A Fuzzy Rule based Approach to Predict Risk Level of Heart Disease

By Kantesh Kumar Oad, Xu DeZhi & Pinial Khan Butt
Central South University, China

Abstract - Health care domain systems globally face lots of difficulties because of the high amount of risk factors of heart diseases in peoples (WHO, 2013). To reduce risk, improved knowledge based expert systems played an important role and has a contribution towards the development of the healthcare system for cardiovascular disease. To make use of benefits of knowledge based system, it is necessary for health organizations and users; must need to know the fuzzy rule based expert system’s integrity, efficiency, and deployments, which are the open challenges of current fuzzy logic based medical systems. In our proposed system, we have designed a fuzzy rule based expert system and also by using data mining technique we have reduced the total number of attributes. Our system mainly focuses on cardiovascular disease diagnosis, and the dataset taken from UCI (Machine Learning Repository). We explored in the existing work. The majority of the researcher’s experimentation was made on 14 attributes out of 76. While, in our system we took advantage of 6 attributes for system design. In the preliminary stage UCI, data participated in suggested system that will get outcomes. The performance of the system matched with Neural Network and J48 Decision Tree Algorithm.

Keywords: fuzzy reasoning, heart disease and diagnose, data mining.

GJCST-C Classification : I.5.1
A Fuzzy Rule based Approach to Predict Risk Level of Heart Disease

Kantesh Kumar Oad °, Xu DeZhi ° & Pinial Khan Butt °

Abstract - Health care domain systems globally face lots of difficulties because of the high amount of risk factors of heart diseases in peoples (WHO, 2013). To reduce risk, improved knowledge based expert systems played an important role and has a contribution towards the development of the healthcare system for cardiovascular disease. To make use of benefits of knowledge based system, it is necessary for health organizations and users; must need to know the fuzzy rule based expert system's integrity, efficiency, and deployments, which are the open challenges of current fuzzy logic based medical systems. In our proposed system, we have designed a fuzzy rule based expert system and also by using data mining technique we have reduced the total number of attributes. Our system mainly focuses on cardiovascular disease diagnosis, and the dataset taken from UCI (Machine Learning Repository). We explored in the existing work. The majority of the researcher's experimentation was made on 14 attributes out of 76. While, in our system we took advantage of 6 attributes for system design. In the preliminary stage UCI, data participated in suggested system that will get outcomes. The performance of the system matched with Neural Network and J48 Decision Tree Algorithm.

Keywords: fuzzy reasoning, heart disease and diagnosis, data mining.

I. Introduction

Just recently peoples are pressured over their health and wellness troubles, in the majority of the countries proportion of cardiovascular disease enhancing really quick and it has actually become the leading cause / death worldwide [1][2], and it is came to be taken into consideration a "second epidemic," changing transmittable conditions as the leading cause of death [3][4]. Health domain application is one of one of the most active study area nowadays. Ideal example of health domain application is the detection system for cardiovascular disease based on computer system assisted diagnosis strategies, where the information acquired from numerous other sources and is evaluated based on computer-based application. Before it was very time consuming job to get knowledge from physician and include this knowledge to computer system program by hand into data base medical decision support system and this was totally depending upon clinical experts' concepts which may be subjective.

This trouble has really been resolved using expert systems; get physician, group, knowledge and certain human client details, intelligently. In boosting outcomes at a couple of healthcare companies and strategy internet sites, expert system has actually operated by making needed clinical knowledge quickly readily available to know-how users [5] Taking care of clinical needs, such as making certain specific medical diagnoses, evaluating in a quick manner for avoidable health problem, or avoiding undesirable drug occasions, are the most standard exploitation of Expert System [6]. Expert System could also be possibly lessen costs, progression performance, and reduce client stress. These systems are classified into 2 groups namely (1) Knowledge based and (2) non-knowledge based [7]. The knowledge based system consists of rules (if-then statements). Expert system that is implemented with the assistance of artificial intelligence has the ability to support in a new setting and to learn for instance [8][9]. Given that the concept of computer-based Clinical Decision Support System aroused at first, significant research has actually been made in both academic and practical areas. Many obstacles are longer to impede the effective application of expert systems in scientific environments, among which portrayal and reasoning concerning clinical understanding predominantly under anxiety is the locations that require improved methodologies and strategies [10][11].

