International Conference on Computing, Communication System and Informatics Management (ICCCSIM)

Venue : Hotel RAMADA , Bur Dubai, UAE

Date : 29 – 30 July, 2012

Analysis of ECGs Survey Data Using Threshold Inference Engine

Saria Safdar

[13]Computer Software Engineering Department

Military College of Signals (MCS), NUST

Islamabad, Pakistan

sariasafdar@yahoo.com

Shoab Ahmad Khan

Computer Engineering Department

Electrical and Mechanical Engineering (EME), NUST

Islamabad, Pakistan

kshoab@yahoo.com

Fahim Arif

fahim.mcs@mcs.edu.pk

Abstract :

Cardiovascular diseases are increasing day by day in Pakistan and now reach to a ratio of around 35 to 40 per cent of the total disease burden in Pakistan. This increasing ratio needs a detailed analysis of the overall geographical distribution of heart patients and also the most aggregating attributes (age, weight, income etc). To cater this situation a Threshold Inference Engine is designed which generates the association rules to extract the city wise more risk increasing attributes, and the common heart disease in that city. Automated Minnesota code is used for the verification of the collected ECGs. The generated results of the Threshold Inference Engine successfully and efficiently generate a detailed report of each city describing the common heart disease and the attributes.

Keywords - *Aggregation, Cardiac Arrhythmias, Centroid, ECG, Fuzzification, Inference Engine, Membership etc.*

I. INTRODUCTION

Sudden cardiac arrest is the leading cause of death worldwide and its ratio in Pakistan is on the highest [1, 2]. Statistics of heart patients in Pakistan is increasing day by day. The situation is getting worst and needs a detailed analysis of the common lifestyle attributes which are aggregating the disease and also the citywise most common heart disease.

To achieve this task a Threshold Inference Engine is designed which takes clustered data as input and generates statistical rules. These rules are further used for the inference purpose. Collected ECGs are verified

with the automated Minnesota code. The result is a detailed report which highlights the geographical areas with maximum disease and the attributes (weight, age, income, drugs etc) contributing more in the heart disease in the particular area.

The generated report will provide great benefit for use by Health Organization and International Health Organization like ministry of health and WHO.

II. LITERATURE SURVEY

Automated ECG survey data analysis from last few years has been an active area of research. Many fuzzy approaches are proposed for ECG diagnosis and classification [3,4] including Fuzzy Adaptive-Resonance TheoryMAP which has been employed to classify cardiac arrhythmia. A hybrid neurofuzzy system was used for ECG classification of myocardial infarction (MI) [5].

A lot of techniques have been developed by different researchers including; T. M. NAZMY et al, presents an intelligent diagnosis system for (ECG) classification by using hybrid approach of adaptive neuro-fuzzy inference system (ANFIS) [6]. Adaptive neuro-fuzzy inference system is commonly used for the classification of ECG signals. To detect premature contraction ANFIS was used in [7].

Sucharita Mitra et al in [8], describes a rule-based rough-set decision system in which an inference engine is used for the detection of the heart disease. Vijilal et al proposed “a hybrid soft computing technique called Adaptive Neuro-Fuzzy Inference System (ANFIS) to separate the Electroencephalogram (EEG) signal from its Electrooculogram (EOG)” [9].

III. THRESHOLD INFERENCE ENGINE (TIE)

Threshold Inference engine is designed to extract the most common lifestyle attributes that are aggregating more in the heart disease. It consists of a knowledge base, inference engine and set of rules. Rules are extracted from the knowledge base and inference engine is applied on individual rule to extract the attributes.

A. A .Algorithm:

The steps that are designed for the algorithm of the proposed Threshold based Inference Engine are presented in table 1. The rest of the section will explain in detail each step, and the outcome of each step.

TABLE 1 TIE ALGORITHM

Step 1: Rules Generation on cities
Step2: Fuzzification on rules attributes
Step 3: Apply Threshold on Membership
Step 4: Implication of Rules
Step 5:Calculate Centroid

Step 6: Aggregation of centroids

\

Step 7 : Report Generation

B. Description of Steps in TIE

Detail description of steps of TIE is presented in this section.

