DATA MINING TECHNIQUES TO DIAGNOSE HEART DISEASES

[1] **AKSHAYA,** [2] **M. NIVETHA PRIYA,** [3] **Ms. S. KARTHIKA,**
[1, 2, 3] Assistant professor,
[1, 2, 3] M. Sc Software Systems,
[1, 2, 3] Sri Krishna Arts and Science College.

ABSTRACT:-

Heart disease is one of the major cause of gruesome and transience in modern society. Though the medical diagnosis is extremely important, complicated tasks should be performed accurately and efficiently. For heart disease further investigation is needed even though diagnosis and treatment are made. There is a huge data available within the healthcare systems. The availability of variousamounts of medical data implies the necessity for powerful data analysis tools to extract useful knowledge. There is a task of effective analysis tools to identify hidden relationships and trends in data. Knowledge discovery and data mining have identified infinite applications in business and scientific field.

Heart disease diagnosis is one the applications where data mining tools are proving its successful results. In this research paper, to diagnose heart diseases throughdata mining tools such as, support vector machine (SVM), rough set theory, neural networks, association rules, genetic algorithm.

In this paper, it will be examined that decision tree and SVM are the most effective ones for heart disease. So, data mining helps to predict of high and low occurrence of risks in heart diseases.

KEYWORDS: - Heart Disease, data mining, SVM, rough set theory, clustering and association rule.

1. INTRODUCTION
1.1   An overview of data mining

Data mining is the core step, which results in the discovering of hidden but useful knowledge from massive databases. Knowledge discovery in databases is a process that consists of several distinct steps.Knowledge that is discovered can be used by the healthcare organization to improve the quality of service. The discovered knowledge can be implemented by the medical practitioners that reduce the number of drug effect and suggest at cheap cost for therapeutically equivalent alternatives. Data mining provides a user approach to novel and hidden patterns in the data.Theservicequality implies diagnosing patients in correct way and provides treatment that is effective in curing diseases. It is the major challenge to be faced in health care organization to provide a quality services at affordable costs.

The data of health care is monumental. It contains resource management datapatients' data and transformed data. To analyze data, health care organization must have ability. Patients' treatment records are stored and maintained. Data mining technology helps in answering critical questions which relates to health care. Medical diagnosis is important and tough task needed to be carrying in a precise and efficient way. Based on doctors' excellence and experience, clinical decisions

are made instead the knowledge rich data is hidden in its database. A manual decision leads to errors, unwanted biases and costlier medical fees which affect the quality of service for patients. By viewing this, Wu, et al proposed formulaof computer based patient records with clinical decision support reduces all problems. The development of Information Technology results in huge amount of databases and large data in various areas. Data mining defines a process of extracting important information and patterns from large data and also called as knowledge discovery process. The goal of data mining is to fetch the patterns that were unknown previously.
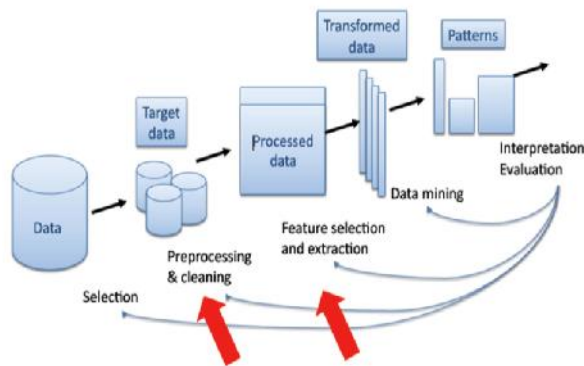

Figure1: KDD process

After patterns being fetched then they can be implemented in making certain decisions for development of businesses. Following three steps involved are:-

- ➢ Exploration
- ➢ Pattern identification
- ➢ Deployment

Exploration: This is the first step of data exploration. In this data is purified and converted into another form, and determined nature of data according to the problem.

Pattern Identification: After exploring data, for the sake of specific variables it is refined and also defined for the formation of pattern identification. Then identify and choose the particularpatterns that predict the best solution.

Deployment: Then patterns that are chosen are deployed for getting desired outcome.

1.2   Causes and impact of heart diseases
WHO reports globally on prevention of cardiovascular disease and control states that cardiovascular disease (CVD) are the greatest causes of death in the world. Even though large probability of CVD can be prevented, they rise mainly because preventive measures are average.

