International Journal of Computer Trends and Technology
ISSN: 2231-2803 / https://doi.org/10.14445/2231280453/IJCTT-V67I3P114
Volume 67 Issue 3, 72-74, March 2019

Original Article

Predicting Dengue Using Fuzzy Association Rule Mining

Anitha A[1], Freeda Jebamalar S[2]

[1]*Professor & Department of InformationTechnology & Anna University*
[2]*PG Scholar & Department of InformationTechnology & Anna University*
Francis Xavier Engineering College, India

Abstract *- Dengue is a fast emerging pandemic-prone and the most dreadful disease spread worldwide. Aedes aegypti and Aedes albopictus mosquitoes cause it. It affects tropical and subtropical countries. Mosquitoes act as transmitters, and the main victim will be humans. It consists of four distinct serotypes(DEN-1,2,3,4).It may belong to the genus flavivirus [6]. The main reason for the propagation of vectors and pathogens is changing in the globalization of travel and trade. Temperature and precipitation will play an important factor in dengue. In 2007 Intergovernmental panel on climate change prescribed and believed that the range of 1.5 and 3.5 people worldwide will face the risk of dengue fever in 2080[7]. Classification will compare and analyze the accuracy level. In this research, fuzzy association rule mining is used to predict the dengue in the earlier phase with the formation of association rule along with the prediction. It may help in quick diagnosis and save human life.*

Keywords *- Climate, Dengue, Pathogen, Prediction, Mining*

I. INTRODUCTION

Data mining is mining the data or digging deep into data efficiently, which may be huge and difficult to analyze data patterns. The mining means sorting the large dataset to identify the patterns and relationships in it. They may help in problem-solving in the data analysis process. There are four stages in data mining. They are data sources, data exploration, modeling, and deploying models. Its technique is used in many fields such as mathematics, cybernetics, genetics, education, sports, medicine, marketing, etc.; web mining is a type of data mining used in customer relationship management. They may predict the customer behavior over the website[8]. They consist of parameters like association rule mining, sequence or path analysis, classification, clustering, and data forecasting.

Fuzzy logic is used in uncertainty in engineering. It has four main parts. They are the Fuzzification module, Knowledge base, Inference engine, and Defuzzification module. In the Fuzzification, inputs(crisp numbers) into fuzzy sets. It may split the input signal into five parts. The knowledge base store the If-then rules. The Inference engine will stimulate the human reasoning process with the help of fuzzy inference. The fuzzy set is converted into the crisp value in the defuzzification module. The membership function will quantify and represent the fuzzy set. The triangular membership function is the most common.

II. LITERATURE SURVEY

In this paper, the author proposed the system by collecting 1055 instances and 18 attributes for training and 29 instances for the testing phase by extracting 20 rules from doctors about DF and DHF for a testing purpose[1]. In the training and testing dataset, they applied the multilayer perception, a part of a neural network that achieves high accuracy. They implemented the system by proposing fuzzy logic in the detection of dengue. It involves the following modules. The fuzzifier will accept the input from the patient's symptoms. The rules phase will store the experience from the doctor. The inference phase will compare the rules and the input of the patients. Defuzzifier will show the output by comparing the rules and the patients' input. The rules can be converted into CSV format. They achieved the mean absolute error of 0.0009, and the root means square deviation is 0.0011. The mean absolute error is used to detect how close to the predicate results. RMSD (root mean square deviation) is the difference between the predicted and observed values.

In this paper, the author proposed the system by collecting data by interviewing doctors on the internet[2]. They import the data after preprocessing into the Matlab 2013a.In the missing value filling phase, I visited the various hospitals in Hyderabad. They conducted interviews with the doctors and patients infected by dengue regarding the symptoms like headache, vomiting, body pain, etc. Data were collected in the form of tables, and missing fields were filled with approximate values for accurate results. In the data representation phase, data includes mixed, i.e., numerical and categorical data. For neural networks, data must be in the form of numerical from categorical data. The neural network involves

collecting the required data by initializing the weights by developing and organizing the network. Then train the dataset with known values and test the data to confirm using these as a new dataset.In the neural model,there are three function p,wp,a.where p is the first function, wp is the second one, added to the scalar product to obtain the input n. Finally, it is passed via transfer function f, which will produce the output a. The dataset was divided into 3 samples. They are training samples, validation samples, and testing samples. Among them, 143 samples were involved in the training phase, and 30 samples were used for validation and testing. In the training phase, the network was adjusted according to the errors.