A. Preprocessing for the TIE

This step involves the preparation of the ECGs database.

B. Minnesota Validation

Minnesota code is a well known medical science paper which is used by the cardiologists for the verification of the ECGs [10]. It has number of clauses which have ranges for the attributes P,QRS, T . These ranges must be satisfied by the selected ECG. This research automated the Minnesota code so that ECGs can be verified and through an automated system and may save time of cardiologists.

When the ECG is uploaded it needs verification i.e whether this ECG is accurate to consider; it contains all the parameters or not. If the ECG is verified by the Minnesota code then those ECG parameters, after calculation are saved in the database however if it does not pass the verification criteria it is discarded.

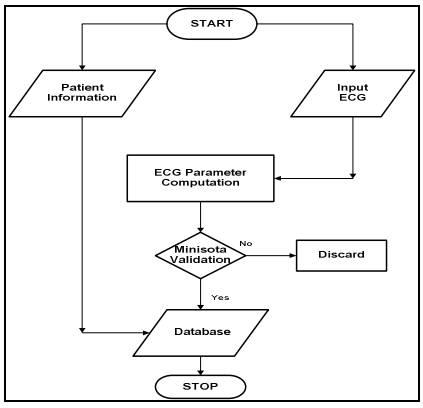

FIGURE 1 MINNESOTA FLOWCHART

The figure 1 represents the flowchart of the preprocessing of the collected ECGs. The system starts with the patient data and the input ECGs. These ECGs are further input to the parameters computations where all the P,QRS,T parameters onset and offset is calculated. These calculated parameters are fed into the automated Minnesota code where these parameters calculated values are matched with the values prescribed in

the Minnesota code. If both calculated and prescribed values are same than that ECG is saved for further computation. This verification through the automated code saves the time of the cardiologists.

C. Rules Generation on Cities

Mean and standard deviation is calculated for each attribute in the database. To achieve cities that have more diseased people threshold is selected. The selected threshold shows the number of diseased people in the province e.g if threshold value is five it will retrieve all the cities in which percentage of diseased people is five or above five. The result is the set of rules, each rule against each city.

"*If weight (93), height (5'5''), age (59) ,education (Fsc), and employment (retired), and income (35k), and drugs (level never used), and BP (110< B <140), and hypertension with stage (permanent), then heart disease (ventricular fibrillation), with stage (serious).*"

This rule describes that in Lahore these are the attributes which are aggregating the disease. This rule is a detailed one so there is a need of a more generalized rule. To achieve this further steps of TIE are applied.

D. Fuzzification on Rules Attributes

The process of applying membership on attributes is fuzzification.. A *membership function* (MF) is defined as "The degree an object belongs to a fuzzy set is denoted by a membership value between 0 and 1. μ*(x)* is called the membership function (or MF) of *x"[11]*. In TBIE trapezoidal membership function is used to calculate membership of each attribute. Table 2 shows the membership calculated for each attribute.

City ID	Age	Weight	Education	Employment	Total FI	Drugs	BP
373	0.39	0.33	0.25	0.7	0.50	0	0.75

TABLE 2 MEMBERSHIP VALUES

E. Maximum Membership holding Attributes

Attributes with maximum membership are selected to achieve most risk increasing attributes. Table 3 shows the selected attributes with there membership values against each city.

TABLE 3 MAX. MEMBERSHIP ATTRIBUTES

City ID	BP	Employment	Total FI	Age	Weight
373	0.75	0.7	0.50	0.39	0.33

F. Implication

When all the statistical analysis is done on the rules antecedent part (age, weight etc) same needs to be applied on the rues consequent part (disease with stages serious, mild, symptoms). This statistical analysis on rules consequent part finally generalizes the rule i.e city with first five more aggregating attributes with most common disease with its stage. Table 4 shows the result of allegation.

TABLE 4 IMPLICATION

City ID	BP	Employment	Age	Weight	Disease	SOD
373	0.75	0.7	0.39	0.33	Sinus Rhythm	Serious

G. Province Level Aggregation

Province level aggregation shows the common attributes and the common disease at the province level. To achieve this task first centroid calculation is done on each city, and then the aggregation of all the collected centroids is done.