1.2.1   Protect heart health
➢ Intake of tobacco, having an unhealthy diet, and doing physical inactivity leads to heart attacks and heart strokes at a high risk.
➢ Doing physical activity for about 30 minute's daily help in relieving from heart attacks and strokes.
➢ Intake lots of fruits and vegetables and limiting in having salt will prevent from heart attack and strokes.

1.2.2   Cardiovascular diseases
Cardiovascular disease is leads in disordering of the heart and blood vessels. Due to that it results in heart attacks, stroke, hypertension, heart failure and so on. Cardiovascular diseases are caused due to use of tobacco, physical inactivity, an unhealthy diet and consumptionof alcohol. Data mining techniques are used to find these heart diseases as it is the leading disease which causes in death.

2. DATA MINING ALGORITHMS AND TECHNIQUES
2.1. Classification
Classification is the common technique applied in data mining, used to develop a model that is used for classification of the population of records at huge. Forgery, fraud detection and credit risk applications are detected by using this technique. This approach employs decision tree. This process involves classification and also learning. In learning process, the training data are analyzed by classification technique. In

classification process, tested data are estimated for the accuracy. There are types of classification models:

➢ Support Vector Machines (SVM)
➢ Classification Based on Associations

2.2. Clustering

Clustering is defined as identifying objects of similar classes. The use of clustering technique identifies sparse region and sparse dense in objects space is done and can discover distribution pattern.Classification approach is sometimes expensive so clustering is used for preprocessing approach for attribute subset selection. Types of clustering methods:

➢ Model-based technique
➢ Partitioning technique
➢ Divisive technique

3. SURVEY OF LITERATURE (DIFFERENT DATA MINING TECHNIQUES TO FIND OUT HEART DISEASES)

3.1 Decision tree classification algorithm

Heart disease is a term that affects the heart. For clinical decision support systems development researchers have made use of artificial intelligence and data mining techniques. These techniques have achieved high classification accuraciesfor about 77% or higher. The support vector clustering method is used to identify heart disease. Zheng Yao introduced and implemented a new model called R-C4.5 which showed improvement in the attribution selection and dividing models. Gang Kou used data separation techniques for giving privacy in the classification of medical data. Karolis M.A. et.at developed an efficient data mining system for the assessment of heart event in the reduction of coronary heart disease. The analysis of heart was done using C4.5 decision tree algorithm. The highest percentages of accurate classification were75% for the myocardial Infarction (MI) and 25%. In rural India,uneducated and average educated people are having a higher proportion of coronary heart disease due to risk factors of smoking and hypertension. CVD affects all the people at every level.

1.3 Association rule discovery

Association rule is used to improve to predict heart disease. When association rule is implemented on medical data set, it produces numerous rules. Such rules are mostly time required to find that is impractical and medical irrelevant, four constraints that reduces the number of rules: item filtering constrain, attribute grouping constrain, maximum item set size constrain, and antecedent rule filtering constrain. Association rules are applied on the entire data set without validation.The doctors use specificity and sensitivity as two basic statistics that validates result. Sensitivity correctly identifies sick patients, whereas specificity correctly identifies healthy individuals.

3.3 Rough set theory

Rough set theory (RST) is a new artificial intelligent and mathematical technique which was developed by Zdzislaw Pawlak. RST discovers the relationships in data. The identification of relationship in the data is also called as knowledge discovery or data mining. The knowledge discovery produces result in meaning and understandable knowledge from data. RST method is a mathematical tool that manages vagueness, ambiguity and uncertainty from incomplete, inexact and noisy information. Association rule which is applied on real data set contains patients' medical records suffering from heart disease.

CONCLUSION

In this paper, we can examine the classification techniques, clustering, association rule and rough set theory. In these classifications of data mining techniques accuracy among these techniques has discussed. The result shows the difference in error rates. Decision tree and SVM perform

classification more accurately than the other methods. We suggest that the sex, age, blood pressure, chest pain, previous history, personal history, fasting blood sugar, cholesterol, ECG, slope, maximum heart rate, etc. is used as perfect indicators for the prediction of presence of heart disease. In future, we will try to increase in finding even best techniques for diagnosing of heart disease and also curing them and turn India into a healthy country.

REFERENCES

➢ http://www.engpaper.com/data-mining-research-papers-2012-107.htm
➢ http://ijitee.org/attachments/File/v1i4/D0253081412.pdf
➢ http://www.nhlbi.nih.gov/educational/hearttruth/lower-risk/what-is-heart-disease.htm