Heart Disease Prediction using Evolutionary Rule Learning

Aakash Chauhan
Department of Information Technology
Amity University Uttar Pradesh
aakashchauhan501@gmail.com

Aditya Jain
Department of Information Technology
Amity University Uttar Pradesh
jainaditya224@gmail.com

Purushottam Sharma
Department of Information Technology
Amity University Uttar Pradesh
psharma5@amity.edu

Vikas Deep
Department of Information Technology
Amity University Uttar Pradesh
vikasdeep8@gmail.com

***Abstract*—In modern society, Heart disease is the noteworthy reason for short life. Large population of people depends on the healthcare system so that they can get accurate result in less time. Large amount of data is produced and collected by the healthcare organization on the daily basis. To get intriguing knowledge, data innovation permits to extract the data through automization of processes. Weighted Association Rule is a type of data mining technique used to eliminate the manual task which also helps in extracting the data directly from the electronic records. This will help in decreasing the cost of services and also helps in saving lives. In this paper, we will find the rule to predict patient's risk of having coronary disease. Test results have shown that vast majority of the rules helps in the best prediction of coronary illness.**

Keywords-Heart Disease Prediction System, Association Rule, Heart Disease, Data Mining.

I. INTRODUCTION

The Process of finding knowledge and information from a vast database is known as data mining. To turn large amount of data into useful information, data mining is used [1]. Association rules give the relationship between the items that are present in large database and is a vital way of knowledge representation. As Association rule is becoming one of the most researched area, the database community is now giving more attention to the association rule. Association rule mining was given by A.Swan, T.Imiclinski and R.Agrawal in 1993 [2].

Due to large amount of data and using that data to extract useful information is the major reason that data mining has attracted huge attention in recent years in information industry. The knowledge that we gain from extracting the data from large database can be used in various field such as production control, management, market analysis and health data analysis.

Small blood vessels are responsible for the supply of oxygen and blood to heart and when these blood vessels starts narrowing it causes coronary heart diseases. The flow of the blood starts to slow down as the arteries narrow, which can cause heart attack, chest pain and shortness of breath [3]. As per World Health Organisation (WHO) coronary disease take life of more than 12 million people every year. 1.73 crores people died in 2008 due to coronary disease, and 80% of the death occur are caused by heart disease [4]. Due to heart disease almost 23.6 million people will die in 2030 as estimated by World Health Organisation.

The analysis of coronary illness relies upon clinical information. When the clinical data of a patient is present, then the heart disease prediction system can help in predicting coronary disease accurately. The healthcare industry is collecting data of patients in large amount which can be mined to discover hidden information that can help medical professionals in effective decision making.

There are many reasons for heart disease such as stress, high blood pressure, drug abuse, lack of exercise, food habit, cholesterol, etc. Our blood vessel become weak due to fatty food which can lead to heart disease. The walls in heart become thicker when more pressure is applied to our arteries. As the walls become thick, it can slow down the flow of blood and can also make the block which lead to heart disease [5] [6]. So we are introducing a method for predicting the heart disease. The pattern that appear frequently in a dataset are called frequent pattern. To find interesting patterns from large database, frequent item set play an important part in information mining.

The records of crores of people can be stored and also the information about their treatment. These along with the data mining strategies can help in answer the most important questions which are related to health of a patient [7]. This paper is roused by the perspective and the previously mentioned issues and proposes an arrangement of methodologies for heart disease prediction.

II. RESEARCH FRAMEWORK

In this research paper, we will analyse data mining algorithm for accurately predicting the heart disease which will also benefit analyst and medical practitioners. We will examine on how learning extraction model can be of great help in heart disease expectation. Recently, relationship between different disease and patient is discovered using the concepts of computational intelligence. So we also uses computational intelligence approach [8] [9].

Using rule mining algorithm we will present rule extraction experiment on heart disease data. We start our examination by cognizance the information source and knowing the quality of the data. Then, we find the first bits of learning into the information. Initial step is to pre-process the information from harsh dataset to the last, and getting the data arranged. We then use the data mining techniques on the patient dataset. Then, we study the result and survey the authenticity of result.

III. DATA SET DESCRIPTION

A. Title: Cleveland Database

Creators of Dataset:

- Hungarian Institute of Cardiology. Budapest: Andras Janosi, M.D.
- University Hospital, Basel, Switzerland: Matthias Pfisterer, M.D.