A. City Level Centroid Calculation

The centroid formula is applied on each cities attributes; the result is a single index which represents the index among the five selected attributes. The attribute on the calculated index is retrieved; this formula is applied on all the cities to get one attribute from all cities. By this the province Punjab will get a set of attributes which shows that these attributes are common in cities of Punjab.

$$C = \frac{\sum_{i=1}^{n} x_{i(index)} \cdot \mu(x_i)}{\sum_{i=1}^{n} \mu(x_i)}$$

$x_{i\,(index)}$= index of attribute

$\mu(x_i)$ = membership of attribute

$\sum_{i=1}^{n} \mu(x_i)$ = sum of membership values

$\sum_{i=1}^{n} x_{i\,(index)}.\mu(x_i)$ = sum of attributes index value and its membership

The table 5 shows the centroids calculated from all cities of each province.

TABLE 5 CENTROID CALCULATION

City ID	Weight	Drugs	Blood Pressure	Age
373	86	Still smoking	100<A>130	45-49

City ID	Hypertension	Income	Weight	Blood Pressure
332	Persistent	15000	63.5	120<A>150

City ID	Employment	Income	Weight	Drugs
351	Retired	22000	82	Quitted

So the centroid calculation is done on each city of each province.

B. Aggregation

It is the process in which the output sets of each Province are joined to make a decision set. The decision set represents the attributes that are common in Pakistan for the heart diseased people. The input to the aggregation process is the truncated sets returned by the individual Provinces and the output is a single set. Aggregation occurs only once just for the final step i.e the aggregation of all the centroids. From all the sets the unique attributes will be selected.[12]

Output Set= [Age(35-39), Employment(Retired) Blood Pressure(100<A<120), Hypertension (Persistent),]

B. H. Generation of Report

This section describes the report generation in which a complete details of all cities is shown with there attributes and the stage of disease. The generated report is very useful as it visually shows the city wise distribution of heart diseases in Pakistan shown in figure 2.

To generate the report standard crystal report format is used

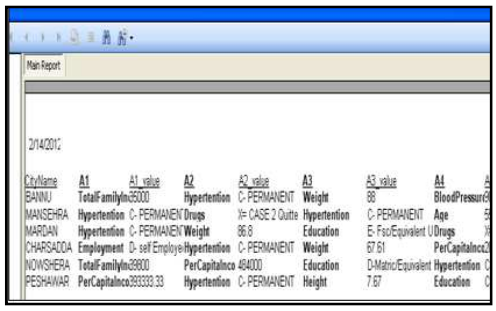

FIGURE 2 REPORT OF CITIES

C. A. Discussion through Graphs

This section shows the graphical representation of graph.

D. B. Weight

Weight is important in some cities, as the graph shows the weight above 70kg are on high risk shown in figure 3.

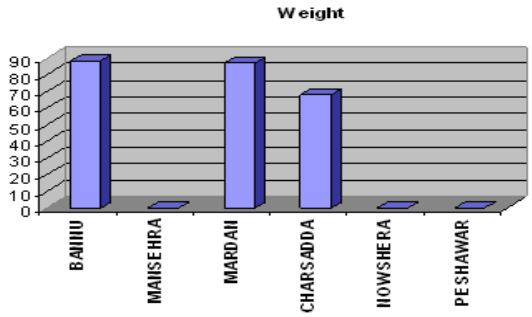

Figure 3 WEIGHT

C. Drugs

The graph shows that drug is an important factor for the heart diseased people and its ratio is high in all cities as shown in figure 4.

