Engineering

# Sournal or Assearch But AARIPET

### **ORIGINAL RESEARCH PAPER**

## AN IMPROVED CARDIAC RISK PREDICTION SYSTEM USING DATA MINING TECHNIQUES

**KEY WORDS:** Cardio vascular diseases, risk prediction, Bayesian classifier, clustering, foot disease,

Anjana.p		Vedavyasa Institute of Technology, Malappuram, Kerala, Indiav	
S.Kavitha		Vedavyasa Institute of Technology, Malappuram, Kerala, Indiav	
ABSTRACT	Today CardioVascular Disease is one of the atrocious disease found in most of the people irrespective of their ages. This commonly due to the change in their life style which includes their day to day activites, stress, foodetc. This paper presents a card risk prediction system, based on the clinical information of the patients that will be useful for the medical team in the diagnosis heart diseases. Implementation of data mining techniques in the health care is a developing filed which is of higher importance understanding the medical data in depth. The data mining approaches included in this study are clustering and Bayes algorithm. First of all 14 important clinical attributes are selected then the risk level of the patient is predicted using the Bayes algorithm. Then they are cluster into 3 group namely high, medium and normal. The accuracy level was near 80%. Finally whe new attribute which indicated the foot problems was added the accuracy level was increased. Our approaches are efficient predicting the heart disease risk level in a patient. The heart-foot attribute included in this study is a unique approach that can used in the heart diseases classification.		

### I INTRODUCTION

Data mining is a process of using variety of techniques to identify nuggets of information or decision –producing knowledge in the database and using these in such a manner they can be used in areas like the relational databases,,prediction,advance systems and so on. The data mining system self-learns from the previous history of investigated system, formulating and testing hypothesis about rules which system obeys. The knowledge discovered can be used for variety of applications like in healthcare. Presently health industry spawns huge amount of data about disease analysis, patients and so on. Data mining offers a set of approaches to determine hidden patterns from data. The heart is a muscular organ in humans and other animals whose function is to pump blood through the blood vessels. In humans heart is divided into four chambers. The data mining approach performs a important role in heart disease research. According to the survey by the world health organization around 17.6million people die each year from CVD's which is around 31% of all deaths world wide. Cardiovascular diseases(CVD) is a highly significant contributor to loss of quality and quantity all over the world.Early detection and prevention is very important for patients treatment and doctors diagnosis[1]. Hypertension, cholesterol, age, etc are the main risk factors of heart diseases. Hypertension or commonly called as the high blood pressure multiply the chaces of heart disease and stroke, also this can cause a everlasting damage to heart long before one can identify any symptom, this is the main reason why hypertension is called as the silent killer. If the blood pressure ih high than the normal level then keeping track of it is very important task. Diabetes also play a key role in the increasing the heart disease. Obesity is another factor, which is commonly an increasing metabolic disorder. Change in life style which include physical inactivity, family history of heart disease, eating habits alcohol consumption, smokingetc is also a major factor that leads to heart disease. The techniques used for the diagnosis of the heart disease are usually high quality. These data mining techniques have good accuracy.

According to a statistic report of 2012 by the world health organizationone in three adults worldwide have hypertension, which is a major cause of death. The heart disease also includes many functional problems like the heart valve problem or abnormality of valve.different or irregular heart rythms. Thse problems can also lead to heart failure.so the need of an effective heart disease prediction system is complimentary.hence we could say that heart disease prediction system has been the main concern in the scientific researches in modern times.Different types of data mining techniques like KNN Algorithm, Decision tree, Neural network are used for predicting the risk level of heart

www.worldwidejournals.com

diseases [2]. Our work attempts to provide a betterheart disease risk prediction system using Bayesian algorithm and clustering techniques.Bayesian classifiers are geometric classifiers. They are used to predict the class membership probabilities, such as the likelihood that a given tuple belongs to a particular class. Bayesian classification is purely based on Baye's theorem, hence the name.

### **II.LITERATURE REVIEW**

Various types of studies have been prepared to focus on the risk prediction of heart diseases. Many data mining approaches are used for the prediction and analysis of different accuracy level for variety of methods.