The validation phase was for network generalization, which might stop the training. Testing was an independent phase in the neural network. The percentage error must be 0, then only there is no misclassification.ROC(Receiver Operating Characteristic) will ensure the quality of classifiers. There are TPR(True Positive Rate) for each threshold, greater or equal to the threshold divided by one target.FPR(False Positive Rate) is less than the threshold divided by zero targets. The Confusion Matrix is also known as the error matrix, which denotes the algorithm's performance. The column denotes the predicted instance, and the row denotes real class instances or vice versa.

In this paper, the author considered the four classifications of the dataset for the prediction of dengue[3]. The classification depends on the clinical test, climate factor or time series, genes, and protein. Classifiers like Tree-based model, neural network model, and Ensemble classifier. The decision tree is used for the approximation of discrete esteemed targets. Objective variables will accept the limited set of qualities. The regression decision tree can take the continuous values, which means considering the predicate as real. The Neural Network model can observe the knowledge via an intermediate unit. The Evolutionary based classifiers are a subset of the hereditary populace based on the meta-heuristic streaming algorithm, which is an example of reproduction, mutation, recombination, and selection. Genetic algorithms, Neuro-evolution, and gene expression are powerful techniques that can also be applied. The Ensemble classifier was used for the multimodel for the accuracy rate. It consolidates the models M1, M2, and M3,…..MK into M*.In the training phase, Applied methods are involved for prediction. In the testing phase, the predicate models of the training phase were given as input to produce predicate values, then accompanied by these values by the Combining rule. Finally, the output will be produced with an interesting measure.

In this paper, the author improved association inference rule mining with the fuzzy c means algorithm were subjected to identify the frequent patterns. The dataset consists of 1000 patients having different symptoms. The Fuzzy c means clustering denotes the multiple clusters depending on the degree of membership. In this algorithm, o Object function calculates in each iteration.

$$O_m = \sum_{p=1}^{N} \sum_{q=1}^{C} U_{pq}^{m} \, || X_p - C_q ||^2 \qquad (1)$$

where,

m – a real number greater than 1

N - Number of data points

C – Number of required clusters

XP-data point

Cq– center vector for cluster q

XP-Cq- a similar data point

The first phase is data preprocessing, and it consists of two steps[4]. Initially, assign the index for symptoms uniquely. Next, mapping was done between the index table and the original dataset. Finally, the new table is used for smooth processing. Association rule mining was used to identify the frequent patterns and store the pattern. In the fuzzy association rule mining, the additional inference was deployed because more frequent patterns lead to difficulty in conclusions. Fuzzy c means were used to match the frequent symptoms with factual knowledge. They are clustered as low, mid, or high, depending upon the degree of membership. The highly matched pattern will be put into the HIGH class, and the unrelated class will be put into the MID class.

The author described the fuzzy logic approach in job scheduling[5]. The proposed job scheduling system approach. Fuzzy sets are in the form of if-then statements. They become the rule base, which will be the decision-making in the control actions. The fuzzy controller design includes four steps, initially identifying the linguistic input and output variables and definition of fuzzy sets. The inference evaluates the rules and merges the output rules. The defuzzification will convert the fuzzy rules into crisp outputs. They implemented the above technique in the single tailor machine job shop problem. It consists of the due date, customer priority, and processing time. They help in the decision-making of rules. The customer priority consists of bad, low, medium, high, and very high. The processing time includes the variables such as short, medium, and long. The membership function is closed and distance as C, D respectively. The process proceeds by calculating the fuzzy close function and fuzzy distance function, then selecting the job and identifying the job's close and distance function. Afterward, select the sequence priority using fuzzy customer priority concerning the close and distance function. Repeat the following step until the allocation of the job to the machine completely.

III.PROPOSED SYSTEM

The proposed system includes three modules: data collection, Fuzzy Inference System, and Prediction.

A. Data Collection

The dataset was collected from Tamilnadu regarding dengue-like illness. The symptoms include fever, chills, throat pain, cough, loose tools, etc.,.and are stored in an excel file.

B. Fuzzy Inference System

It consists of four parts. They are rule base, fuzzification, inference engine, and defuzzification. In the rule base, the decision will make based upon the If-Then conditions. At the same time, in the fuzzification, the crisp input, which is temperature, pressure, etc., the Inference engine will trigger the condition by matching the input with rules. Defuzzification converts the rule into crisp input. There are three types of fuzzifier includes singleton, Gaussian, trapezoidal or triangular fuzzifier.

C. Prediction

Using the fuzzy association rule mining, predicting the disease through the generation of the rules using a rule generator and predicting the possibility of the disease.

