ISSN 2063-5346

HEART DISEASE PREDICTION USING RANDOM FOREST ALGORITHM

A. Piragathi[1], Dr.M. Sakthivel[2], Dr.B. Sujatha[3], Dr.R. Vijayarangan[4]

==========

Article History: Received: : 19.04.2023 Revised: 02.05.2023 Accepted: 10.06.2023

Abstract

Heart disease is a disease that effects on the function of heart. There are number of factors which increases risk of heart disease. At the present days, in the world heart disease is the main cause of deaths. The World Health Organization (WHO) has expected that 12 million deaths occur worldwide, every year due to the heart diseases. Prediction by using data mining techniques gives us accurate result of disease. IHDPS (intelligent heart disease prediction system) can find out and extract hidden knowledge related with heart disease from a historical heart disease database. It can answer complex queries for diagnosing heart disease and thus help healthcare analysts and practitioners to make intelligent clinical decisions which conventional decision support systems cannot. A few kinds of heart disease are cardiovascular diseases, heart attack, coronary heart disease and Stroke. Stroke is a type of heart disease; it is caused by narrowing, blocking, or hardening of the blood vessels that go to the brain or by high blood pressure. System based on the risk factors would not only help medical professionals but also it would give patients a warning about the probable presence of heart disease even before he visits a hospital or goes for costly medical Checkups. Machine learning will help in predicting and making decisions from the large amount of data produced by healthcare industries and hospitals. We have also seen Machine Learning techniques are being used in many fields in different areas. It discovered a new method that will help in finding significant features by applying machine learning techniques such as Random Forest algorithm that results in improving the accuracy in the prediction of cardiovascular disease. The prediction model will contain different types of machine learning algorithms.

Keywords: *Heart Disease, Data Mining, Medical Diagnosis.*

[1] 2nd M.E CSE, Sengunthar Engineering College, Tiruchengode.

[2,3,4] Professor/CSE, Sengunthar Engineering College, Tiruchengode.

[1] piragathi98@gmail.com

DOI: 10.48047/ecb/2023.12.si12.093

1. Introduction

Data Mining

Data mining (the analysis step of the "Knowledge Discovery in Databases" process, or KDD), a field at the intersection of computer science and statistics is the process that attempts to discover patterns in large data sets. It utilizes methods at the intersection of artificial intelligence, machine learning , statistics, and database systems .The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use aside from the raw analysis step, it involves database and data management aspects, data preprocessing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a few analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

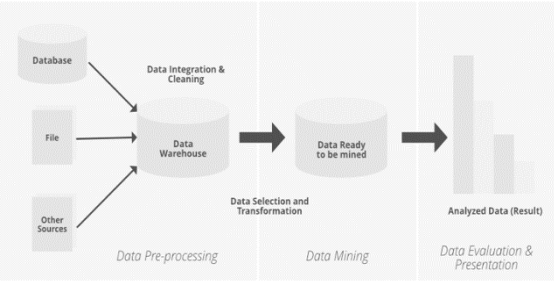

Figure 1.1: Process of Data Mining

Data

Data are any facts, numbers, or text that can be processed by a computer. Today, organizations are accumulating vast and growing amounts of data in different formats and different databases. This includes:

- Operational or transactional data such as, sales, cost, inventory, payroll, and accounting
- Nonoperational data, such as industry sales, forecast data, and macro-economic data
- Meta data - data about the data itself, such as logical database design or data dictionary definitions

Information

The patterns, associations, or relationships among all this data can provide information. For example, analysis of retail point of sale transaction data can yield information on which products are selling and when.

Knowledge

Information can be converted into knowledge about historical patterns and future trends. For example, summary information on retail supermarket sales can be analyzed in light of promotional efforts to provide knowledge of consumer buying behavior. Thus, a manufacturer or retailer could determine which items are most susceptible to promotional efforts.