In our proposed system, mainly focus on cardiovascular disease diagnosis. We have taken dataset taken from UCI (Machine Learning Repository). UCI database consists of 76 attributes; we investigated in the existing work. Most of the experiments were made by using a subset of 14 from UCI. While, in our system we reduced the number of input attributes that will reduce the number of diagnostic results and we used seven attributes to experiment. From seven attributes, we used six attributes as input, and one attribute for output. Numerical data will enter in into suggested system, and in the last; system will get prediction results. The primary objective of our research is to make and carry out fuzzy rule based system for heart disease peoples.
II. RELATED WORK


III. A FUZZY RULE BASED APPROACH TO PREDICT RISK LEVEL OF HEART DISEASE

As you can see in figure (1) the process begins with data processing (see section 3.1). In a second step, we reduced the number of attributes and then this processed data (symptoms) inserted into fuzzy system using MATLAB programming. After the fuzzy model is successfully developed, the prediction of the symptoms will start and lastly performance based on the result will be analyzed at the end of the development phase.

a) Data Processing

The function of data processing is to draw out the significant information from the raw data collection useful for heart disease prediction, and these data sets should be transformed into the needed style for the level of risk prediction. Due to large amount of data there are chances of errors on it so before processing heart disease datasets, the original raw datasets must be processed; as a result in data processing phase, we cleaned, transformed and analyzed it into the row column format after taking out the unnecessary ones.

b) Data Formats

For recap or view purpose we should make data in a format so it can be easily reviewed. We have converted data to a .csv (Comma Separated Values) documents format also called comma delimited file. .csv file is a particularly configured text file; which establishments spreadsheet or standard database-style information in a quite easy style, with one record for each line, and each field within that document separated by a comma.

c) Attributes Selection

To minimize disease data sets (feature selection), we have used data reduction technique. Attribute selection lessens the data set size by getting rid of redundant or unnecessary and extracting with a minimized set of characteristics has an extra benefit [22]. Primarily we have actually made use of the data mining tool to get decreased collection of datasets. An attributes selection method contains four steps, named (1) subset generation; (2) subset evolution, (3) stop criterion (4) outcome validation [23] subset generation is a searching technique and we have made use of Best First Search Method using DM tool. Each new subset evaluated and matched with earlier best one according to a certain evolution criterion. It changes the previous finest subset if the new subset turns out to be much better. The process of subset generation and evolution is duplicated till a given quitting criterion is pleased.

By using selected method, we have obtained 6 attributes from a total of 14 attributes.

d) System Phases

Fuzzy system used in circumstances where we have a trouble of uncertainty. When the trouble has a
dynamic behavior; Fuzzy Rule Base is an appropriate system for handling this issue. In the primary step of fuzzy expert system design; figure out the input and result variables. As it is described in section (4.3) we selected six inputs and one output variable. Then, we have actually made the membership functions for all variables used in the system. Membership function is primarily a visual portrayal of a fuzzy set; and determines the membership degree of objects to fuzzy sets. Rule evolution and defuzzification process described in the next sections.

**Figure 2**: General Schema of Fuzzy Logic

### i. Fuzzification and Membership Function
Fuzzification is a process of fuzzifying all inputs and output. Determine the degree to which these inputs and outputs belong to each of the suitable fuzzy sets.

#### Age:
We have actually separated age into 3 fuzzy sets (Young, Middle and Old) and ranges of these fuzzy sets are determined in Table (1). We have used triangular and trapezoidal Membership function for fuzzy sets.

- \( \mu_{young}(x) = 1 \) when \( x \in [1, 29] \)
- \( \mu_{young}(x) = (38- x)/(38-29) \) when \( x \in [29, 38] \)
- \( \mu_{young}(x) = 0 \) otherwise.
- \( \mu_{middle}(x) = (x-33)/(39-33) \) when \( x \in [33, 39] \)
- \( \mu_{middle}(x) = (45- x)/(45-39) \) when \( x \in [39, 45] \)
- \( \mu_{middle}(x) = 0 \) otherwise.
- \( \mu_{old}(x) = 0 \) when \( x \in [1, 40] \)
- \( \mu_{old}(x) = (x-40)/(60-40) \) when \( x \in [40, 60] \)
- \( \mu_{old}(x) = 1 \) when \( x \in [60, 100] \).

#### Blood Pressure:
We have actually separated this input fuzzy set into 4 levels called (Low, Medium, High and Very high). Trapezoidal MFs are used for (Low and Very high) and MFs of (medium and high) sets are triangular.

- \( \mu_{low}(x) = 1 \) when \( x \in [1, 111] \)
- \( \mu_{low}(x) = (134- x)/(134-111) \) when \( x \in [111, 134] \)
- \( \mu_{low}(x) = 0 \) otherwise.
- \( \mu_{medium}(x) = (x-126)/(139-126) \) when \( x \in [126, 139] \)
- \( \mu_{medium}(x) = (152- x)/(152-139) \) when \( x \in [139, 152] \)
- \( \mu_{medium}(x) = 0 \) otherwise.
- \( \mu_{high}(x) = (x-142)/(157-142) \) when \( x \in [142, 157] \)
- \( \mu_{high}(x) = (172- x)/(172-157) \) when \( x \in [157, 172] \)
- \( \mu_{high}(x) = 0 \) otherwise.
- \( \mu_{veryhigh}(x) = 0 \) when \( x \in [1, 154] \)
- \( \mu_{veryhigh}(x) = (x-154)/(172-154) \) when \( x \in [154, 172] \)
- \( \mu_{veryhigh}(x) = 1 \) when \( x \in [172, 300] \).

