A NOVEL METHODOLOGY FOR DIAGNOSING THE HEART DISEASE USING FUZZY DATABASE

K. Sudhakar[1], M. Manimekalai[2]

[1] *Research Scholar, Department of Computer Application,*
[2] *Head and Director, Department of Computer Application*
[1,2]*Shrimati Indira Gandhi College, Tamilnadu, India*

Abstract

A familiar method used for information storing is using a Database. Sometimes the user's criteria may not be fulfilled due to the presence of vast amount of data in the system of regular database and for the decision making they should be provided with the exact information. In this paper, in the database of the heart disease where there is inaccuracy occurs, accurate information is provided for the users to help them by introducing a medical fuzzy database. The occurrence of uncertainty in the available information for the decision making, the vague values or imprecise in the representation of the data is called Inaccuracy in given data. In this paper, the severity of the patient with the heart disease is diagnosed by introducing a new database system called a fuzzy database management system by utilizing an existing data in the common database systems.

Key Words: *Fuzzy database, inaccurate data, membership function, linguistic representation, SQL.*

---***---

1. INTRODUCTION

In the present scenario, the major role to step forward in the each and every sphere from earth to ocean, earth to sky and earth to planet for the modern civilization is Information. Based on the information, the astounding inventions are discovered at a rapid pace by the people. As a result of that the information obtained should be accurate and it is necessary.

Now a days, the size of the data are decorous large. Therefore, it is a challenge to deal with these large amounts of data stored in database. The traditional database is based on Boolean logic, which attests that the information is either completely true or completely false. In further, it is impossible to compact with imprecise information in classical database management systems. And it is vital to discover expedient ways to store and administer the human perception based data which is often vague and indecisive in traditional database system.

Due to the recent developments in the field of technology, there is no sufficient analysis tools for detecting the concealed relationships among the data and within the health care systems, there is an availability of huge data management tools. As a result of that, almost the medical information is imprecise, uncertain and vague. Through this extorting the correct information from the data is considered an art. It can be whispered to be an art since it is complicated by many features and its solution engrosses accurately all of a human's abilities including intuition and subconscious.

Medical diagnosis is a intricate task that entails the process exactly and professionally. According to the survey taken by World Health Organization, each year 12 million deaths occur due to heart diseases. It becomes one of the major grounds behind deaths in adults. In United States, 50% of deaths crop up due to cardio vascular disease. In USA due to cardiovascular disease, one person dies in every 34 seconds. Likewise, in other developed countries heart disease turned as one of the main reasons in the wake of adult death. Consecutively to decrease the transience rate of cardiovascular disease, it is obligatory for the disease to be diagnosed at the beginning stage.

By containing the fuzzy data management capability in a database is key to be able to store the vague datas. Ignoring vague data management means the risk of losing important information, which may be functional for some applications. A database which supports vague, imprecise and uncertain information is called a fuzzy database. It is supported on fuzzy logic and fuzzy set theory which is introduced by Zadeh [1].

The presented investigation is organized as follows: Section 2 describes related work, Section 3 talks about Expert's Opinion regarding the risk factors of Heart Disease and fuzzy sets, Section 4 deals with the methodology and proposed system which includes Algorithm of the system, Membership functions, Input and Output of the system, SQL (Structured Query Language), several snap shots of the system, System testing and forms, lastly, Section 5 will conclude and discuss future outcomes.

2. LITERATURE SURVEY

There are numerous papers have been published in association to the diagnosis of various diseases.

Shrandhanjali [2] developed a Fuzzy Petri net application for Heart Disease Diagnosis. The rule is associated with transition for certain factors. The fuzzy Petri net is drowning for the rule base and get decision of the disease, truth value proposition is used.

Lavanya et al. [3] designed a Fuzzy rule based inference system for detection and diagnosis of lung cancer. The dataset is used from domain expert with symptoms, stages and treatment facilities to provide an efficient and easy method to diagnose lung cancer.

For the diagnosis of heart disease, a fuzzy expert system was proposed by [4] Adeli et al. The fuzzy logic is used by the developed system. To get the fuzzy values, the crisp values are fuzzified in their system. Fuzzy is also their output and fuzzy values are used by this expert system. To get the output of crisp, the defuzzification of fuzzy output takes place.

In [5] Sony et al. using a weighted associate classifiers, an Effective and Intelligent Heart Disease Prediction System is designed. The backend and front end as MS Access and Java are used by them. Only two consideration are taken for the prediction (i.e. No Heart Disease and Heart Disease).