1. Dengue Diagnosis Tool GUI:

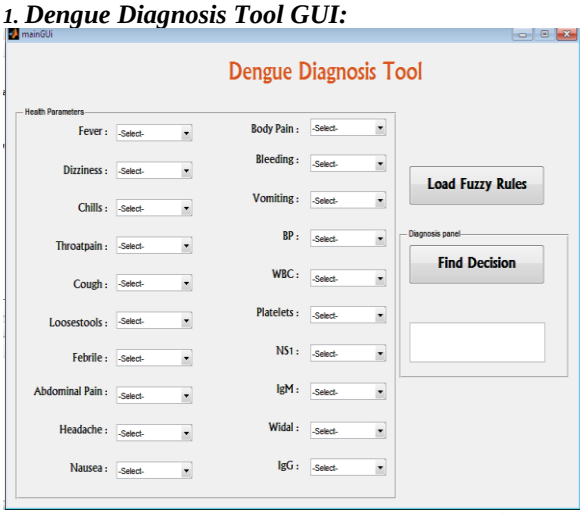

Fig. 1 Dengue diagnosis tool graphical user interface

2. Disease Prediction Rule Generator

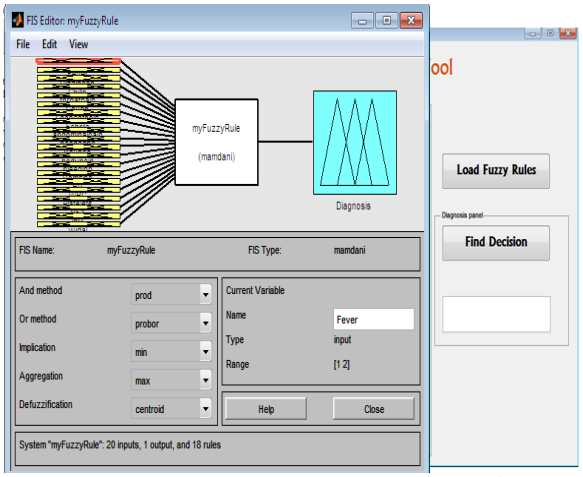

Fig. 2 Disease prediction rule generator using an inference engine

3. Diagnosis of Disease:

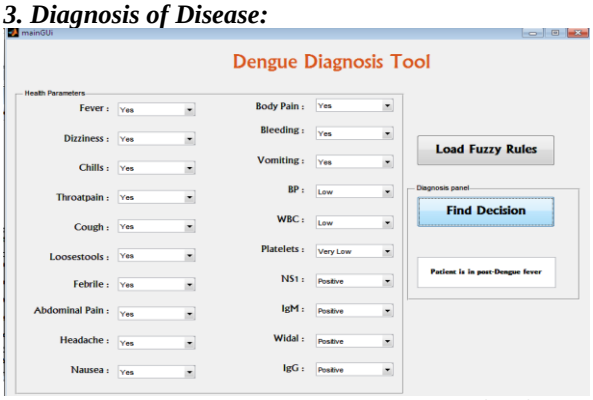

Fig. 3 Dengue diagnosis tool graphical user interface for prediction of disease.

IV. CONCLUSION

From the proposed system, it has been concluded that the fuzzy association rule mining technique with the rule generation will provide the possibility of disease prediction was performed. In the future, it has been improved with more latest techniques.

REFERENCES

[1] N.A Husin,A Alharogi,N.Mustapha, H.Hamdan, U.A Husin "Early self-diagnosis of dengue symptoms using fuzzy and data mining approach", AIP Conference 2016 Sep, 2018.

[2] M.V.J Reddy, B.Kavitha "Expert system to predict the type of fever using data mining techniques on medical databases"IJCSE, vol.3, Issue.09, Sep. 2015

[3] N.Iqbal,M.Islam"Machine learning for dengue outbreak prediction :An Outlook"IJACS,vol .8,no. 1,Jan-Feb. 2017,pp. 93-102

[4] K.Chaturvedi, R.Patel, D.K Swami"A Fuzzy Inference approach for association rule

[5] Odii J.N,Onwuama T.U,Okpalla C.L,Ejem A,"Job scheduling system using fuzzy logic approach"IJCTT,vol.42.no.2,Dec 2016. mining"IOSR,vol.16,Issue.6,Nov-Dec.2014,pp.57-66.

[6] http://www.who.int/denguecontrol/monitoring/en/

[7] https://www.ncbi.nlm.nih.gov/pmc/srticles/pmc5583666/

[8] https://searchsqlserver.techtarget.com/efinition/data-mining