Survey Of Data Mining Techniques Used In Healthcare Domain And Diagnosis Of Dengue Fever

Pradeep Raj. D[1], Mythili. R[2]

[1](Department Of Computer Application, Indo-American College/ Thiruvalluvar University, India)
[2](Department Of Computer Application, Indo-American College/ Thiruvalluvar University, India)

Abstract: *Healthcare organisations and industries produces huge amount of data everyday. These vast information can be extracted and analysed to obtain patterns which can be used to forcast or predict the future events. Many healthcare organizations have already started collecting healthcare records to systematically use those data to identify patterns and improve health management of a particular population thereby providing improved care and reduced manual work and cost. This paper focuses on the survey of different dengue fever classification and study of a particular population where people are mostly diagnosed with dengue infection. It also discusses critical issues and challenges associated with data mining and healthcare in general. The research found a growing number of data mining applications, including analysis of health care centres for better health policy-making, detection of disease outbreaks. We have collected data from government hospitals. We have applied these data in our fuzzy logic and generated decision tree and compared the performance of other techniques.*

Keywords: *Data Mining,, Decision Tree, Dengue, Fuzzy Logic, Healthcare*

I.   Introduction

Dengue is one of the most widespread mosquito-borne viral diseases. More than 100 million cases of dengue fever are reported annually. Dengue fever, also known as break bone fever, is caused by four different viruses and spread by Aedes mosquitoes. Severe symptoms include Dengue Shock Syndrome (DSS) and Dengue Hemorrhagic Fever (DHF). The Centers for Disease Control and Prevention (CDC) estimate that 400 million people are infected each year. Around 2.5 billion people, or 40 percent of the world's population, live in areas where there is a risk of dengue transmission. Unlike malaria, dengue can happen in both urban areas and rural areas, but research published in 2011 suggested that is more common in rural areas. Common symptoms of dengue fever are: Red spots or points on the skin, bleeding from the nose or gums, Frequent vomiting, Vomiting with blood, Black stools, Sleepiness, Constant crying, Abdominal pain, Excessive thirst (dry mouth), Pale, cold and clammy skin, Difficulty in breathing.

1.1 Dengue Hemorrhagic fever

Symptoms of this type of fever may be mild at first, but within a few days they gradually worsen. The patient may experience: Bleeding from the mouth, gums or nose, Damage to lymph and blood vessels, Small blood spots under the skin, weak pulse.

1.2 Dengue shock syndrome

DSS is a severe form of dengue. It can be fatal. The patient may experience: Intense stomach pain, Sudden hypotension or a fast drop in blood pressure, Heavy bleeding.

II.   Data Mining In Healthcare

The knowledge Discovery can applied to clinical data which comprise the process of data collection, data integration, data analysis. Health care systems reduce the cost of clinical tests. Healthcare organization the quality of service at reasonable cost is one of the major challenges. The effectiveness of eminence service to diagnosing patients treatments. Data mining is used in healthcare for customer relationship management decisions, physicians identify treatment and best practices, patients receive improved and more affordable healthcare services and healthcare insurers detect deception and exploitation. (10)

2.1 Diabetes healthcare

Data mining plays important role in healthcare for diabetes patients. Six types of treatments were identified in the 2005 WHO. They are Drug, Diet, Weight reduction, Smoke cessation, Exercise and Insulin [10].Regression based data mining techniques are used to investigate which mode of treatment is more effective for each age group. Data Mining architecture of diabetes fig1 [10].