Neshat et al. [6] developed a Fuzzy Expert System for diagnosis of Liver Disorder. They considered two cases, people with healthy liver, and people with unhealthy liver along with calibration of disease risk intensity measure. The fuzzy inference system is developed in MATLAB software.

Kadhim et al. [7] implemented a fuzzy expert system which was to diagnosis the back pain. The rules were developed by experts and decision sequence is illustrated by a decision tree.

All the works explained above are the developed and existing systems that use Fuzzy Logic to diagnose disease. Most of them are based on initialising a fuzzy inference engine by using MATLAB.

The difference between our work and other work is we erect our system based on the available data existing in the database. As a result, by focusing on database concepts we developed a system that uses SQL on common database management systems such as Microsoft Access.

3. FUZZY SETS OF THE SYSTEM AND RISK FACTORS OF THE HEART DISEASE BY THE EXPERTS OPINION

Here I have presented a very brief explanation of Fuzzy theory in medical application. A crisp property P can be defined $\mu: x \rightarrow \{0, 1\}$ whereas a fuzzy property can be described by $\mu: x \rightarrow [0, 1]$. $\mu(x)$ indicates the degree to which x has the property. To elucidate fuzzy logic in medical science, a simple example illustrated in [8] [9] this is explained below.

For example: in case of fever, the linguistic variable is high fever. If the body temperature(x) is greater than 39 degree Centigrade then in medical concept $\mu(x)$ for high fever is 1 it means x has surely 'High Fever'. If x is less than 38.5 degree centigrade then $\mu(x)$ of medical concept ' High Fever' is 0, which means x surely does not have 'High Fever'. If x is in the interval [38.5 degree Centigrade, 39 degree Centigrade] then x has a property of high fever with some degree (i.e., membership function) between [0, 1].

In this proposed system, the inputs have fuzzy sets according to the series that they fall in. In that case, to achieve more precise result the inputs are connected with the

membership function proliferated by output membership function to breed the rules strength.

As we acquire the union of all rules strength then we have 44 rules. According to the input the rule strength is generated and we take the highest from the union fuzzy set to generate the Heart Disease Condition as desired output.

According to the American Heart Association [10], the chief risk factors of heart diseases are that it drastically increases the risk of heart and blood vessel (cardio vascular). In addition, the risk factors are exposed to, the better the chance of budding a heart disease. Some of the risk factors such as aged people and gender are related to birth and cannot be changed. The Majority of the People at 65 or older died largely due to heart disease in comparison to the other ages. Men have a superior risk of developing a heart disease compared to women. Though, older women have the same risk similar to men. The Cholesterol level is one of the major factors which affected by age and gender.

Blood sugar is another underlying factor that increases the menace of developing heart disease. If blood sugar is not controlled then a high risk remains. However, Blood pressure is another risk factor which augments the heart's workload and roots the heart to be thickened and hardened. This stiffening of the heart shrinks the normal function. Stress, inactive lifestyle, smoking, alcohol, diet and nutrition are other threat factors that manipulate heart disease in the long run.

Hence, in our system input variables are: Age, Gender, Chest Pain, Cholesterol, Blood Pressure, Blood Sugar, Heart Beat, Electrocardiography (ECG), Exercise, Old Peak and Thallium Scan. The Output of the Heart Disease Condition has shown in the following linguistic terms.

4. PROPOSED SYSTEM AND RESEARCH METHODOLOGY

In this section, first the algorithm and methodology are explained and then the snapshots of the system functions and testing are illustrated.

Algorithm:

Step 1: 11 inputs: Age, Gender, Chest Pain, Cholesterol, Blood Pressure, Blood Sugar, Heart Beat, ECG, Exercise, Old Peak, Thallium Scan, and Category of Heart Disease.

Step 2: Output: Heart Disease Condition shown as the linguistic terms.

Step 3: Each input has fuzzy variables.

Step 4: The member function is associated in each fuzzy Variable.

Step 5: The membership function is calculated for each fuzzy variable.

Step 6: On the basis of the fuzzy variable member function, the rules strength is calculated.

Step 7: Heart Disease condition is calculated by taking the maximum of the selected output set as the final result.

4.1   Input, Output and Member Function:

On behalf of the input variable Cholesterol, we deploy Low Density Lipoprotein (LDL). Nevertheless, it is also

promising to use High density Lipoprotein (HDL). In case of Blood Pressure, Systolic Blood Pressure is employed. Membership function is very essential for each fuzzy variable. Also based on the membership function, the rules strength is calculated.

Age: This input consists of four fuzzy sets i.e. Linguistic variable (Very old, Mid, Old, Young). The membership function is associated with them in the each Linguistic variable. The range of the fuzzy sets for age is shown in Table 1.

