

Review on Heart Disease Prediction System using Data Mining Techniques

Mrs. Dukitha.M¹, Dr. A.Banumathi ²

Assistant Professor¹, Assistant Professor²,

¹Department of MCA, Er.Perumal Manimekalai College of Engineering, Hosur

²Department of Computer Science, Government Arts College, Karur

ABSTRACT

Data mining is the computerized process of analyzing vast amounts of data and then extracting meaning from the data. Data mining tools predict future trends and enable organizations to make proactive, knowledge-based decisions. Data mining tools can answer business questions that traditionally take a long time to answer. The vast amounts of data involved in predicting heart disease are too complex and voluminous to process and analyze using traditional methods. Data mining provides the methodology and technology to transform these data sets into useful information for decision making. By using data mining techniques, disease can be predicted in less time and with higher accuracy. In this paper, we review several works in which one or more data mining algorithms are used for heart disease prediction. In one work, the hit rate using neural networks. So, prediction using data mining algorithms gives efficient results. The application of data mining techniques to data for the treatment of heart disease can provide as reliable performance as in the diagnosis of heart disease.

Keywords:-Heartdisease, Datamining, Datamining techniques

1. INTRODUCTION

The main objective of our work is to learn about the different data mining techniques used in predicting heart diseases using different data mining tools. Life depends on the efficient working of the heart because heart is an essential part of our body. When the heart is not working properly, it affects other parts of the human body such as the brain, kidneys, etc. Heart disease is a disease that affects the function of the heart. There are a number of factors that increase the risk of heart disease. Nowadays, heart disease is the leading cause of death worldwide. The World Health Organization (WHO) estimates that 12 million people die each year worldwide due to heart disease. In 2008, 17.3 million people died from heart disease. More than 80% of deaths in the world are due to heart disease. WHO estimates that nearly 23.6 million people will die from heart disease by 2030, as described in [10]. Prediction using data mining techniques gives us an accurate result of the disease. IHDP (intelligent heart disease prediction system) can discover and extract hidden knowledge related to heart diseases from a historical database of heart diseases. It can answer complex queries related to heart disease diagnosis, thus



helping analysts and healthcare clinicians to make intelligent clinical decisions, which is not possible with traditional decision support systems. In this paper, we analyze various data mining techniques that are helpful for medical analysts and physicians to accurately diagnose heart diseases and present them in tables.

THE RISK FACTOR FOR HEART DISEASE

Family history of heart disease: - most people know that heart disease can run in families. If someone in the family has heart disease, there is a higher risk of heart attack, stroke and other heart diseases.

Smoking: - Smoking is the leading cause of heart attack, stroke and other peripheral artery disease. Nearly 40% of all people who die from tobacco use die from heart and blood vessel disease. A smoker's risk of suffering a heart attack drops rapidly after just one year of not smoking.

Cholesterol: - Abnormal levels of lipids (fats) in the blood are a risk factor for heart disease. Cholesterol is a soft, waxy substance that is one of the lipids in the bloodstream and in all cells of the body. High triglyceride levels (the most common type of fat in the body) combined with high low-density lipoprotein (LDL) cholesterol levels accelerate atherosclerosis and increase the risk of heart disease.

High Blood Pressure: - High blood pressure, also known as HBP or hypertension, is a widely misunderstood disease. High blood pressure increases the risk of overstretching and injuring the walls of our blood vessels. It also increases the risk of heart attack or stroke, as well as the development of heart failure, kidney failure and peripheral vascular disease.

Obesity: the term obesity is used to describe the health condition of individuals who are significantly over their ideal healthy weight. Obesity increases the risk for health problems such as heart disease, stroke, high blood pressure, diabetes and others.

Physical inactivity: - Physical inactivity is a risk factor for developing coronary heart disease (CAD). Lack of exercise increases the risk of CAD, as it also increases the risk of diabetes and hypertension.