Data Warehouses

In computing, a data warehouse (DW or DWH) is a database used for reporting and data analysis. It is a central repository of data which is created by integrating data from multiple disparate sources. Data warehouses store current as well as historical data and are commonly used for creating trending reports for senior management reporting such as annual and quarterly comparisons. The data stored in the warehouse are uploaded from the operational systems (such as marketing, sales etc., shown in the figure to the right). The data may pass through an operational data store for additional operations before they are used in the DW for reporting. The typical ETL based data warehouse uses staging, integration, and access layers to house its key functions. The staging layer or staging database stores raw data extracted from each of the disparate source data systems. The integration layer integrates the disparate data sets by transforming the data from the staging layer often storing this transformed data in an operational data store (ODS) database. The integrated data are then moved to yet another

database, often called the data warehouse database, where the data is arranged into hierarchical groups often called dimensions and into facts and aggregate facts. A data warehouse constructed from integrated data source systems does not require ETL, staging databases, or operational data store databases. The integrated data source systems may be considered to be a part of a distributed operational data store layer. Data federation methods or data virtualization methods may be used to access the distributed integrated source data systems to consolidate and aggregate data directly into the data warehouse database tables. Unlike the ETL based data warehouse, the integrated source data systems and the data warehouse are all integrated since there is no transformation of dimensional or reference data. This integrated data warehouse architecture supports the drill down from the aggregate data of the data warehouse to the transactional data of the integrated source data systems. Data warehouses can be subdivided into data marts. Data marts store subsets of data from a warehouse. This definition of the data warehouse focuses on data storage. The main source of the data is cleaned, transformed, catalogued and made available for use by managers and other business professionals for data mining, online analytical processing, market research and decision support However, the means to retrieve and analyse data, to extract, transformant load data, and to manage the data dictionary are also considered essential components of a data warehousing system. Many references to data warehousing use this broader context. Thus, an expanded definition for data warehousing includes business intelligence tools, tools to extract, transform and load data into the repository, and tools to manage and retrieve metadata. Dramatic advances in data capture, processing power, data transmission, and storage capabilities are enabling organizations to integrate their various databases into data warehouses. Data warehousing is defined as a process of centralized data management and retrieval. Data warehousing, like data mining, is a relatively new term although the concept itself has been around for years. Data warehousing represents an ideal vision of maintaining a central repository of all organizational data. Centralization of data is needed to maximize user access and analysis. Dramatic technological advances are making this vision a reality for many companies. And, equally dramatic advances in data analysis software are allowing users to access this data freely. The data analysis software is what supports data mining. It enables these companies to determine relationships among "internal" factors such as price, product positioning, or staff skills, and "external" factors such as economic indicators, competition, and customer demographics. And, it 4 Pre-processing Data mining Data cleaning Evaluation Resourcing Interpretations Problem specification Exploration inaesthetic determine the impact on sales, customer satisfaction, and corporate profits. Finally, it enables them to "drill down" into summary information to view detail transactional data.

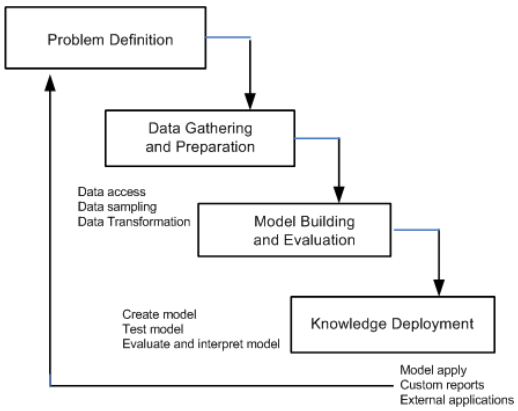

Figure 1.2: Levels of Data Mining

Data Mining Elements

- Extract, transform, and load transaction data onto the data warehouse system.
- Store and manage the data in a multidimensional database system.
- Provide data access to business analysts and information technology professionals.
- Analyze the data by application software.
- Present the data in a useful format, such as a graph or table.

Different Level of Analysis

- Artificial neural networks: Non-linear predictive models that learn through training and resemble biological neural networks in structure.
- Genetic algorithms: Optimization techniques that use processes such as genetic combination, mutation, and natural selection in a design based on the concepts of natural evolution.

