



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 11, November 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.625



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com



Identifying Cardiac issues with Data Mining Techniques

Mrs.P.Vanitha¹, Ms.Vijayashree V²

Assistant Professor, Department of Computer Applications (UG), Hindusthan College of Arts & Science, Coimbatore, India¹

Student, Department of Computer Applications (UG), Hindusthan College of Arts & Science, Coimbatore, India²

ABSTRACT: In today's culture, heart disease is a significant contributor to morbidity and mortality. Medical diagnosis is a crucial but challenging job that needs to be completed accurately and effectively. Even though there has been a lot of advancement in the detection and management of heart illness, more research is still required. In order to extract usable knowledge from the vast amounts of medical data that are readily available, strong data analysis tools are required. The healthcare systems have access to a lot of info. Effective analysis tools are needed to unearth Heart hidden relationships and trends in data, though. Data mining and knowledge discovery have many applications in the commercial and scientific fields. For a long time, researchers have been interested in using statistical and data mining methods to enhance data.

KEYWORDS: Data Mining, Heart Disease, SVM, rough sets techniques, association rules & clustering.

I. INTRODUCTION

1.1 Overview of Data Mining

Database knowledge discovery is a well-defined procedure with a number of distinct steps. The main process, known as data mining, enables concealed but valuable knowledge to be uncovered from huge databases. The following is a formal description of knowledge discovery in databases: Data mining, according to [17], is the process of extracting implicit information about data that was previously unknown and possibly useful. A user-oriented method to discovering new and hidden patterns in the data is provided by data mining technology. The healthcare administrators can use the learned information to raise the level of support. Medical professionals can use the newly discovered information to lessen the frequency of negative drug side effects and to propose less expensive therapeutic alternatives. predicting the patient's future actions based on the provided. The provision of high-quality services at reasonable prices is a significant challenge confronting healthcare organisations (hospitals, medical facilities). Correct patient diagnosis and efficient treatment delivery are essential components of quality care. Poor clinical judgement can result in catastrophic outcomes, which is unacceptable. Clinical diagnostic costs must be kept to a minimum by hospitals. By using the proper computer-based information and/or decision support systems, they can accomplish these results. There is a vast amount of health care info. Patient-centric data, resource management data, and modified data are all included. Organizations in the healthcare industry must be able to analyse data. Millions of patient treatment records can be stored and computerized, and data mining techniques may aid in resolving a number of crucial and essential health care-related questions.. The availability of integrated information via the sizable patient repositories has caused a shift in how clinicians, patients, and payers perceive clinical data visualisation from qualitative to more quantitative, with support from all clinical and imaging data. For example, it might now be feasible for doctors to compare the diagnostic data of different patients with the same conditions. Similarly, doctors can confirm their conclusions by comparing them to those of other doctors handling the same situation around the globe [18]. Medical diagnosis is thought of as an important but complex task that needs to be carried out accurately and successfully. The same could be automated, which would be very advantageous. Instead of using the knowledge-rich data buried in the database, clinical choices are frequently made based on doctors' experience and intuition. This practise results in unintended biases, mistakes, and excessive medical expenses, which have an impact on the standard of care given to patients. Clinical decision support and computer-based patient data could be integrated to decrease medical errors, improve patient safety, reduce unwelcome practise variation, and enhance patient outcomes, according to a study by Wu et al. [19]. This suggestion holds promise because tools for data modelling and analysis, such as data mining, have the potential to create a



International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

knowledge-rich environment that can help to considerably raise the calibre of clinical decisions. Large databases and vast amounts of data have been produced in many different fields as a result of the development of information technology. An approach to storing and manipulating this priceless data for future decision-making has emerged as a result of study in databases and information technology. Data mining is the process of removing relevant information and trends from massive amounts of data. Data/pattern analysis, knowledge mining from data, knowledge extraction, and knowledge discovery procedure are other names for it.

Data Mining Techniques to Find Out Heart Diseases: An Overview

Once these patterns are found they can further be used to make certain decisions for development of their businesses. Three steps involved are

- Exploration
- Pattern identification
- Deployment

Data cleaning and transformation are the first steps in data exploration. Next, significant variables and the structure of the data based on the issue are identified. Pattern identification is the second stage after data has been examined, clarified, and defined for the particular variables. Find and select the designs that offer the most reliable predictions.