Attribute Information:

Age: Age in years

Thalach: maximum heart rate achieved

Sex: sex (1 = male; 0 = female)

Exang: exercise induced angina (1 = yes; 0 = no)

Cp: chest pain type
-- Value 1: typical angina
-- Value 2: atypical angina
-- Value 3: non-anginal pain
-- Value 4: asymptomatic

Oldpeak: ST depression induced by exercise relative to rest

Trestbps: resting blood pressure

Slope: the slope of the peak exercise ST segment
-- Value 1: upsloping
-- Value 2: flat
-- Value 3: downsloping

Chol: serum cholestoral in mg/dl

Ca: number of major vessels

Fbs: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)

Thal: 3 = normal; 6 = fixed defect; 7 = reversable defect

Restecg: resting electrocardiographic results

Fig 1: Attribute used

This work is performed using java based tool called KEEL. [10]. KEEL is used for the simulation of evolutionary learning as well as soft computing based techniques.

IV. CLASSIFICATION MODEL BY ASSOCIATION RULES

==========

Model generated

==========

Seed=1286082570
Training Accuracy=0.610738255033557
Validation Accuracy=0.6040268456375839
Test Accuracy=0.5337837837837838

V. EVALUATION RESULTS

Set: training

Total successes percentage:

0.612

Successes percentage:

1 - 0.604

2 - 0.621

Confusion matrix

(rows -> real c1ass; co1umns->obtained class):

159	0	1	0	0
40	0	8	6	0
17	0	14	4	0
17	0	9	9	0
5	0	2	6	0

Set: test

Total successes percentage:

0.531

Successes percentage:

1 0.533

2 0.53

Confusion matrix

(rows->real c1ass; co1umns->obtained class):

153	0	3	4	0
47	0	5	2	0
27	0	1	7	0
25	0	6	4	0
10	0	0	3	0

VI. CONCLUSION

Recently, there is a rapid growth in cardiovascular disease in India and other developing countries. The prediction state that coronary disease will be the main cause of death in India by 2020. So, there is a need to minimize its impact. Hence, Heart disease prediction system is proposed to identify the risk of heart disease accurately. In this paper using data mining techniques, we propose a system for heart disease prediction. To generate strong association rules, we have applied frequent pattern growth association mining on patient's dataset. The method will help doctor to explore their data and predict heart disease accurately.