- Decision trees: Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID). CART and CHAID are decision tree techniques used for classification of a dataset. They provide a set of rules that you can apply to a new (unclassified) dataset to predict which records will have a given outcome. CART segments a dataset by creating 2-way splits while CHAID segments using chi square tests to create multi-way splits. CART typically requires less data preparation than CHAID.
- Nearest neighbor method: A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most like it in a historical dataset (where k 1). Sometimes called the k-nearest neighbor technique.
- Rule induction: The extraction of useful if-then rules from data based on statistical significance.
- Data visualization: The visual interpretation of complex relationships in multidimensional data. Graphics tools are used to illustrate data relationships.

Data Mining Techniques

There are several major data mining techniques have been developed and used in data mining projects recently including association, classification, clustering, prediction and sequential patterns.

Association

Association is one of the best-known data mining techniques. In association, a pattern is discovered based on a relationship of a particular item on other items in the same transaction. For example, the association technique is used in market basket analysis to identify what products that customers frequently purchase together. Based on this data businesses can have corresponding marketing campaign to sell more products to make more profit.

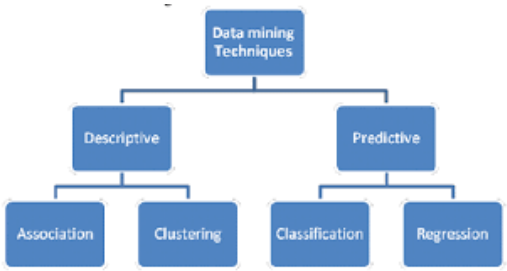

Figure 1.3: Techniques of Data Mining

Classification

Classification is a classic data mining technique based on machine learning. Basically, classification is used to classify each item in a set of data into one of predefined set of classes or groups. Classification method makes use of mathematical techniques such as decision trees, linear programming, neural network and statistics. In classification, make the software that can learn how to classify the data items into groups. For example, can apply classification in application that "given all past records of employees who left the company, predict which current employees are probably to leave in the future." In this case, divide the employee's records into two groups that are "leave" and "stay".

Clustering is a data mining technique that makes meaningful or useful cluster of objects that have similar characteristic using automatic technique. Different from classification, clustering technique also defines the classes and put objects in them, while in classification objects are assigned into predefined classes. To make the concept clearer, can take library as an example. In a library, books have a wide range of topics available. The challenge is how to keep those books in a way that readers can take several books in a specific topic without hassle. Other disciplines Information science Machine learning Data mining Visualization

Goal of the Project

In today's modern world cardiovascular disease is the most lethal one. This disease attacks a person so instantly that it hardly gets any time to get treated with. So, diagnosing patients correctly on timely basis is the most challenging task for the medical fraternity. A wrong diagnosis by the hospital leads to earn a bad name and loosing reputation. At the same time treatment of the said disease is quite high and not affordable by most of the patients particularly in India. The purpose of this paper is to

develop a cost-effective treatment using data mining technologies for facilitating data base decision support system. Almost all the hospitals use some hospital management system to manage healthcare in patients. Unfortunately, most of the systems rarely use the huge clinical data where vital information is hidden. As these systems create huge amount of data in varied forms but this data is seldom visited and remain untapped. So, in this direction lots of efforts are required to make intelligent decisions. The diagnosis of this disease using different features or symptoms is a complex activity. In this System using varied data mining technologies an attempt is made to assist in the diagnosis of the disease in question.

Predictive Data Mining for Medical Diagnosis

An Overview of Heart Disease Prediction The successful application of data mining in highly visible fields like e-business, marketing and retail has led to its application in other industries and sectors. Among these sectors just discovering is healthcare. The healthcare environment is still „information rich" but „knowledge poor". There is a wealth of data available within the healthcare systems. However, there is a lack of effective analysis tools to discover hidden relationships and trends in data. This research paper intends to provide a survey of current techniques of knowledge discovery in databases using data mining techniques that are in use in today's medical research particularly in Heart Disease Prediction. Number of experiments has been conducted to compare the performance of predictive data mining technique on the same dataset and the outcome reveals that Decision Tree outperforms and sometime Bayesian classification is having similar accuracy as of decision tree but other predictive methods like KNN, Neural Networks, Classification based on clustering are not performing well. The second conclusion is that the accuracy of the Decision Tree and Bayesian Classification further improves after applying genetic algorithm to reduce the actual data size to get the optimal subset of attribute sufficient for heart disease prediction.