Table-1: Age Table

Input Field	Range	Linguistic Representation
Age	<38	Young
	33-45	Mid
	40- 58	Old
	5	Very Old
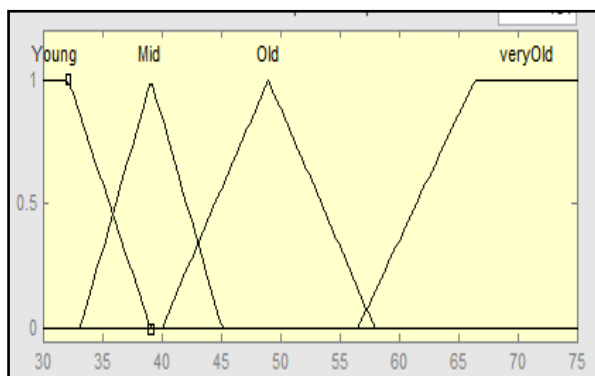
Fig-1: Member Function for Age

Chest Pain: There are four types of chest pain in this input field: Non Angina, Asymptomatic, Atypical Angina and typical Angina. At a time only one type of this chest pain can attack in one patient. The chest pain representation will be, 1= Non Angina, 2 = Asymptomatic, 3= Atypical Angina and 4 = Typical Angina.

Cholesterol: To compare with the other fields, the result is manipulated in this input field. There will be High Density Lipoprotein (HDL) and Low Density Lipoprotein (LDL) Cholesterol can be Low Density Lipoprotein (LDL) and High density Lipoprotein (HDL). LDL alone is considered in our deployed structure. However, it is possible to consider HDL instead of LDL. At a time, only one type can be used. There are four fuzzy sets in this field. The membership function is associated in each fuzzy variable. The range of the fuzzy sets for Cholesterol is given in Table 2.

Table-2: Cholesterol Table

Input Field	Range	Linguistic Representation
Cholesterol	< 197	Low
	188-250	Medium
	217-307	High
	281>	Very High

Gender: There are two representations in this input field Male and Female. Male is represented by 1 and female is indicated by 0.

Blood Pressure: Another chief risk factor is Blood Pressure. Diastolic, Systolic and Mean types are maintained. The systolic blood pressure is considered in our system. Any type of blood pressure can be selected probably. There are four fuzzy sets in this field. In the table 3, the Linguistic variable representation is given. The membership function is calculated based on the range.

Table-3: Blood Pressure Table

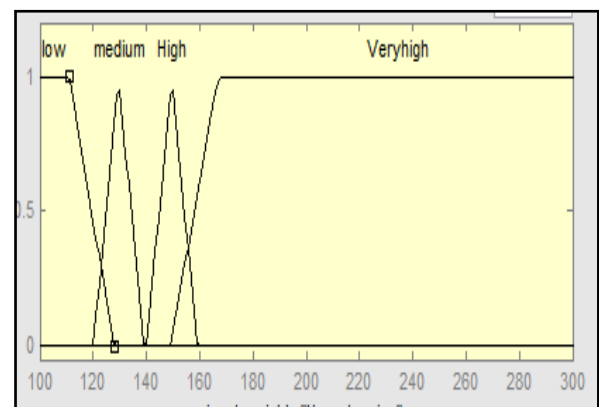| Input Field	Range	Linguistic Representation
Blood Pressure	< 134	Low
	127- 153	Medium
	142-172	High
	154>	Very High

Fig-2: Member Function for Blood Pressure

Heart Rate: This field has three fuzzy sets. Each Linguistic representation is coupled with membership function. The ranges for each linguistic representation are given in Table 4.

Input Field	Range	Fizzy Sets
Heart rate	< 141	Low
	111-194	Medium
	152>	High

Table-4: Fuzzy Set of Heart Rate

Blood Sugar: For the changing of the results, a vital role is acted by this field. The linguistic representations are of two types. Each fuzz variable is allied with membership function based on the range. The ranges of fuzzy sets are given in Table 5.

Input Field	Range	Fizzy Sets
Blood Sugar	>=120	Yes(1)
	<120	No(0)

Table-5: Blood Sugar Table

Electrocardiography (ECG): ECG is measured by numerous waves in a graph paper such as T wave, Q wave, P wave, S wave of electric in pulse of Heart muscle. In general, the waves stay upper bound of the graph. If any of the wave goes down then it is thought as abnormality. In order to develop our system, we assume that S wave and T wave go down to represent the abnormality and this abnormality is named as ST_T abnormal in our designed system. This input field has three fuzzy sets: Normal, ST_T abnormal and Hypertrophy. The ranges for fuzzy sets representation are given in Table 6.