#### Cholesterol:
Cholesterol has 3 fuzzy sets (Low, Medium and High). Ranges for these fuzzy sets are determined in Table (1).

- \( \mu_{low}(x) = 1 \) when \( x \in [1, 151] \)
- \( \mu_{low}(x) = (197- x)/(197-151) \) when \( x \in [151, 197] \)
- \( \mu_{low}(x) = 0 \) otherwise.
- \( \mu_{medium}(x) = (x-188)/(219-188) \) when \( x \in [188, 219] \)
- \( \mu_{medium}(x) = (250- x)/(250-188) \) when \( x \in [219, 250] \)
- \( \mu_{medium}(x) = 0 \) otherwise.
- \( \mu_{high}(x) = 0 \) when \( x \in [1, 217] \)
- \( \mu_{high}(x) = (x-217)/(263-217) \) when \( x \in [217, 263] \)
- \( \mu_{high}(x) = 1 \) when \( x \in [263, 500] \).

#### Heart Rate:
Heart Rate split into 3 fuzzy sets named (Low, Medium and High). Ranges for these fuzzy sets are trapezoidal and MF of (Medium) is triangular.

- \( \mu_{low}(x) = 1 \) when \( x \in [1, 100] \)
- \( \mu_{low}(x) = (141- x)/(141-100) \) when \( x \in [100, 141] \)
- \( \mu_{low}(x) = 0 \) otherwise.
- \( \mu_{medium}(x) = (x-111)/(152-111) \) when \( x \in [111, 152] \)
- \( \mu_{medium}(x) = (194- x)/(194-111) \) when \( x \in [111, 154] \)
- \( \mu_{medium}(x) = 0 \) otherwise.
- \( \mu_{high}(x) = 0 \) when \( x \in [1, 154] \)
- \( \mu_{high}(x) = (x-154)/(172-154) \) when \( x \in [154, 172] \)
- \( \mu_{high}(x) = 1 \) when \( x \in [172, 300] \).

#### Old Peak:
Old Peak divided into 3 fuzzy sets (Low, Risk and Terrible). These fuzzy sets have actually been shown in Table (1) with their ranges.

- \( \mu_{low}(x) = 1 \) when \( x \in [0, 1] \)
- \( \mu_{low}(x) = (2- x)/(2-1) \) when \( x \in [1, 2] \)
- \( \mu_{low}(x) = 0 \) otherwise.
- \( \mu_{risk}(x) = (x-1.5)/(2.8-1.53) \) when \( x \in [1.5, 2.8] \)
- \( \mu_{risk}(x) = (4.2- x)/(4.2-2.8) \) when \( x \in [2.8, 4.2] \)
- \( \mu_{risk}(x) = 0 \) otherwise.
- \( \mu_{terrible}(x) = 0 \) when \( x \in [0, 2.55] \)
- \( \mu_{terrible}(x) = (x-2.55)/(4.2-2.55) \) when \( x \in [2.55, 4.2] \)
- \( \mu_{terrible}(x) = 1 \) when \( x \in [4.2, 6] \).
Thallium Scan:
This input field includes 3 fuzzy sets: (Normal, Fix Defect and Reverse Defect). For each and every fuzzy set we have defined a value that we use them for system testing. These fuzzy sets with their values are shown in Table (1).

\[
\begin{align*}
\mu_{\text{normal}}(x) &= \frac{(x-1)}{(2-1)} \text{ when } x \in [1, 2] \\
\mu_{\text{normal}}(x) &= \frac{(3-x)}{(3-2)} \text{ when } x \in [2, 3] \\
\mu_{\text{normal}}(x) &= 0 \text{ otherwise.}
\end{align*}
\]

\[
\begin{align*}
\mu_{\text{Fix Defect}}(x) &= \frac{(x-3)}{(4.5-3)} \text{ when } x \in [3, 4.5] \\
\mu_{\text{Fix Defect}}(x) &= \frac{(6-x)}{(6-4.5)} \text{ when } x \in [4.5, 6] \\
\mu_{\text{Fix Defect}}(x) &= 0 \text{ otherwise.}
\end{align*}
\]

\[
\begin{align*}
\mu_{\text{Rev Defect}}(x) &= \frac{(x-6)}{(6.5-6)} \text{ when } x \in [6, 6.5] \\
\mu_{\text{Rev Defect}}(x) &= \frac{(7-x)}{(7-6.5)} \text{ when } x \in [6.5, 7] \\
\mu_{\text{Rev Defect}}(x) &= 0 \text{ otherwise.}
\end{align*}
\]

Over we have actually selected chosen features, now we have split all inputs into fuzzy sets; we have actually utilized trapezoidal and triangular membership functions in system.