An Analysis of Heart Disease Prediction using Different Data Mining Techniques

Heart disease is a term that assigns to many medical conditions related to heart. These medical conditions describe the abnormal health conditions that directly influence the heart and all its parts. Heart disease is a major health problem in today's time. This paper aims at analyzing the various data mining techniques introduced in recent years for heart disease prediction. The observations reveal that Neural networks with 15 attributes has outperformed over all other data mining techniques. Another conclusion from the analysis is that decision tree has also shown good accuracy with the help of genetic algorithm and feature subset selection. This paper exhibits the 9 analysis of various data mining techniques which can be helpful for medical analysts or practitioners for accurate heart disease diagnosis. The main methodology used for our work was by examining the publications, journals and reviews in the field of computer science and engineering, data mining and cardiovascular disease in recent times. Association Rule Discovery with the Train and Test Approach for Heart Disease Prediction 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 _Nd 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, we introduce an algorithm that uses search constraints to reduce the number of rules, searches for association rules on a training set and validates them on an independent test set. The medical of discovered rules is evaluated with support, and lift. Association rules are applied on a real data set containing medical records of patients with heart disease. In medical terms, association rules relate heart perfusion measurements and risk factors to the degree of disease in four arteries. Search constraints and test set validation reduce the number of association rules and produce a set of rules with high predictive accuracy. We exhibit important rules with high rules, high lift, or both, that remain valid on the test set on several runs. These rules represent valuable medical knowledge.

HDPS: Heart Disease Prediction System

A complex combination of clinical and pathological data. Because of this complexity, there exists a significant amount of interest among clinical professionals and researchers regarding the efficient and accurate prediction of heart disease. In this paper, we develop a heart disease predict system that can assist medical professionals in predicting heart disease status based on the clinical data of patients. Our approaches include three steps. Firstly, we select 13 important clinical features, i.e., age, sex, chest pain type, treetops, cholesterol, fasting blood sugar, resting ecg max heart rate, exercise induced angina, old peak, slope, number of vessels colored, and then. Secondly, we develop an artificial neural network algorithm for classifying heart disease based on these clinical features. The accuracy of prediction is near 80%. Finally, we develop a user-friendly heart disease predict system (HDPS). \The HDPS system will be consisted of multiple features, including input clinical data section, ROC curve display section, and prediction performance display section (execute time, accuracy, sensitivity, specificity, and predict result). Our approaches are effective in predicting the heart disease of a patient. The HDPS system developed in this study is a novel approach that can be used in the classification of heart disease.

A Data Mining Approach for Prediction of Heart Disease Using Neural Networks

Heart disease diagnosis is a complex task which requires much experience and knowledge. Traditional way of predicting heart disease is doctor's examination or number of medical tests such as ECG, Stress Test, and Heart MRI etc. Nowadays, Health care industry contains huge amount of heath care data, which contains hidden information. This hidden information is useful for making effective decisions. Computer based information along with advanced Data mining techniques are used for appropriate results. Neural network is widely used tool for predicting heart disease diagnosis. In this research paper, a Heart Disease Prediction system (HDPS) is developed using Neural network. The HDPS system predicts the likelihood of patient getting a heart disease. For prediction, the system uses sex, blood pressure, cholesterol like 13 medical parameters. Here two more parameters are added i.e obesity and smoking for better accuracy. From the results, it has been seen that neural network predict heart disease with nearly 100% accuracy.