Input Field	Range	Fizzy Sets
Electrocardiography(ECG)	<0.4 0.4 - 1.8 1.8>	Normal ST_T abnormal Hypertrophy

Table- 6 ECG Table
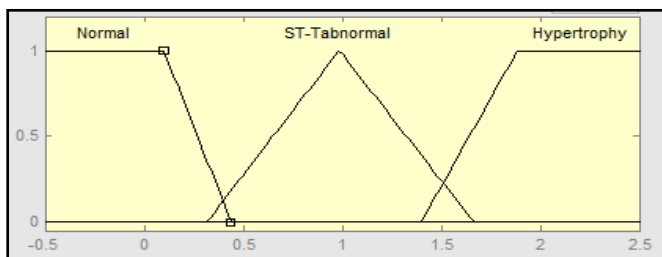
Fig-3: Member Function for ECG

Exercise: The exercise test is whether required for the patient is signified in this field. There is the representation of two fuzzy sets in this input field. The value is entered as 1 for the patient who obligates an exercise test otherwise the value will be 0. 1 is represented as "Yes" and 0 by "No" in the linguistic representation. Each fuzzy set has membership function associated with that

Old Peak: By the exercise which is relative to test, ST depression is carried in this field. The meaning of ST depression is associated to the ECG field. Formerly, in the ECG graph paper of the T wave and S wave of the patient were down. For assuring the existing condition of the S wave and T wave of the ECG, an Old Peak is essential. It has three fuzzy sets representation. Each fuzzy variable is allied with membership function. The range for the fuzzy sets is given in Table 7.

Input Field	Range	Fizzy Sets
Old Peak	<2 1.5 - 4.2 2.5>	Low Risk Terrible

Table- 7 Old Peak Table

Thallium Scan: The heart image redeployment is the Thallium scan. There are three linguistic representations in this input field: Fixed defect, Normal and Reversible Defect. For calculating these representations, the time taken for the appearance of the heart image in the Gamma camera on the screen. The radioactive dye in the body is sensed by this

Gamma camera. In extension of our system we believe that the Thallium scan linguistic representation whenever the heart image appears within 3 hours then it is normal, within 6 hours then it is Reversible defect and when it is appears within 7 hours then it is fixed defect. In the table 8, the Thallium scan linguistic representation is given.

Input Field	Range	Fuzzy sets
	3 6 7	Normal Fixed Defect Reversible Defect

Table- 8 Thallium Scan Table

Output: The output is the presence of Heart disease valued from 0 (no presence i.e. Healthy condition) to 4. If the integer value increases then the heart disease risk increases. We carve up the Output fuzzy sets as {Healthy, Mild, Moderate, Severe, Very Severe}.The ranges and membership function for output variable are presented below:

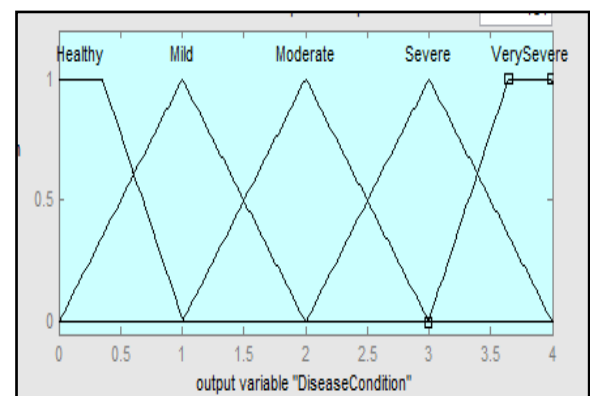| Output Field	Range	Fuzzy sets
Result	< =1 0-2 1- 3 2-4 3>	Healthy Mild Moderate Severe VerySevere

Table- 9: Output Table

Fig- 4: Member function for Output

4.2   SQL (Structure Query Language):

By implementing our proposed system, all the inputs are connected together through SQL and therefore we attain our desired output. We fabricate one Patient table with all inputs and then connect the inputs together by SQL to get the desired output. One SQL from our designed system is given below:

SELECT Patient. Age, IIf([Age]<=29,1,IIf([Age] Between 29 And 38,Format((38-[Age])/9,"#.00"),0)) AS YoungMF,

IIf([Age]=38,1,IIf([Age] Between 33 And 38,Format(([Age]-33)/5,"#.00"),IIf([Age] Between 38 And 45,Format((45-[Age])/7,"#.00"),0))) AS MidMF, IIf([Age]=48,1,IIf([Age] Between 40 And 48,Format(([Age]-40)/8,"#.00"),IIf([Age] Between 48 And 58,Format((58-[Age])/10,"#.00"),0))) AS OldMF, IIf([Age]>=60,1,IIf([Age] Between 52 And 60,Format(([Age]-52)/8,"#.00"),0)) AS VeryOldMF;