1.2 Causes and impact of heart diseases

The main causes of death and disability worldwide, according to the WHO report Global Atlas on Cardiovascular Disease Prevention and Control, are cardiovascular diseases (CVDs). Although a significant portion of CVDs are preventable, their rise is primarily due to insufficient preventive steps. 17.3 million plus In 2008, an estimated 17.3 million individuals passed away from CVDs. 23.6 million, or more than 80%, of CVD fatalities occur in low- and middle-income nations. Nearly 23.6 million individuals will pass away from CVDs by 2030.

1.2.1 Protect heart health

- Tobacco use, an unhealthy diet, and physical inactivity increase the risk of heart attacks and strokes.
- Engaging in physical activity for at least 30 minutes every day of the week will help to prevent heart attacks and strokes.
- Eating at least five servings of fruit and vegetables a day, and limiting your salt intake to less than one teaspoon a day, also helps to prevent heart attacks and strokes.

1.2.2 Cardiovascular Diseases (Cvds) Key Facts

Globally, cardiovascular diseases (CVDs) are the leading cause of mortality, killing more people each year than any other cause, including heart attacks, strokes, hypertension, peripheral artery disease, rheumatic heart disease, congenital heart disease, and heart failure. The main contributors to cardiovascular disease are unhealthy eating, tobacco use, physical inactivity, and drink abuse. These three factors—chest pain, stroke, and heart attack—are what trigger heart diseases.[14] The purpose of this research studies is to readily identify heart-related diseases through the use of various data mining techniques for the prevention and identification of these diseases. In the last ten years, heart disease has been the top cause of death worldwide. Numerous academics use statistical Several researchers are assisting medical workers in the diagnosis of heart disease by using statistical and data mining tools. [11]

1.2.3 Cardiovascular Diseases

Cardiovascular disease, which encompasses coronary heart disease CVDs, is brought on by heart and blood vessel disorders.

- According to estimates, 17.3 million deaths worldwide in 2008 were attributable to CVDs, or 30% of all fatalities. Estimates place the causes of these fatalities at 6.2 million strokes and 7.3 million coronary heart disease. Over 80% of deaths from cardiovascular disease (CVD) occur in low- and middle-income nations, and men and women are nearly equally affected by these diseases. Nearly 23.6 million individuals will pass away from CVDs by 2030, primarily from heart disease and stroke. According to projections, these will continue to be the major causes of death.



International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

II. DATAMINING TECHNIQUES

The most widely used data mining method is classification, which uses a set of previously classified examples to create a model that can classify the majority of records. Applications for assessing credit risk and detecting fraud are especially well adapted for this kind of analysis. This method commonly uses classification algorithms based on neural networks or decision trees.

2.1. Classification

Learning and classification are both involved in the data classification procedure. The training data are analysed by a classification programme in learning. Data from classification tests are used to gauge how accurate the guidelines are. The rules can be applied to the new data tuples if the accuracy is deemed appropriate. This would include comprehensive records of both legitimate and fraudulent activities identified record-by-record for a fraud detection application. These pre-classified instances are used by the classifier-training algorithm to establish the parameters necessary for accurate discrimination. The algorithm then incorporates these factors into a classifier model. Types of classification models:

- Classification by decision tree induction
- Bayesian Classification
- Neural Networks
- Support Vector Machines (SVM)
- Classification Based on Associations

2.2. Clustering

Identification of similar classes of items is referred to as clustering. We can further define dense and sparse areas in object space and learn about the general distribution pattern and correlations between data attributes by employing clustering methods. The classification technique can also be used to differentiate between groups or classes of objects, but it is more expensive, so clustering can be used as a preprocessing method before choosing an attribute subset and classifying it. For instance, categorising genes with comparable functionality or grouping customers based on their buying habits. Types of clustering methods

- Partitioning Methods
- Hierarchical Agglomerative(divisive) methods
- Density based methods
- Grid-based methods
- Model-based methods

2.3. Prediction

Prediction can be accomplished using the regression method. The connection between one or more independent variables and dependent variables can be modelled using regression analysis. Independent variables in data mining are characteristics that are already known, whereas reaction variables are what we want to forecast. Unfortunately, not all real-world issues can be predicted. Sales volumes, stock prices, and product failure rates, for instance, are all very challenging to forecast because they may be influenced by intricate interactions among numerous predictor factors. Therefore, it may be required to forecast future values using more sophisticated techniques (such as logistic regression, decision trees, or neural networks). Frequently, the same model categories can be applied to classification and regression. Types of regression methods

- Linear Regression
- Multivariate Linear Regression
- Nonlinear Regression
- Multivariate Nonlinear Regression