<table>
<thead>
<tr>
<th>Input Field</th>
<th>Range</th>
<th>Fuzzy Sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>&lt;38</td>
<td>Young</td>
</tr>
<tr>
<td></td>
<td>33-45</td>
<td>Middle</td>
</tr>
<tr>
<td></td>
<td>40&gt;</td>
<td>Old</td>
</tr>
<tr>
<td>Blood Pressure</td>
<td>&lt; 138</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>126-152</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td>142-172</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>154&gt;</td>
<td>Very High</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>&lt; 197</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>188-250</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td>217&gt;</td>
<td>High</td>
</tr>
<tr>
<td>Heart rate</td>
<td>&lt; 141</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>111-194</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td>152&gt;</td>
<td>High</td>
</tr>
<tr>
<td>Old Peak</td>
<td>&lt; 2</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>1.5 - 4.2</td>
<td>Risk</td>
</tr>
<tr>
<td></td>
<td>2.5&gt;</td>
<td>Terrible</td>
</tr>
<tr>
<td>Thallium Scan</td>
<td>3</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Fixed Defect</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>Reversible Defect</td>
</tr>
</tbody>
</table>

Table 1: Risk factors and Ranges

ii. Rules Evolution
In Fuzzy Rule Base System, rules play an important role in the prediction. The rules deliver/provide a sense to linguistic variables and MF (membership function). So we have occupied these fuzzyfied inputs in antecedent part of the rules. In this research, we have actually utilized 19 rules to predict heart disease in the patient.

In our system antecedent part of the rule consist of only single part that will opinion result of antecedent development. Fuzzy logic will govern risk level and this prediction indeed relies on rules that we have made. The made rules we have applied in using Mat Lab R2012a in the rule editor. After then fuzzification, crisp will certainly examine by passing in rule instance.

iii. Output (Defuzzification)
In this system, we have one output variable, which divided to 2 fuzzy sets (healthy and sick). For defuzzification procedure, designed system makes use of the Centroid method, determines the area of membership functions within the range of (output) variable.

\[
\text{CoA} = \frac{\int_{\text{min}}^{\text{max}} f(x).dx}{\int_{\text{min}}^{\text{max}} f(x) dx}
\]

CoA is the center of area/gravity; x is the linguistic variable and x (min) and x (max) signifies the arrays of variables.

IV. System Testing
To compare the performance of our system with Neural Network and J48 Decision Tree, we have divided Cleveland Heart Disease datasets into 2 parts e.g. training data 60% and testing data 40%. And this efficiency/performance is usually matched in term of sensitivity, specificity and accuracy. These terms normally took advantage of diagnostic approaches to enhance analysis results.

\[
\begin{align*}
\text{Sensitivity} &= \frac{TP}{(TP + FN)} \\
\text{Specificity} &= \frac{TN}{(FP + TN)} \\
\text{Accuracy} &= \frac{(TP + TN)}{(TP + TN + FP + FN)}
\end{align*}
\]
We use specificity to analyse and assess the amount of true positives predicted accurately. Specificity analyses and measure the amount of true negative predicted accurately. Accuracy can be obtained by sum of True Positive and True Negative divided with the total number of instances.

*Example*: No of healthy (True Positive) and sick (True Negative) peoples predicted correctly.

**Figure 4**: Training Datasets Performance

The performance testing has been performed on both training and testing data sets, and we merely used Cleveland heart disease datasets establishes to evaluate our system. And, in last we compared the performance of our system with Neural Network and J48 Decision Tree model using very same data sets. The final result we obtained for training and testing data are shown in figure (5-2a and 5-2b).

**V. Conclusion**

This proposed system “Fuzzy Rule Based Support System” modelled to predict heart disease intelligently and efficiently, and to replace manual efforts. Experts system can be more proficient and fast so it can be more accurate then manual work. Our system modelled to diagnosis and detecting cardiovascular diseases, the system involves two major phases, one that performs classification and diagnosis, the other one that detects the rate of risks of the respiratory diseases. For this system we have used mamdani inference system. In final this system tested and compared with Neural Network and J48 Decision Tree model to check performance of the system.

**VI. Acknowledgement**

This work is supported by National Natural Science Foundation of China under grant No.60970096 and 90818004, the Postdoctoral Science Foundation of China under Grant No.20080440988 and the Natural Science Foundation of Hunan Province, China Under Grant No.09JJ4030.

**References Références Referencias**