2.2. Heart disease prediction

The data mining techniques for diagnosis of many diseases as heart disease, diabetes, stroke and cancer. Many data mining techniques used in the diagnosis of heart disease good accuracy. The detected of a heart disease several factors or symptoms is a multi-layered .The effective technique is to exploit the knowledge and experience of several specialists in assisting Diagnosis process .Data mining techniques as naïve bayes, neural network, decision tree and support vector machine for prediction and diagnosis of heart diseases.
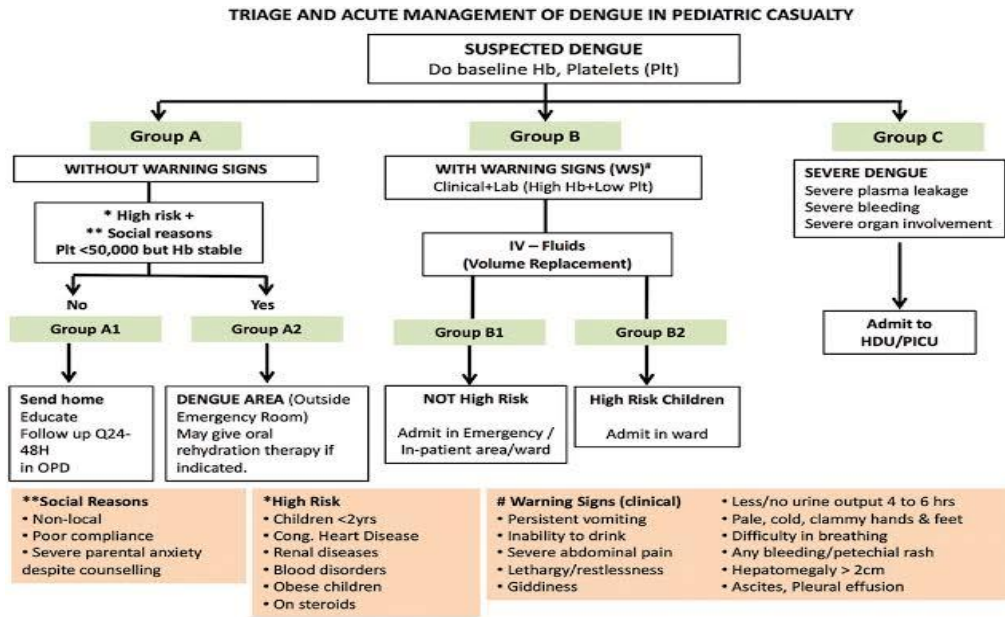
III.   Data Mining Techniques Disease Prediction

3.1. Decision Tree Algorithm

The decision trees one of the most frequently used techniques of data analysis [5].Decision tree easy to visualize and recognize and opposing to noise in data. Normally, decision trees are used to classify records to an appropriate class. Besides are applicable in both regression and associations tasks. The heart disease or coronary artery disease (CAD) or coronary heart disease (CHD) or ischemic heart disease (IHD) [4] is a broad term that can refer to any condition affects the heart. The Clinical decision support systems, literature presents a number of researches that have made use of artificial intelligence and data mining techniques. The popular decision tree algorithm is C4.5. It can make accurate predictions from the data but explain the patterns. The problems of the numeric attributes, missing values, pruning, estimating error rates complexity of decision tree induction and generating rules from tree [7].In term of projecting accuracy, C4.5 performs slightly better than CART and ID3[6].The learning and classification of C4.5 are generally [8].The parameter of C4.5 algorithm changing confidence threshold responsible for tree pruning, minimum numbers of instance are permitted at a leaf. It is possible to set the size of pruning set is the number of data part from which the last is used for tree pruning. C4.5 is used in classification problems and it is the most used algorithm for building DT. It is suitable for real world problems as it deal with numeric attributes and missing values.

Decision tree performs classification or regression of collected data set in the form of tree structure. In our model it helped to breakdown the dataset in to smaller and smaller subset to come closer for prediction based on the symptoms recorded from the patient. We also developed an associated decision tree based on these regressions. The decision tree is built top-down from root node and partitioning of data into subsets is done. The information gained based on these regression dataset is split on an attribute. Now choose the attribute with the closest prediction for dengue fever and repeat the same process for every set of records.

S.No	Type of disease	Data mining tool	Technique	Algorithm	Traditional Method	Accuracy level(%) from DM application
1	Heart Disease	ODND, NCC2	Classification	Naïve	Probability	60
2	Cancer	WEKA	Classification	Rules. Decision Table		97.77
3	HIV/AIDS	WEKA 3.6	Classification, Association Rule Mining	J48	Statistics	81.8
4	Blood Bank Sector	WEKA	Classification	J48		89.9
5	Brain Cancer	K-means Clustering	Clustering	MAFIA		85
6	Tuberculosis	WEKA	Naïve Bayes Classifier	KNN	Probability, Statistics	78
7	Diabetes Mellitus	ANN	Classification	C4.5 algorithm	Neural Network	82.6
8	Kidney dialysis	RST	Classification	Decision Making	Statistics	75.97
9	Dengue	SPSS Modeler		C5.0	Statistics	80
10	IVF	ANN, RST	Classification			91
11	Hepatitis C	SNP	Information Gain	Decision rule		73.20

TABLE 1: DATAMINING APPLICATIONS IN HEALTHCARE [9]

TRIAGE AND ACUTE MANAGEMENT OF DENGUE IN PEDIATRIC CASUALTY

SUSPECTED DENGUE
Do baseline Hb, Platelets (Plt)

Group A
WITHOUT WARNING SIGNS

* High risk +
** Social reasons
Plt <50,000 but Hb stable

No → **Group A1**
Yes → **Group A2**

Group A1
Send home
Educate
Follow up Q24-48H
in OPD

Group A2
DENGUE AREA (Outside Emergency Room)
May give oral rehydration therapy if indicated.