2.4. Association rule

In order to discover frequent item set findings among large data sets, association and correlation are typically used. Business decisions like catalogue design, cross-marketing, and customer purchasing behaviour analysis are all aided by these kinds of findings. Algorithms for association rules must be able to produce rules with confidence levels



International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

below one. Although there are a lot of potential Association Rules for a particular dataset, the majority of them are typically of little (if any) value. Types of association rule

- Multilevel association rule
- Multidimensional association rule
- Quantitative association rule

2.5. Neural networks

Each link in a neural network has a weight associated with it. It is made up of connected input/output units. The network adjusts its weights as it learns in order to anticipate the correct class labels for the input tuples. Neural networks have the remarkable ability to extract patterns and identify trends from complex or ambiguous data that are too subtle to be noticed by people or other computer techniques. These are ideal for inputs and outputs with constant values. For instance, rearranging handwritten characters, teaching a computer to read English text, and many other real-world business issues have all been effectively implemented in numerous industries. For instance, rearranging handwritten characters, teaching a computer to read English text, and many other real-world business issues have all been effectively implemented in numerous industries.

III. SURVEY OF LITERATURE (DIFFERENT DATA MINING TECHNIQUES TO FIND OUT HEART DISEASE)

3.1 Decision Tree Classification Algorithm

Any illness that affects the heart can be referred to as having heart disease, also known as coronary artery disease (CAD), coronary heart disease (CHD), or ischemic heart disease (IHD)[4]. There are several studies in the literature that have used data mining and artificial intelligence to create clinical decision support systems. Numerous research on the diagnosis of heart disease have been published to date. These studies used a dataset taken from the UCI machine learning repository to apply various methods to the given problem and reach high classification accuracies of 77% or higher. According to experimental findings, the discriminate function derived from logistic regression had a correct classification accuracy of about 77%. On the Cleveland database, the John Gennari LASSIT conceptual clustering algorithm had a 78.9% accuracy rate.. To detect heart illness, a fuzzy support vector clustering method was employed. For the purpose of diagnosing heart disease, Resul Das developed a methodology that makes use of SAS base software 9.13. In order to increase the effectiveness of attribution selection and partitioning models, Zheng Yao implemented a new model known as R-C4.5. Data separation-based methods were used by Gang Kou partition to protect privacy when classifying medical data.

Coronary Artery Bypass Graft (CABG) and intervention (PCI) [16]. Among Indians, CHD has epidemic dimensions. With the prevalence of CHD increasing, India is going through a fast health transformation. Additionally, compared to other groups, the long-term case fatality rate following acute coronary syndrome is significantly higher in Indians. In the Indian community, there has also been a reversal of socioeconomic gradients for CHD risk factors. In this study, we have discovered a decision tree classifier-based system for automated medical diagnosis of heart disease risk. The North Karelia study² was the first to show the effectiveness of population-based interventions that addressed numerous risk factors for CHDs through lifestyle-related community programmes.

Such steps may in fact be effective in developing nations like India for a variety of reasons. First, Indians are at a greater risk due to their high risk factor levels. The high risk population is more prone to be affected by interventions [8]. In both developed and developing nations, cardiovascular disease (CVD) is the main cause of death and disability. In order to react to the rapidly rising burden of CVD morbidity and mortality in India, a paradigm shift away from the biomedical model is therefore necessary [6]. Rural Indians with less education and less access to education are more likely to have coronary heart disease and to smoke and have elevated blood pressure [5]. Data analysis suggests that a cohort of highly educated doctors who are in the highest salary quintile are at epidemic levels of risk for CVD and stroke [7,10]. People with CVD can be from any socioeconomic level[5,7,10].

3.2. UCI Database Description about Decision Tree Classification

The University of California, Irvine (UCI) archive's database on heart illness is used. The Cleveland Clinic Foundation, Hungarian Institute of Cardiology, V.A. Medical Center, and University Hospital of Switzerland each



International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

contributed four data files to this database. In total, 920 entries are provided. The database's initial raw property count was 76. However, only 13 of these variables—Age, Sex, P, Trstbps, Chol, Fbs, estecg, Thalach, Exang, OldPeak, Slope, Ca, Thal, and Num—are mentioned in any of the published studies.[10]