VII. CLASS DISTRIBUTION WITH REPECT TO DIFFERENT ATTRIUTES

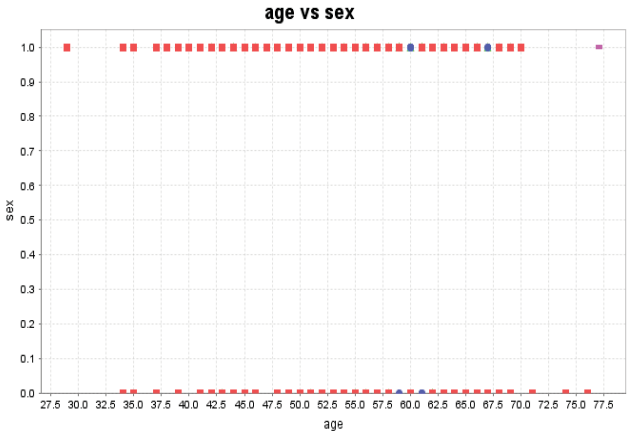

Fig 2: Class distribution with respect to sex

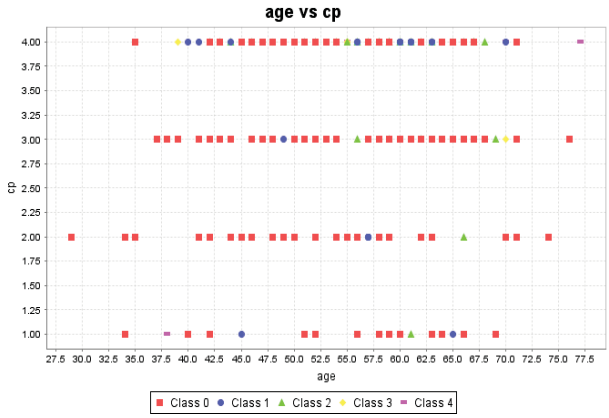

Fig 3: Class distribution in relation with chest pain

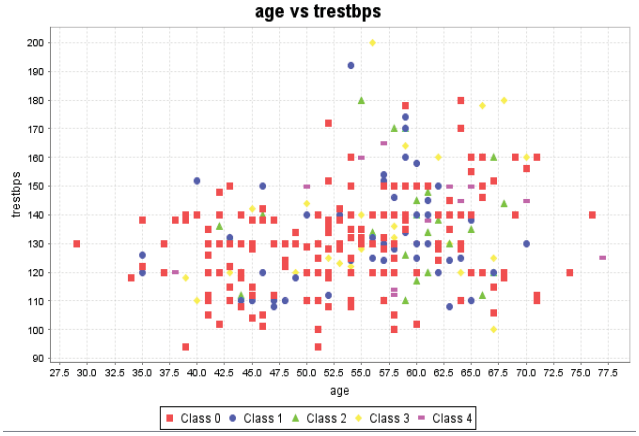

Fig 4: Class distribution in relation with resting blood pressure

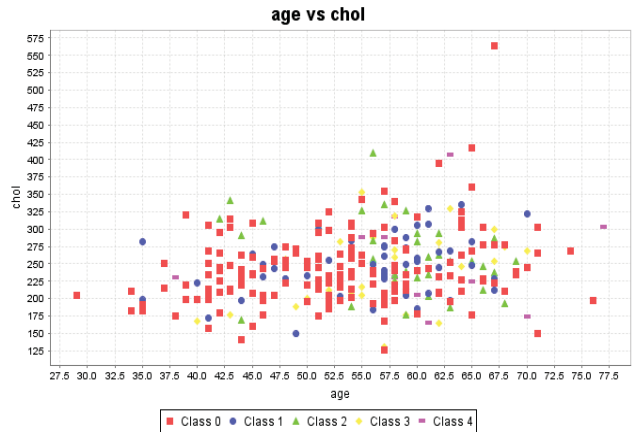

Fig 5: Class distribution with respect to serum cholesterol

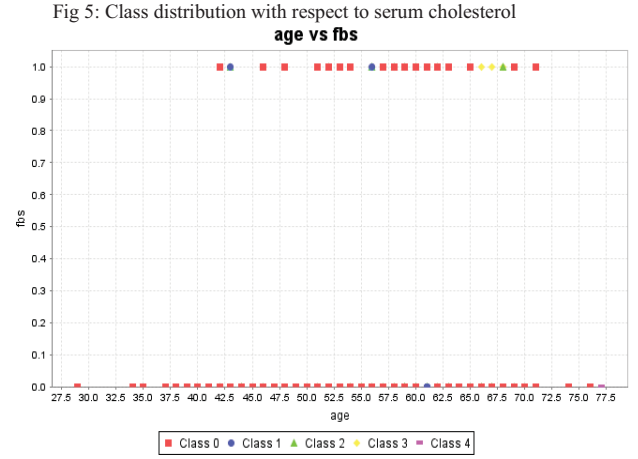

Fig 6: Class distribution in relation with fasting blood sugar

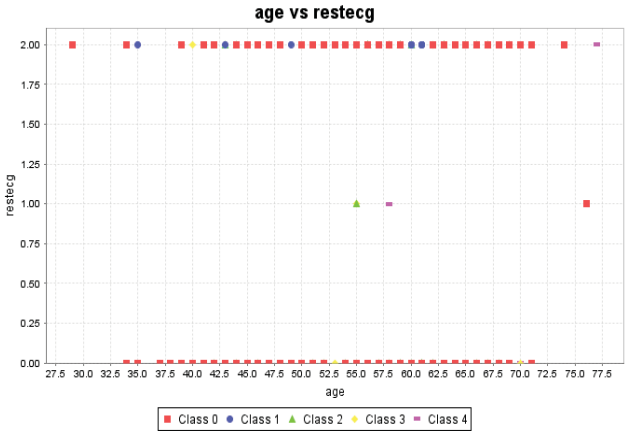

Fig 7: Class distribution in relation with resting electrocadiographic result

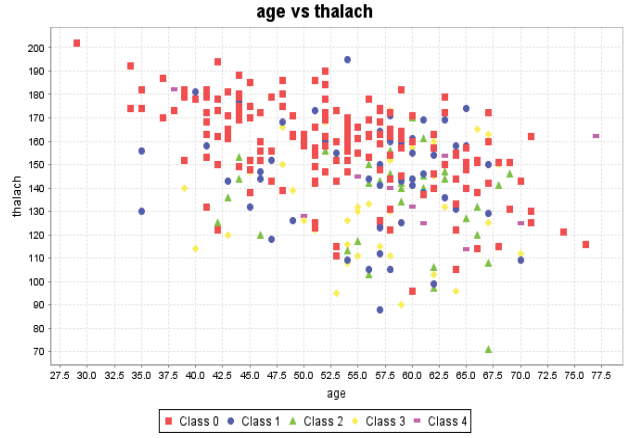

Fig 8: Class distribution in relation with maximum heart rate achieved

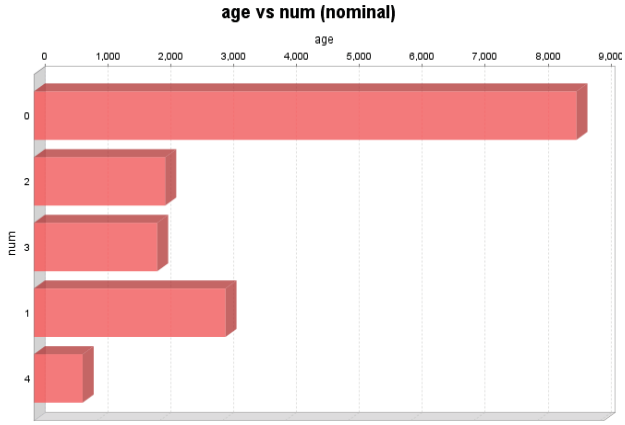

Fig 9: Class distribution in dataset cleveland

REFERENCES

[1] Aswathy Wilson, Gloria Wilson, Likhiya Joy, "Heart Disease Prediction Using the Data mining Techniques", International Journal of Computer Science Trends and Technology, Jan-Feb 2014, pp. 84-88.

[2] Shashi Chhikara, Pururshottam Sharma, "Weighted Association Rule Mining: A Survey", International Journal for Research in Applied Science and Engineering technology, 2014, pp. 85-88.

[3] M.Akhil, Dr.Priti Chandra, Dr.B.L Deekshatulu, "Heart Disease Prediction System using Associative Classification and Genetic Algorithm", International Conference on Emerging Trends in Electrical, Electronics and Communication Technologies-ICECIT, 2012.