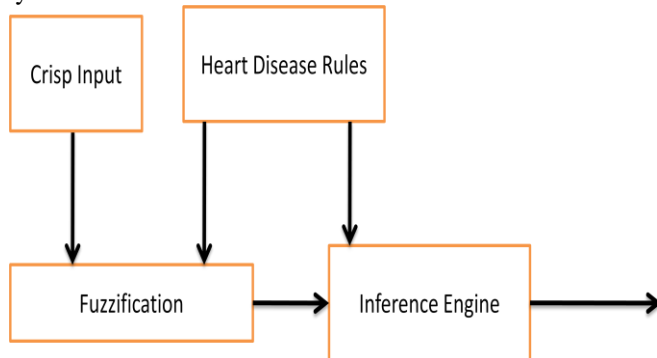The following figure indicates the overall summary of our system:

Fig- 5: Developed Fuzzy Expert Database System

4.3 Snapshots Of The System

Some snap shots for the developed system are shown below:

Fig-6: Input field Age with membership function

4.4 System Testing

We have tested our developed system with the following values and also for the Cleveland Clinic Foundation's processed Heart dataset [11].

Fig-7: Diagnostic of Heart Disease with risk factors

Fig- 8: The Input Form for the Fuzzy Diagnostic of Heart Disease with all the risk factors

5. CONCLUSION

Through the above investigation it is hereby to conclude that there are numerous applications based on fuzzy data analysis. However, building fuzzy inference system based on recent database applications is atypical. In the medical diagnosis, there are huge collections of data available in the regular databases. The problem of using these database is that there are some inaccuracy in data that makes searching for some information impossible. But our proposed fuzzy database system we have deployed the same data with the regular SQL to build fuzzy rules. The major benefits of our developed system are that we can exploit the available data existing in the current database systems for decision making. We utilized the common SQL language, therefore we do not necessitate to use any special tools such as SQLF or FSQL. This investigation provides the perfect platform in developing the data base management. In future, this research ample an opportunity extends our research on using fuzzy database for other diseases like cancer and diabetes.

REFERENCES

[1] Zadeh, L.A, B.1965 "Fuzzy sets", Information and Control, Vol. 8, pp.338- 353.

[2] Shradhanjali, 2012. "Fuzzy Petry Net Application: Heart Disease Diagnosis", International Conference on Computing and Control Engineering.

[3] Lavanya, K., Durai, M.A., N.Ch.Sriman Narayan lyengar, 2011, "Fuzzy Rule Based Inference System for Detection and Diagnosis of Lung Cancer", International journal of Latest Trends in computing, vol- 2, Issue 1, pp.165-169.

[4] Adeli, A., Mehdi.Neshat, 2010."A Fuzzy Expert System for Heart Disease Diagnosis". Proceedings of the International Multi Conference of Engineers and Computer Scientists, Vol I, ISSN 2078-0966.

[5] Soni, J., Ansari, U., Dipesh Sharma, 2011, "Intelligent and Effective Heart Disease Prediction System using Weighted Associative classifiers". International Journal on Computer Science and Engineering, vol 3, No. 6, pp.2385- 2392.

[6] Neshat, M., Yaghobi, M., Naghibi, M.B., Esmaelzadeh, A., 2008, "Fuzzy Expert System Design for Diagnosis of liver Disorders", International

Symposium on Knowledge Acuisition and Modeling, IEEE, pp.252-256.

[7] Kadhim, M., Alam, M., Harleen Kaur, 2011, "Design and Implementation of Fuzzy Expert System for Back pain Diagnosis", International Journal of Innovative Technology & Creative Engineering, Vol. 1, No.9, pp. 16-22.

[8] Klaus-Peter Adlassnig, 1986." Fuzzy Set theory in Medical Diagnosis". IEEE Transactions on Systems, Man and Cybernetics, Vol. Smc-16, No.2, pp. 260-265.

[9] Phuong, N., Vladik Kreinovich, 2001, "Fuzzy Logic and its Applications in Medicine", International Journal of Medical Informatics, Vol- 62, No. 2-3, pp.165-173.

[10] American Heart Association, 2006, Inc."Journal of American Heart Association ", Risk factors and Coronary Heart Disease.

[11] Robert Detrano & M.D & PhD, V.A. Medical center, Long Each and Clevand Clinic Foundation. Available:www.archieve.ics.uci.edu/ml/datasets/Heart +Disease.