3.3 Clustering D.M. Technique Using K- Means Algorithms

Clustering is the division of a data set into subsets or the classification of objects into different groups so that each subset's data share a common characteristic, frequently proximity to a predetermined distance metric. Numerous contexts have recognised the clustering issue, and solutions have been found to be helpful in numerous medical applications. By enabling the extraction of numerous suitable features from each of the clusters, clustering the medical data into small groups with meaningful data can support the application of traditional data mining techniques and assist in the finding of patterns. In this technique, the well-known K-Means clustering algorithm was used, one of many clustering methods that are available in the literature. One of the most well-known clustering tools, the k-means algorithm is used in many scientific and commercial uses. K-means divides the data into k unique clusters based on their characteristic values. The characteristic values of data that belong to the same cluster are the same. You must prepare k, the positive integer indicating the number of groups. The following list of stages constitutes a k-means algorithm: Heart illness prediction using the K-Means clustering method 1. The area is filled with K points that represent the data that will be clustered. The main group centurions are indicated by these locations. 2. The group that is next to the centurion is given the statistics. 3. After all the data have been allocated, the locations of all the K centroids are recalculated. 4. Continue repeating steps 2 and 3 until the centroids cease moving. As a consequence, the data are divided into groups so that the metric to be minimised can be thought through. The preprocessed heart disease data is clustered using the Kmeans algorithm with the K values. Clustering is a type of multivariate statistical analysis also known as cluster analysis, unsupervised classification analysis, or numerical taxonomy. K-Means clustering generates a specific number of disjoint, flat (non-hierarchical) clusters. It is well suited to generating globular clusters. The K-Means method is numerical, unsupervised, non-deterministic and iterative.

3.4 K-Means and derivatives

The k-Means algorithm assigns each point to the cluster whose center (also called centroid) is nearest. The center is the average of all the points in the cluster — that is, its coordinates are the arithmetic mean for each dimension separately over all the points in the cluster. Example: The data set has three dimensions and the cluster has two points $X = (x_1, x_2, x_3)$ and $Y = (y_1, y_2, y_3)$. Then Z becomes $Z = (z_1, z_2, z_3)$, where $z_1 = x_1 + y_1$, $z_2 = x_2 + y_2$ and $z_3 = x_3 + y_3$

3.5 Advantages to using this technique

The major benefits of this algorithm are its speed and simplicity, which enable it to process huge datasets. b. K-Means may be computationally quicker than hierarchical clustering when there are a lot of variables. (if K is small). c. If the clusters are globular, K-Means may generate tighter clusters than hierarchical clustering.[3]

3.6 Data Mining Through Genetic Algorithms

We start out with a randomly selected first generation. Every string in this generation is evaluated according to its quality, and a fitness value is assigned. Next, a new generation is produced by applying the reproduction operator.

3.7 Classification via clustering

Clustering is the process of grouping similar elements. This technique may be used as a preprocessing step before feeding the data to the classifying model. The attribute values need to be normalized before clustering to avoid high value attributes dominating the low value attributes. Further, classification is performed based on clustering. Experiments were conducted with Weka 3.6.0 tool. Data set of 909 records with 13 attributes. All attributes are made categorical and inconsistencies are resolved for simplicity. To enhance the prediction of classifiers, genetic search is incorporated. Observations exhibit that the Decision Tree data mining technique outperforms other two data mining techniques after incorporating feature subset selection but with high model construction time. Naïve Bayes performs consistently before and after reduction of attributes with the same model construction time.

3.8 Association Rule Discovery

Association rules represent a promising technique to improve heart disease prediction. Unfortunately, when association rules are applied on a medical data set, they produce an extremely large number of rules. Most of such rules are medically irrelevant and the time required to find them can be impractical., four constraints were proposed to



International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

reduce the number of rules: item filtering, attribute grouping, maximum item set size, and antecedent/consequent rule filtering. When association rules are applied on a medical data set, they produce an extremely large number of rules. Most of such rules are medically irrelevant and the time required to find them can be impractical. A more important issue is that, in general, association rules are mined on the entire data set without validation on an independent sample. To solve these limitations, the author has introduced an algorithm that uses search constraints to reduce the number of rules, searches for association rules on a training set, and finally validates them on an independent test set. Instead of using only Support and confidence, one more parameter i.e. lift have been used as the metrics to evaluate the medical significance and reliability of association rules.