Hlaudi Daniel Masethe carried out his experiment using data mining algorithm such as J48, Nadve Bayes in his paper "Prediction of Heart Disease Using Classification Algorithms" [3], for the prediction process. And in his work the result produces an accuracy of 99%. Ms.G.Subbalakshmi et.al developed a "Decision Support in Heart Prediction System(DSHDPS)"[4], by using the data mining techniques like Nadve Bayes. She used the medical attributes like such as age, sex, blood sugar level, blood pressure to identify the relationship between variables in data. This systemcould predict the possibility of patients getting a heart diseases.K.Aravinthan and Dr.MVanitha developed a heart disease prediction system using several data mining algorithms in their paper "A Comparitive Study on Prediction Of Heart Disease Using Cluster And Rank Based Approach" [5]. In this paper they used the UCI heart disease data set for the experiment purpose. Neural network algorithm was used to improve the accuracy. And the experiment produced an accuracy level of 82.57%.

### **III DATA DESCRIPTION**

There are different types of risk factors related with heart disease ,some are controllable ,some are uncontrollable and so on. Understanding the risk factors and categorizing each one according to their priority is a difficult task. So the risk factors were examined from different sources. The data set consist of 13 risk factors. The data of 40 people were gathered by survey. Then using the Bayesian classifier the system was able to predict whether the patient have any risk of heart disease or not. Then the values generated were clustered and grouped into 3 levels namely low risk, medium risk, high risk which will alert the patients about their present risk level. After the prediction additional attribute called the foot disease was added which gave 99% accuracy. It was seen that most of the heart patients have some similarities in the risk factors. The Bayesian classifications are arithmetic classifiers which are capable of predicting the membership possibilities, like the probability of a specific tuple belongs to

### PARIPEX - INDIAN JOURNAL OF RESEARCH

### VOLUME-6 | ISSUE-7 | JULY-2017 | ISSN - 2250-1991 | IF : 5.761 | IC Value : 79.96

particular class. The Bayesian classification do not give effort to **IV..WORK FLOW DESIGN** learn a new explicit decision rule.

The Bayesian classification is completely based on the bayes theorem.

### Bayes theorem:

P(H/X) = P(H/X)P(H)/P(X)

Where, P(H/X) – posterior probability of H conditioned on X P(H) – prior probability of H P(X/H) –posterior probability of X conditioned on H P(X) – prior probability of X

With the help of a Bayesian classifier a system can identify the hidden knowledge connected with disease from the historical records othe patients who has heart disease. One can use a bayes theorem to identify the probability that a planned diagnosis is right, given the observation. According to the naive Bayes classifier the occurrence or any unoccurence of a feature of a class is regarded as independent of presence of any other characteristic. If the dimension of the inputs are very high and if we are expecting an perfect result then the chief Naive Bayes classifier technique is applicable.

A cluster is a grouptof data objects that are relatedwitheach another within the same cluster and are dissimilar to the objects in other clusters. The procedure of grouping a set of physical or abstract objects into classes of alike objects are called as clustering. Basically different types of data are used in cluster analysis which includes interval scaled, binary nominal, ordinal and ratio scaled data. The clustering deals with different types of attributes, noisy data, they often deal with arbitrary shapes etc. The performance of this algorithm largely depends on the value of 'k'., it should be chosen such that it reflects some characteristics of the data set under assessment. The number of clusters should be specified to apply this algorithm, i.e state the value of 'k'. in order to efficiently choose the value of 'k', silhouette measure is used. The value of 'k' with highest average silhouette width is chosen for the data set. It is formulates as:

### $S(I) = b(i)-a(i) / max{a(i).b(i)}$

# The TABLE provides therisk factors and values used in this project.

SL.NO	RISK	VALUES
	FACTORS	
1	Sex	Male(1),female(0)
2	Age	20-34 (-2), 35-50 (-1), 51-60 (0),>61(1)
3	Chest pain type	Typical angina, atypical, asymp, nonangina
4	Resting blood pressure	<120mm-low,120-139(normal) >139- high
5	Serum cholesterol	<200mg- normal,>200mg-high
6	Fasting blood sugar	<120mg/Dl-normal,>120mghigh
7	Resting ECG	Normal(0),left_vent_hyper(1)
8	Max heart rate	<70-low, 70-100-normal, >100-high
9	Exercise induced angina	No(0), yes(1)
10	ST Depression induced by exercise	False(0), True(1)
11	Slope of peak	Flat, up,down
12	No.of major vessels	0,1,2,3
13	Thallium heart scan	Normal, fixed, reversable



