Predicting Dengue Using Fuzzy Association Rule Mining

Anitha A[#1], Freeda Jebamalar S[*2],
Professor & Department of InformationTechnology & Anna University
PG Scholar & Department of InformationTechnology & Anna University
Francis Xavier Engineering college,India

Abstract

Dengue is a fast emerging pandemic prone and most dreadful disease spread all over the world. It is caused by Aedes aegypti and aedes albopictus mosquitoes. It is affect the tropical and subtropical countries. Mosquitoes act as a transmitter and the main viticm will be the human. 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 vector and pathogen because of changes occurred in globalization of travel and trade. Temperature and precipitation will play the important factor in dengue. In 2007 Intergovernmental panel on climate change prescribed and believed that the range of 1.5 and 3.5 people all over the world will face risk of dengue fever in 2080[7]. Classification will compare and analysis 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 helps in quick diagnosis and save the human life.

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

I. INTRODUCTION

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

Fuzzy logic used in uncertainty in engineering.It has four main parts. They are Fuzzification module, Knowelege base, Inference engine, Defuzzification module. In the

Fuzzification ,inputs(crisp numbers) into fuzzy sets.It may splits the input signal into five parts.The knowledge base store the If-then rules.In the Inference engine will stimulates the human resoning process with the help of fuzzy inference.In the defuzzification module,the fuzzy set is converted into the crisp value.The membership function will quantify and represent the fuzzy set.The triangular membership function is the most common membership function.

II. LITERATURE SURVEY

In this paper author proposed the system by collecting 1055 instances and 18 attributes for training and 29 instances for testing phase by extract 20 rules from doctors from about DF and DHF as a testing purpose[1]. In the training and testing dataset they applied the multilayer perception which is a part of neural network achieves high accuracy.They implemented the system by proposing the fuzzy logic in the detection of dengue.It involves the following modules.In the fuzzifier will accept the input from the patient's symptoms.Rules phase will store the experience from the doctor. The inference phase will compare the rules and the input of the patients. Defuzzifier will shows the output by comparing the rules and the input of the patients. The rules can be converted into csv format.They achieved the mean absolute error with 0.0009 and root mean 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 value and the value observed.

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

and organize the network by initialize the weights.Then train the dataset with known values and test the data for confirmation of 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,which is added to scalar product,inorder to obtain the input n.Finally nis passed via transfer function f which will produce the output a.The dataset were divided into 3 samples.They are training samples,validation samples and testing samples.Among them 143 samples involved in the training phase and the 30 samples used for validation and testing phase.In the training phase,network were adjusted according to the errors occurred.Validation phase was for network generalization and it might stop the training .Testing was a independent phase in the neural network.Percentage error must be 0,then only there is no misclassification.ROC(Receiver Operating Characteristic) will ensure the quality of classifiers.For each threshold,there are TPR(True Positive Rate) which is greater or equal to threshold divided by one target.FPR(False Positive Rate) which is less than threshold divided by zero targets.Confusion Matrix is also known as error matrix,which denotes the performance of the algorithm.In which column denotes the predicted instance and the row denotes real class instances or vice versa.

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

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

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

where,
m – 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- similarity data point

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

In this paper the author described using the fuzzy logic approach in the 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. In the fuzzy controller design include four steps, initially identified the linguistic input and output variables and definition of fuzzy sets. The inference evaluates the rules and merge the output rules. In the defuzzification will convert the fuzzy rules into crisp outputs. They implemented the above technique in the single tailor machine job shop problem. It consist of due date, customer priority, processing time. They are helps in decision making of rules. The customer priority consist of bad, low, medium, high, very high. The processing time include the variables such as short, medium and long. The membership function are closed and distance as C ,D respectively .The process proceed as by calculating the fuzzy close function, fuzzy distance function, then select the job, for which identify the job's close and distance function. Afterwards selecting the sequence priority using fuzzy customer priority with respect to the close and distance function. Repeat the following step until the allocation of job to machine completely.

III.PROPOSED SYSTEM

The proposed system includes three modules namely about data collection,Fuzzy Inference system,Prediction.

A. Data Collection

The dataset collected from tamilnadu regarding dengue like illness.The symptoms includes fever,chills,throatpain,cough,loosetools etc.,.and stored in excel file.

B. Fuzzy Inference System

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

C. Prediction

Using the fuzzy association rule mining,predicting the disease through the generation of the rules using rule generator and predict the possibility of the disease.
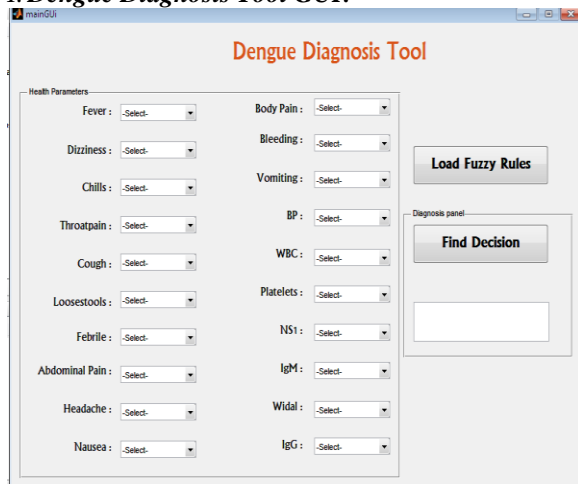
1. Dengue Diagnosis Tool GUI:

Fig. 1 Dengue diagnosis tool graphical user interface

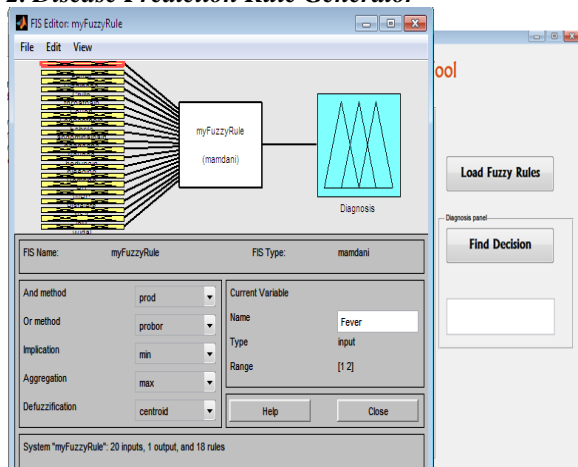### 2. Disease Prediction Rule Generator

Fig. 2 Disease prediction rule generator using inference engine
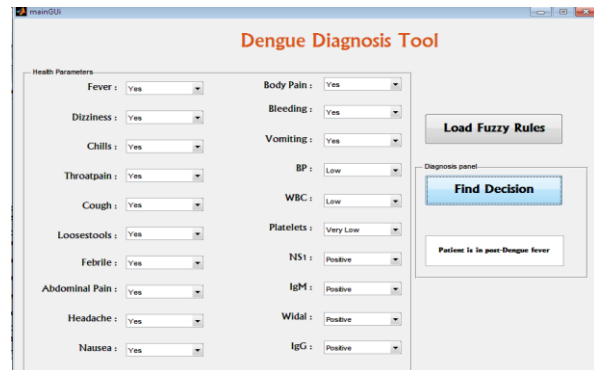
3. Diagnosis of Disease:

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

IV.CONCLUSIONS

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