[4] Ibrahim Umar Said, Abdullahi Haruna Adam, Dr. Ahmed Baita Garko, "ASSOCIATION RULE MINING ON MEDICAL DATA TO PREDICT HEART DISEASE", International Journal of Science Technology and Management, August 2015, pp. 26-35.

[5] Aswathy Wilson, Gloria Wilson, Likhiya Joy, "Heart Disease Prediction Using the Data mining Techniques", International Journal of Computer Science Trends and Technology, Jan-Feb 2014, pp. 84-88.

[6] Purushottam Sharma, Dr Kanak Saxena, Richa Sharma" Efficient Heart Disease Prediction System using Decision Tree" in IEEE International Conference on Computing Communication and Automation (ICCCA-2015),May 2015.

[7] Feng Tao, fionn Murtagh, Mohsen Farid, "Weighted Association Rule Mining using Weighted Support and significance Framework", 2003.

[8] Irina Tudor, "Association Rule Mining as a Data Mining Technique," 2008, pp. 49-56.

[9] M. Padmavalli, Prof.K.Sreenivasa Rao, "An Efficient Interesting Weighted Association Rule Mining", International Journal of Advanced Research in Computer Science and Software Engineering, October 2013, pp. 1059-1064.

[10] Jes´us Alcal´a-Fdez, Salvador Garc´ıa, Francisco Jos´e Berlanga, Alberto Fern´andez, Luciano S´anchez, M.J. del Jesus and Francisco Herrera, "KEEL: A data mining software tool integrating genetic fuzzy systems", IEEE, 2008, pp.83-88.

[11] Purushottam Sharma, Dr. Kanak Saxena, Richa Sharma, "Diabetes Mellitus Prediction System Evaluation Using C4.5 Rules and Partial Tree", IEEE, 2015.