2. LITERATURE REVIEW

The term heart disease refers to a variety of conditions related to the heart. These diseases describe the abnormal health conditions that directly affect the heart and all its parts. Heart disease is a major health problem in today's world. In this paper, we analyze the various data mining techniques that have been introduced in recent years to predict heart diseases. Table 1 shows different data mining techniques used in heart disease diagnosis in different heart disease datasets. In some works, only one technique is used for heart disease diagnosis, such as in Shadab et al [10], Carlos et al [5], etc., while in other research works, more than one data mining technique is used for diagnosis of heart diseases as in Ms. Ishtake et al [3], MA.JABBAR, et al [2], Shantakumar et al [7] etc.

Table 1: The table shows different data mining techniques used in diagnosis of heart diseases for different heart disease datasets.

<i>Author</i>	<i>Technique Used</i>	<i>attributes</i>
Carlosetal	Associationrules	25
Dr. K. UshaRani	Classification	13
	NeuralNetworks	
JesminNahar ,etal	Apriori	14
	Predictive Apriori	
	Tertius	
Lathaetal.	Genetic algorithm	14
	CANFIS	
Majabbaretal	Clustering	14
	Associationrule mining,	
	Sequencenumber,	
Ms.Ishtakeetal.	DecisionTree	15
	NeuralNetwork	
	NaiveBayes	
Nan-Chenet al	(EVAR)	
	Machinelearning	
	Markovblanket	
Olegetal.	Artificial neural network	
	Genetic polymorphisms	
Shadabetal	Naivebayes	15
Shantakumaretal	MAFIA	13
	Clustering	
	K-Means	

3. DATA MINING

Data mining is primarily concerned with the analysis of data, and data mining tools and techniques are used to find patterns from the data set. The main goal of data mining is to find patterns automatically with minimal effort. Data mining is a powerful tool for decision making and predicting future market trends. Data in different areas in different forms. Today, many companies use data mining as a tool to withstand competition in data analysis. By using data mining tools and techniques, different business units can benefit by easily evaluating different market trends and patterns, and producing fast and effective market trend analysis. Data mining is a



very useful tool for diagnosing diseases.

3.1. Techniques used in data mining

A. Association: - Association is one of the most popular data mining techniques. Association involves discovering a pattern based on the relationship between a particular item and other items in the same transaction. For example, association technique is used in heart disease prediction as it shows us the relationship between different attributes used for analysis and picks out the patients with all the risk factors required to predict the disease.

B. Classification: -Classification is a classical data mining technique based on machine learning. Basically, classification is used to classify each element in a dataset into a predefined set of classes or groups. Mathematical techniques such as decision trees, linear programming, neural networks and statistics are used in the classification method.

C. Clustering: -Clustering is a data mining technique that uses automatic techniques to create meaningful or useful clusters of objects with similar characteristics. Unlike classification, clustering technique also defines the classes and classifies the objects into them, whereas classification assigns objects to predefined classes. For example, in predicting heart disease by clustering, we obtain clusters or a list of patients who have the same risk factor. This means that a separate list of patients with high blood sugar and associated risk factors is created and so on.

D. Prediction: - Prediction, as the name suggests, is one of the data mining techniques that reveals relationships between independent variables and relationships between dependent and independent variables. For example, the technique of predictive analysis in sales can be used to predict the profit for the future if we consider sales as an independent variable while profit could be a dependent variable. Based on the historical sales and profit data, we can then create a fitted regression curve that is used to predict profit.

3.2. Comparative statement

The following table provides a comparative representation of various data mining trends from the past to the future, adapted from Venkatadr et al[32].