Group B
WITH WARNING SIGNS (WS)#
Clinical+Lab (High Hb+Low Plt)

IV – Fluids
(Volume Replacement)

Group B1
NOT High Risk
Admit in Emergency / In-patient area/ward

Group B2
High Risk Children
Admit in ward

Group C
SEVERE DENGUE
Severe plasma leakage
Severe bleeding
Severe organ involvement

Admit to HDU/PICU

**Social Reasons
• Non-local
• Poor compliance
• Severe parental anxiety despite counselling

*High Risk
• Children <2yrs
• Cong. Heart Disease
• Renal diseases
• Blood disorders
• Obese children
• On steroids

 # Warning Signs (clinical)
• Persistent vomiting
• Inability to drink
• Severe abdominal pain
• Lethargy/restlessness
• Giddiness

• Less/no urine output 4 to 6 hrs
• Pale, cold, clammy hands & feet
• Difficulty in breathing
• Any bleeding/petechial rash
• Hepatomegaly > 2cm
• Ascites, Pleural effusion

IV. Fuzzy Logic Implementation

Fuzzy logic algorithm was applied to analyze the relevance of various data obtained from the hospital and the other health centers reports. An algorithm for data mining was devised based on this fuzzy logic to predict the dengue fever status by analyzing the symptoms like duration of fever, temperature, red spots, headache, muscle and joint pain. The accuracy of prediction model depends on the quality of data. i.e. higher quality of data will lead to more accurate prediction. This method should be able to predict the dengue fever with reasonable accuracy.
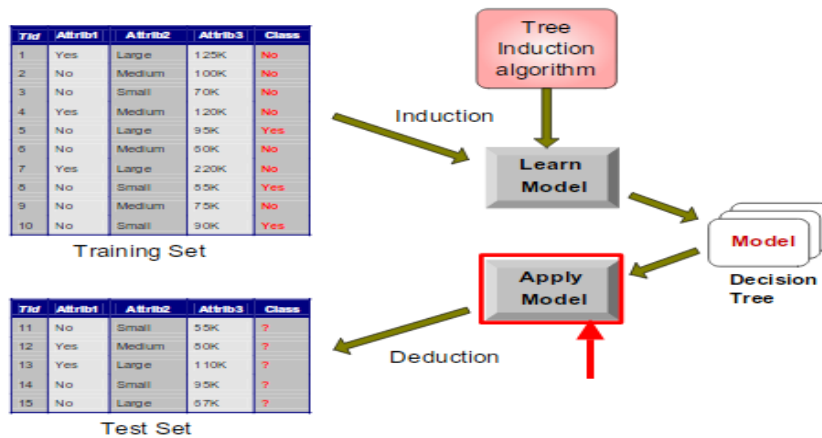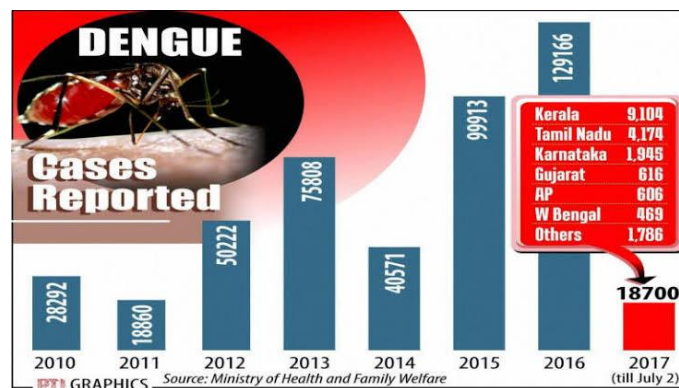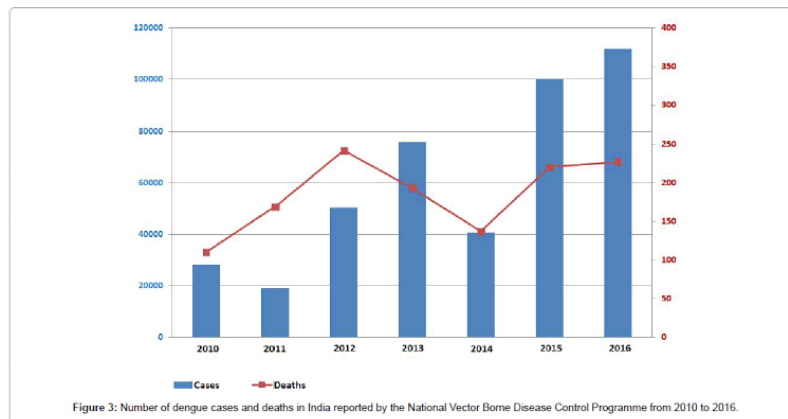