Automated Diagnosis of Coronary Artery Disease Affected Patients Using LDA, PCA, ICA and Discrete Wavelet Transform

Coronary Artery Disease (CAD) is the narrowing of the blood vessels that supply blood and oxygen to the heart. Electrocardiogram (ECG) is an important cardiac signal representing the 11-sum total of millions of cardiac cell depolarization potentials. It contains important insights into the state of health and nature of the disease afflicting the heart. However, it is very difficult to perceive the subtle changes in ECG signals which indicate a particular type of cardiac abnormality. Hence, we have used the heart rate signals from the ECG for the diagnosis of cardiac health. In this work, we propose a methodology for the automatic detection of normal and coronary artery disease conditions using heart rate signals.

The heart rate signals are decomposed into frequency sub-bands using Discrete Wavelet Transform (DWT). Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Independent Component Analysis (ICA) were applied on the set of DWT coefficients extracted from sub-bands in order to reduce the data dimension.

The selected sets of features were fed into four different classifiers: Support Vector Machine (SVM), Gaussian Mixture Model (GMM), Probabilistic Neural Network (PNN) and K-Nearest Neighbor (KNN). Our results showed that the ICA coupled with GMM classifier combination resulted in highest accuracy of 96.8%, sensitivity of 100% and specificity of 93.7% compared to other data reduction techniques (PCA and LDA) and classifiers. Overall, compared to previous techniques, our proposed strategy is more suitable for diagnosis of CAD with higher accuracy

Effective Diagnosis of Heart Disease Through Neural Networks Ensembles

In the last decades, several tools and various methodologies have been proposed by the researchers for developing effective medical

decision support systems. Moreover, new methodologies and new tools are continued to develop and represent day by day. Diagnosing of the heart disease is one of the important misuses and many researchers investigated to develop intelligent medical decision support systems to improve the ability of the physicians. In this paper, we introduce a methodology which uses SAS base software 9.1.3 for diagnosing of the heart disease. A neural networks ensemble method is in the center of the proposed system. This ensemble-based methods creates new models by combining the posterior probabilities or the predicted values from multiple predecessor models. So, more effective models can be created. We performed experiments with the proposed tool.

We obtained 89.01% classification accuracy from the experiments made on the data taken from Cleveland heart disease database. We also obtained 80.95% and 95.91% sensitivity and specificity values, respectively, in heart disease diagnosis.

Review on Heart Disease Prediction System using Data Mining Techniques

The main objective of our paper is to learn the different techniques of data mining used in prediction of heart disease by using different data mining tools. Life is dependent on efficient working of heart because heart is essential part of our body. If operation of heart is not proper, it will affect the other body parts of human such as brain, kidney etc. Heart disease is a disease that effects on the operation of heart. There are number of factors which increases risk of heart disease. Nowadays, in the world heart disease is the major cause of deaths. The World Health Organization (WHO) has estimated that 12 million deaths occur worldwide, every year due to the heart diseases. In 2008, 17.3 million people died due to heart disease. Over 80% of deaths in world are because of heart disease. WHO estimated by2030, almost 23.6 million people will die due to heart disease as written in [10]. Prediction by using data mining techniques gives us accurate result of disease. IHDPS (intelligent heart disease prediction system) can discover and extract hidden knowledge associated with heart disease from a historical heart disease database. It can answer complex queries for diagnosing heart disease and thus help healthcare analysts and practitioners to make intelligent clinical decisions which traditional decision support systems cannot. In this paper analysis of various data mining techniques given in tables which were used and helpful for medical analysts or practitioners for accurate heart disease diagnosis.

Study and Development of Novel Feature Selection Framework for Heart Disease

Prediction Heart disease prediction is designed to support clinicians in their diagnosis. We proposed a method for classifying the heart disease data. The patient's record is predicted to find if they have symptoms of heart disease through Data mining. It is essential to find the best fit classification algorithm that has greater accuracy on classification in the case of heart disease prediction. Since the data is huge attribute selection method used for reducing the dataset. Then the reduced data is given to the classification. In the Investigation, the hybrid attribute selection method combining CFS and Filter Subset Evaluation gives better accuracy for classification. We also propose a new 13 feature selection method algorithm which is the hybrid method combining CFS and Bayes Theorem. The proposed algorithm provides better accuracy compared to the traditional algorithm and the hybrid Algorithm CFS+Filter Subset Eval.