### Fig 1:Flow chart of proposed algorithm

First the patient details are recorded which includes their basic details like name, sex ,age in which age is a vital attribute which plays a major role in the risk prediction system. Then the input attributes are read from the user. Based on these input attributes the Bayesian classification is applied and the risk is predicted. Based on this risk percentage the clustering is done to separate the level of risk into different priorities like high risk, low risk and medium risk. Finally the overall risk is predicted and this will be made available to the users. Predicting the risk level of heart disease in humans will help them to take necessary precautions to reduce the chance of occurrence of any heart disease or when the risk level is higher than the normal level , it would help them to follow a healthy diet and precautions moreover they could understand their present level of the disease.

### **V.RESULT AND DISCUSSION**



### Fig:1: major risk factors

### PARIPEX - INDIAN JOURNAL OF RESEARCH

### VOLUME-6 | ISSUE-7 | JULY-2017 | ISSN - 2250-1991 | IF : 5.761 | IC Value : 79.96

Figure1 :shows the major risk factors. These risk factors are arranged according to their priority and similarity shown in most of the heart patients. With the help of Bayesian algorithm the risk value of the patient is predicted. These are the major attributes which would help in detecting the risk of heart disease in humans.



### Fig:2 predicted value

Figure:2 depicts the predicted value. Next an additional attribute foot problem is added.after adding this risk factor we could observe that the prediction was more accurate and the accuracy also increased. Then clustering is done in order to assign these value into three levels mainly low,medium and high risk. This clustering helps a patient to understand the risk level better.

For the past years Heart disease is a leading cause for the increase in the death rates. Identifying heart disease is not an easy method. It need professional doctors and modern methods for the easy identification of heart disease. Most of the people neglect the swelling and pain on their legs. But this is the first symptom of heart disease which most people ignore. In our project this foot problems is considered as the major attribute which help us to provide a better risk rate which is more accurate.



### Fig:3 accuracy level

Figure:3 provide the accuracy level . when the major risk factors were included and the risk level was calculated the accuracy level was near to 90%. But when the additional attribute the foot problems were added the accuracy level was mlore accurate it showed 99% accuracy.

### VI.CONCLUSION

www.worldwidejournals.com

The main inspiration behind this paper was to provide an vision about detecting the heart disease risk prediction using various data mining approaches. As per study mode different authors uses various approaches and number of attributes for their study.hence it is clear that various technologies provide various accuracy levels depending on the number of attributes taken into consideration. Using Bayesian and clustering techniques the risk rate was detected and accuracy level was also achieved. In future by dropping the number of attributes the risk rate and accuracy can be increased

### REFERENCES

- shanxu, Haoyue Shi, XiaohuiDuan, "Cardiovascular risk prediction method based on test analysis and data mining ensamble system", Big Data Analysis(ICBDA),2016.
- [2] K. Sudhakar , Dr.M.Manimekalai, "Study of Heart Diseases Prediction using Data Mining", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 1, pp. 1157-60, jan 2014.
- [3] Hlaudi Daniel Masethe, Mosima Anna Masethe, "Prediction of Heart Disease using Classification Algorithm", Proceedings of the World Congress on Engineering and Computer Science 2014, Vol II WCECS 2014, 22-24 October, 2014, San Francisco, USA.
- [4] Mrs.G.Subbalakshmi , Mr. K. Ramesh," Decision Support in Heart Disease Prediction System using Naive Bayes", Indian Journal of Computer Science and Engineering (IJCSE) ISSN, Vol. 2 No. 2 Apr-May 2011.
- [5] K. Aravinthan, Dr. M. Vanitha, "A Comparative Study on Prediction of Heart Disease using Cluster and Rank based Approach", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5, Issue 2, February 2016.
- [6] ChaitaliS.Dangare, Sulabha S Apte," improved study of heart disease prediction system using data mining classification techniques", international journal of computer applications, vol47-No.10, june 2012.
  [7] Shadab Adam Pttekari and AsmaParveen," Prediction system for heart disease
- [7] Shadab Adam Pttekari and AsmaParveen," Prediction system for heart disease using naive bayes", international journal of advanced computer and mathematical sciences ISSN 2230-9624. Vol3.