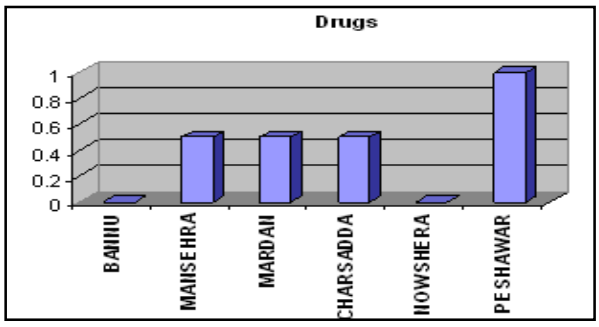

Figure 4 DRUGS

D. Hypertension

The graph in figure shows that the hypertension is the main cause of heart diseases as it is present in all cities of Pakistan as shown in figure 5

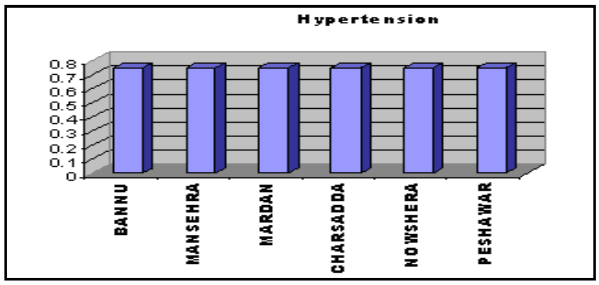

Figure 5 HYPERTENSION

1V. CONCLUSION

The analysis clearly shows that prevalence of heart diseases in Pakistan is on high risk. This analysis provides great benefit to the government as it captures the overall distribution of heart diseases in Pakistan. The result shows that Threshold Inference Engine successfully and efficiently extracted the most aggregating attributes and the common heart diseases in various cities. The results shows that hypertension is present in all cities of Pakistan and is the highest membership holding attribute aggregating the heart disease.

The future work can be is to train the membership values through the resilient back propagation algorithm to make the system more intelligent.

V.ACKNOWLEDGMENT

I would like to pay my gratitude to my all respected teachers and parents who guide me to achieve this success. I would also like to thank Mr Aihab Khan for their guidance and support.

REFERENCES

1. http://www.emro.who.int/pakistan/programmes_ncd.htm
2. World Bank report Heart disease, diabetes on the rise in Pakistan – The Express Tribune.mht by Mustafadon Islamabad Pakistan Feb 10,2011
3. F. M. Ham and S. Han, "Classification of cardiac arrhythmia using fuzzy ARTMAP," IEEE Trans. Biomed. Eng., vol. 43, no. 4, pp. 425–430, Apr. 1996.
4. Y. H. Hu, W. J. Tompkins, J. L. Urrusti, and V. X. Afonso, "Applications of artificial neural networks for ECG signal detection and classification," J. Electrocardiol., vol. 26, pp. 66–73, 1993
5. P. Bozzola, G. Bortolan, C. Combi, F. Pinciroli, and C. BroHet, "A hybrid neuro-fuzzy system for ECG classification of myocardial infarction," in Proc. Comput. Cardiol., Indianapolis, IN, 1996, pp. 241–244.
6. T. M. NAZMY, H. EL-MESSIRY, B. AL-BOKHITY "Adaptive Neuro-Fuzzy Inference System for Classification of ECG Signals" Journal of Theoretical and Applied Information Technology © 2005 - 2009 JATIT. All rights reserved.
7. LAM Si Man, WAN Feng, DONG Mingchui Faculty of Science and Technology, University of Macau Macau China "Automatic Detection of Premature Ventricular Contractions Using Adaptive Neuro-Fuzzy Inference System" 2009
8. Sucharita Mitra, Student Member, IEEE, Madhuchhanda Mitra, and B. B. Chaudhuri, Fellow, IEEE "A Rough-Set-Based Inference Engine for ECG Classification" IEEE Transactions on Instrumentation and Measurement, VOL. 55, NO. 6, DECEMBER 2006
9. C. Kezi Selva Vijilal, P.Kanagasabapathy2, Stanley Johnson' and Vinodh Ewards' "Artifacts Removal in EEG Signal using Adaptive Neuro Fuzzy Inference System" IEEE - ICSCN 2007, MIT Campus, Anna University, Chennai, India. Feb. 22-24, 2007. pp.589-591.
10. The Minnesota Code Classification System= for Electrocardiographic Findings
11. http://www.orsc.edu.cn/~liu/Lecture/USet/Memberfunction/Memberfunction.pdf
12. http://www.mathworks/Fuzzy Inference Systems Tutorial (Fuzzy Logic Toolbox™).mht