FEVER PREDICTION USING CLASSIFICATION TECHNIQUES”, International Research Journal of Engineering and Technology (IRJET),2019.
- 14.S.Gupta, Kumar, A. Mishra (2011), “Data Mining Classification Techniques Applied for Dengue Disease Diagnosis and Prognosis”. IEEE Journal of Computer Science and Engineering. Vol.23, pages 1191-1193.
15. Loshini T1., Vijanth S. Asirvadam , Sarat C. Dass Balvinder S. Gill , ,” Predicting Localized Dengue Incidences using Ensemble System Identification”, International Conference on Computer, Control, Informatics and Its Applications 2015.