3.9 Rough Set Theory

Decision trees, association rules, decision rules, and sequential reasoning are only a few of the knowledge discovery process's outputs. Rules are the most extensive and interpretable kind of knowledge that can be extracted. Many rules are produced by some rule induction algorithms, such as rough set theory. This huge amount reduces the knowledge's ability to be interpreted. The benefits of rule-based systems will be diminished by their lack of interpretability. The noise redundancy in the input and/or training data sets is the reason of the resulting excessive number of rules. Rule pruning is a technique for reducing the number of rules while keeping the system's quality high. Zdzislaw Pawlak of the Warsaw University of Technology created rough set theory (RST), a relatively recent mathematical and artificial intelligence technique, in the first part of 1980. RST is very helpful for identifying relationships in data. Knowledge discovery or data mining is the process of identifying relationships in data. Understandable and significant knowledge from data is the end outcome of knowledge discovery. The RST approach was developed as a mathematical tool to handle ambiguities, vagueness, and uncertainties resulting from imperfect, imprecise, and noisy information. This method allows for the discovery of rules that are assessed with confidence, support, and lift. For an actual data collection containing the medical records of patients with heart disease, association rules are implemented. Pattern [12, 13]

IV. CONCLUSION

This study studies data mining classification strategies and compares classification performance across them. Among these data mining strategies, classification accuracy has been examined. The outcome demonstrates the variations in mistake rates. Different strategies do, however, differ somewhat from one another. SVM and decision trees do classification more precisely than the other techniques. application of data mining to heart disease According to author name et. al., a major benefit of data mining is that it has a 92.1% 91.0% accuracy rate for cardiac disease. Age, sex, chest pain, blood pressure, personnel history, prior medical history, cholesterol, fasting blood sugar, resting ECG, maximum heart rate, slope, etc. are some factors that may be utilised as accurate indications to determine the presence of heart disease. Additionally, we recommend that data be investigated and confirmed by a group of medical professionals with expertise in cardiac disease. In the future, we'll work to improve the accuracy for patients with heart disease by using more of the many metrics recommended by the experts.

REFERENCES

- [1] Mrs. Bharati M. Ramageri, —Data Mining Techniques And Applications|,Bharati M. Ramageri / Indian Journal of Computer Science and Engineering Vol. 1 No. 4 301-305 [2] Heart disease from
- [2]<http://www.nhlbi.nih.gov/educational/hearttruth/lower-risk/what-is-heart-disease.html>
- [3] Bala Sundar V, Bharathiar, —Development of a Data Clustering Algorithm for Predicting Heart| International Journal of Computer Applications (0975 – 888) Volume 48– No.7, June 2012
- [4] <http://heart-disease.emedtv.com/coronary-artery-disease/coronary-artery-disease.html>
- [5] R. Gupta, V. P. Gupta, and N. S. Ahluwalia, —Educational status, coronary heart disease, and coronary risk factor prevalence in a rural population of India, BMJ. pp 1332–1336, 19 November 1994.
- [6] Panniyammakal Jeemon & K.S. Reddy, —Social determinants of cardiovascular disease outcomes in Indians|, pp 617-622, November 2010. mining techniques.
- [7] A Mathavan, MD, A Chockalingam, PhD, S Chockalingam, BSc, B Bilchik, MD, and V Saini, MD, —Madurai Area Physicians Cardiovascular Health Evaluation Survey (MAPCHES) – an alarming status|, The Canadian Journal of Cardiology; 25(5): 303–308, May 2009.



International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

- [8] Vamadevan S. Ajay & Dorairaj Prabhakaran, —Coronary heart disease in Indians: Implications of the INTERHEART study|, Indian J Med Res 132, pp 561-566, November 2010.
- [9] K.S.Kavitha,K.V. Ramakrishnan, Manoj Kumar Singh “Modeling and design of evolutionary neural network for heart disease detection| IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 5, 1694-0814, September 2010.
- [10] Rajeswari K, Vaithiyanathan V, P. Amirtharaj —Application of Decision Tree Classifiers in Diagnosing Heart Disease using Demographic Datal American Journal of Scientific research ISSN 2301-2005 pp. 77-82 EuroJournals Publishing, 2012.
- [11] Yanwei Xing —Combination Data Mining Methods with New Medical Data to Predicting Outcome of Coronary Heart Disease| IEEE Transactions on Convergence Information Technology, pp(868 – 872), 21-23 Nov. 2007
- [12] Ordonez C,| Association rule discovery with the train and test approach for heart disease prediction|, IEEE Transactions on Information Technology in Biomedicine, P(334 – 343), April 2006
- [13] AbuKhoua, E —Predictive data mining to support clinical decisions: An overview of heart disease prediction systems|, IEEE Transaction on Innovations in Information Technology (IIT), pp(267 - 272) March 2012.
- [14] Srinivas, K.,|Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques|, IEEE Transaction on Computer Science and Education (ICCSE), p(1344 - 1349), 2010.
- [15] Shen, Z., Clarke, M. Jones, R. Alberti, T. —A neural network approach to the detection of coronary artery disease|,IEEE Transaction on Computers in Cardiology, 5-8 Sep 1993.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details