Datamining Trends	Algorithms/Techniques Employed	Data formats	Computing Resources
Past	Statistical Techniques	Numerical data and structured data stored in traditional databases	Evolution of 4GPL and various related techniques
Present	Statistical, Machine Learning, Artificial Intelligence, Pattern Reorganization Techniques	Heterogeneous data formats include structured, semi structured and unstructured data	High speed networks, High end storage devices and Parallel Distributed computing etc...
Future	Soft Computing techniques like Fuzzy logic, Neural Networks and Genetic Programming	Complex data objects includes high dimensional, high speed data streams, sequence, noise in the time series, graph, Multi instance objects etc.	Multi-agent technologies and Cloud Computing



3.3.Data mining is used in various applications, some of which are listed below

Business Intelligence: business intelligence is a set of theories, methods, architectures, and technologies that transform raw data into meaningful and useful information for business purposes. BI can process enormous amounts of unstructured data to identify, develop, and create new opportunities. BI in simple terms, it makes interpreting voluminous data easy. Taking advantage of new opportunities and implementing an effective strategy can provide a competitive advantage and long-term stability. BI Technologies provide historical, current, and predictive views of business operations.

Student Performance Analysis: - Classification task is used to evaluate student performance and since there are many approaches used for data classification, decision tree method is used here. Information such as attendance, class work, seminars and assignments were collected from the students' management system to predict the performance at the end of the semester. In this paper, the accuracy of decision tree techniques for predicting student performance is investigated.

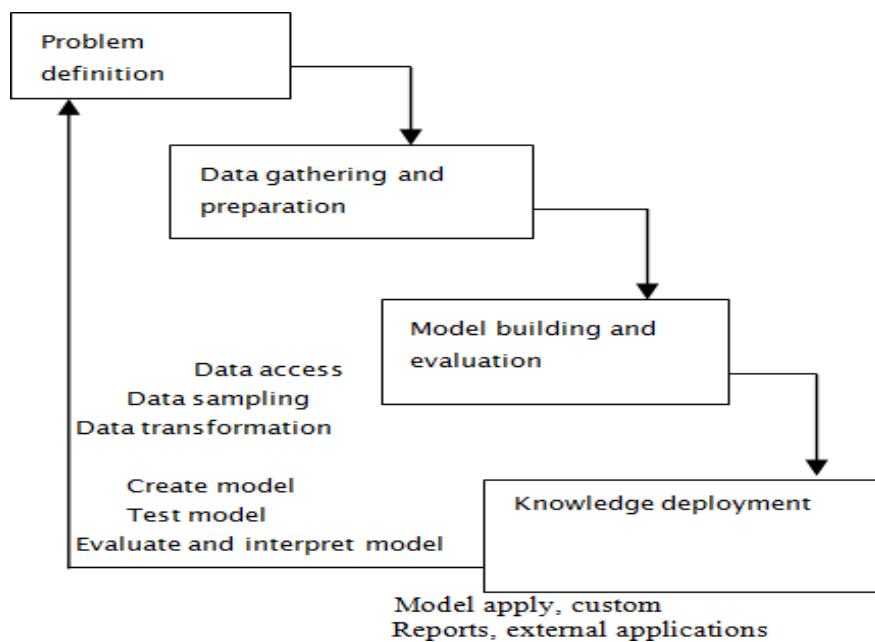
Table 3: Table shows different data mining tools used on heartdisease predictions withaccuracy.

Author	Techniqueused	Data miningtool	Accuracy	Objective
Abhishek et al	J48	Weka 3.6.4	95.56%	HDPSystemUsingDM Techniques
	Naive		92.42%	
	Bayes		94.85%	
	J48			
Chaitraliet al	NeuralNetwork	Weka 3.6.6	100%	PredictionofHD
MonaliEtal	C4.5	WEKA		StudyandAnalysisof DataminingAlgorithmforHealthcareDecisionSupport System
	MultilayerPercep			
	tr on NaïveBayes			
Nidhi etal	NaiveBayes	Weka .6.6	90.74%,99.62%,100%	AnalysisofHDPusingDifferentDMTechniques
	DecisionTrees	TANAGRA	52.33%,52%,45.67%	
		Weka .6.0	86.53%,89%,85.53%	
	Neural	.NET	96.5%,	