Platelet count:

Platelet Count	Less than 25,000	25,000-49,000	50,000-74,000	75,000-1,00,000	1,00,0000
No. Of Cases	382	276	254	189	153

V. Previous Work

There are some researchers who work on dengue classification such as Tanner, et al and Tarig, et al. The team of Tanner classified 1,200 patients using decision tree approach. They found 6 significant features and they got 84.7 % correctness [2]. A combination of the self-organizing map (SOM) and multilayer feed-forward neural networks (MFNN) was employed for the risk prediction of dengue patients in the Tarig's research. They clustered patients into 2 groups which are low risk and high risk using three criteria [3]. They used only examples from Day0 until Day2 (Day2 refers to 2 days after the day of defervescence of fever).On the other hand, they got only 70% correctness. Fatimah Ibrahim et al. [1] predicted the day of defervescence of fever (day0) from 252 dengue patients (4 DF and 248 DHF). They used Multi-Layer Perceptrons (MLP) and got 90% correctness.

Figure 3: Number of dengue cases and deaths in India reported by the National Vector Borne Disease Control Programme from 2010 to 2016.

VI.   Conclusion

The data obtained from the fuzzy logic and the decision tree implementation provides predicable analysis of dengue fever. The record set collected from the healthcare organization is subjected to regression and the most absolute data is obtained for the analysis. An algorithm for data mining was devised based on this fuzzy logic to predict the dengue fever status by analyzing the symptoms like duration of fever, temperature, red spots, headache, muscle and joint pain. This method should be able to predict the dengue fever with reasonable accuracy.

References

[1]   F. Ibrahim, M. N Taib, W. A. B. Wan Abas, C.G. Chan and S. Sulaiman, A novel dengue fever (DF) and dengue haemorrhagic fever (DHF) analysis using artificial neural network (ANN), *Computer Methods and Programs in Biomedicine*, No.79, 2005, pp. 273-281.

[2]   L. Tanner, M. Schreiber, J.G. Low, A. Ong, T.Tolfvenstam, Y.L. Lai, L.C. Ng, Y.S. Leo, L.Thi Puong, S.G. Vasudevan, C.P. Simmons,M.L. Hibberd and E.E. Ooi, Decision Tree Algorithms Predict the Diagnosis and Outcome of Dengue Fever in the Early Phase of Illness, *PLoS Neglected Tropical Disease*, Vol.2, 2008.

[3]   T. Faisal, F. Ibrahim and M.N. Taib, A noninvasive intelligent approach for predicting the risk in dengue patients, *Expert Systems with Application,*Vol.37, No.3, 2010, pp. 2175- 2181.

[4]   World Health Organization, *Guideline for Treatment of Dengue Fever/Dengue Haemorrhagic Fever*, 1999

[5]   M. Anbarasi, E. Anupriya, N.ch.s.n.Iyengar, *Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm*, International Journal of Engineering Science and Technology,2010.

[6]   Marcel A.J. van Gerven, *Predicting carcinoid heart disease with the noisy-threshold classifier*, Elsevier, 2007.

[7]   Jesmin Nahar, Tasadduq Imam, *Computational intelligence for heart disease diagnosis: A medical knowledge driven approach*, Elsevier, 2013.

[8]   Nan-Chen Hsieh & Lun-Ping Hung & Chun-Che Shih, *Intelligent Postoperative Morbidity Prediction of Heart Disease Using Artificial Intelligence Techniques*, J Med System, 2012.

[9]   M.Durairaj, V.Ranjani, Data Mining Applications in Health care sector: A study , International Journal of scientific & technology research Vol-2,oct 2013,ISSN 2277-8616.

[10]   Abdullah A.Alijumah, Mohammed Gulam Ahamad, Mohammad khubab siddiqui, Application of data mining : Diabetes health care in young and old patients, Journal of king saud university ,computer & information sciences(2013) 25, 127-136.

[11]   Dr.Arun kumar P.M, Karthick.P, Ganesan.M and Madhan A.S, Dengue Disease Prediction Using Decision Tree and Support Vector Machine, International Journal of computer science and Engineering-(ICET'17)-March 2017.

[12]   R.Tamilarasi, Dr.R.Porkodi, Disease Prediction Techniques in data mining for health care, International Journal of Emerging research in management & technology ,ISSN:2278-9359(Vol-4,Issue-3)-March 2015.