3.4. Methodology used in data mining

Data mining is a core component of the Knowledge Discovery Database (KDD). Many people treat data mining as synonymous with KDD because it's an important part of the KDD process. Knowledge discovery as a process is shown in Figure 1 and consists of an iterative sequence of the following steps:

- data cleansing - To remove noise or irrelevant data.
- Data integration - Where multiple data sources can be combined.
- Data selection - Where data relevant to the analysis task is retrieved from the database.
- Data transformation - Data is transformed or consolidated into a form suitable for mining through summary or aggregation operations.
- Data Mining - An essential process in which intelligent methods are applied to extract data patterns.
- Pattern evaluation - Identification of the truly interesting patterns representing knowledge based on some measures of interest.
- Knowledge Presentation - Knowledge representation techniques are used to present the extracted knowledge to the user.



4. CONCLUSION

The aim of our work is to provide a study on different data mining techniques that can be used in automated systems for heart disease prediction. In this work, we define different techniques and data mining classifiers that have been developed in recent years for efficient and effective diagnosis of heart diseases. The analysis shows that different technologies are used in all the contributions, considering a different number of attributes. Thus, the different technologies used have different accuracy. In some papers, neural networks are shown to achieve 100% accuracy in heart disease prediction. On the other hand, it is also shown that the decision tree also



obtained good results with 99.62% accuracy using 15 attributes [6]. Thus, the accuracy of the different technologies used depends on the number of attributes used and the tool used for implementation. Motivated by the increasing mortality rate of patients with heart diseases worldwide and the availability of big data, researchers are using data mining techniques in the diagnosis of heart diseases. Although the application of data mining techniques to assist healthcare professionals in the diagnosis of heart disease has shown some success, less attention has been paid to the use of data mining techniques to determine appropriate treatment for heart disease patients.

REFERENCES

- [1] Mohammad Taha Khan, Dr. Shamimul Qamar, and Laurent F. Massin, A Prototype of Cancer/Heart Disease Prediction Model Using Data Mining, International Journal of Applied Engineering Research, 2018.
- [2] Ma.jabbar, Dr.priiti Chandra, B.L.Deekshatulu, cluster based association rule mining for heart attack prediction, Journal of Theoretical and Applied Information Technology,2019.
- [3] Ms. Ishtake S.H ,Prof. Sanap S.A., "Intelligent Heart Disease Prediction System Using Data Mining Techniques", International J. of Healthcare & Biomedical Research,2019.
- [4] Dr. K. Usha Rani, Analysis of heart disease datasets using neural network, International Journal of Data Mining & Knowledge Management Process, 2017.
- [5] Carlos Ordonez, Edward Omiecinski, Mining Constrained Association Rules to Predict Heart Disease, IEEE. Published in International Conference on Data Mining (ICDM), pp. 433- 440, 2017.
- [6] Nidhi Bhatla Kiran Jyoti, An Analysis of Heart Disease Prediction using Different Data Mining Techniques, International Journal of Engineering Research & Technology (IJERT), 2017.
- [7] Shantakumar B.Patil, Dr.Y.S. Kumaraswamy, Extraction of Significant Patterns from Heart Disease Warehouses for Heart Attack Prediction, (IJCSNS) International Journal of Computer Science and Network Security ,2019.
- [8] Abhishek Taneja, Heart Disease Prediction System Using Data Mining Techniques, Oriental Scientific Publishing Co, India, 2019.
- [9] M. Anbarasi, E. Anupriya, N.ch.s.n.Iyengar, Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm, International Journal of Engineering Science and Technology,2018.
- [10] Miss. Chaitrali S. Dangare, Dr. Mrs. Sulabha S. Apte, A data mining approach for prediction of heart disease using neural networks, international journal of computer engineering and technology, 2016.