Heart Disease Prediction System Using Supervised Learning Classifier

Cardiovascular disease remains the biggest cause of deaths worldwide and the Heart Disease Prediction at the early stage is importance. In this paper Supervised Learning Algorithm is adopted for heart disease prediction at the early stage using the patient's medical record is proposed and the results are compared with the known supervised classifier Support Vector Machine (SVM). The information in the patient record is classified using a Cascaded Neural Network (CNN) classifier. In the classification stage 13 attributes are given as input to the CNN classifier to determine the risk of heart disease. The proposed system will provide an aid for the physicians to diagnosis the disease in a more efficient way. The efficiency of the classifier is tested using the records collected from 270 patients. The results show the CNN classifier can predict the

likelihood of patients with heart disease in a more efficient way.

2. Output & Impact

Life is dependent on competent functioning of heart, because heart is necessary part of our body. If function of heart is not suitable, it will affect the other body parts of human such as brain, kidney etc. Heart disease is a disease that effects on the function of heart. There are number of factors which increases risk of heart disease. At the present days, in the world heart disease is the main cause of deaths. The World Health Organization (WHO) has expected that 12 million deaths occur worldwide, every year due to the heart diseases. Prediction by using data mining techniques gives us accurate result of disease. IHDPS (intelligent heart disease prediction system) can find out and extract hidden knowledge related with heart disease from a historical heart disease database. It can answer complex queries for diagnosing heart disease and thus help healthcare analysts and practitioners to make intelligent clinical decisions which conventional decision support systems cannot. A few kinds of heart disease are cardiovascular diseases, heart attack, coronary heart disease and Stroke. Stroke is a type of heart disease; it is caused by narrowing, blocking, or hardening of the blood vessels that go to the brain or by high blood pressure. System based on the risk factors would not only help medical professionals but also it would give patients a warning about the probable presence of heart disease even before he visits a hospital or goes for costly medical Checkups. Hence this system presents a technique for prediction of heart disease. These techniques involve one successful data mining technique named Random Forest algorithm.

3. System Analysis

Existing System

The Existing system using naïve bayes is that it requires a small amount of training data to estimate the parameters. Naive bayes is used to compute posterior probabilities given observations. For example, a patient may be observed to have certain symptoms. Bayes theorem can be used to compute the probability that a proposed diagnosis is correct, given that observation. In simple terms, a naïve Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature. Generally, all machine learning algorithms need to be trained for supervised learning tasks like prediction. Here training means to train them on inputs in such a way that, if later we may test them for unknown inputs (which they have never seen before) for which they may predict based on their learning. According to Naive bayes algorithm first we must convert the data set into a frequency table. Create a frequency table for all the features against the different classes. Likelihood table is created by finding the probabilities. Naïve Bayes Testing Phase will be used to compute posterior probabilities. For example, a patient may be observed to have certain symptoms. Bayes' theorem is used to compute the probability that a proposed diagnosis is correct, given that observation. Naïve Bayes technique recognizes the characteristics of patients with heart disease. It shows the possibility of each 15-input attribute for the predictable state. The main goal of this system is to predict heart disease using data mining technique such as Naive Bayesian Algorithm. Raw hospital data set is used and then preprocessed and transformed the data set. Then apply the data mining technique such as Naïve Bayes algorithm on the transformed data set. After applying the data mining algorithm, heart disease is predicted and then accuracy is calculated.

Drawbacks

If your test data set has a categorical variable of a category that wasn't present in the training data set, the Naive Bayes model will assign it zero probability and won't be able to make any predictions in this regard.

This algorithm is also notorious as a lousy estimator.

Proposed System

The Proposed system using the Random Forest algorithm, It is a supervised learning algorithm that builds multiple decision trees and merges them together to get a more accurate and stable prediction. The Random Forest algorithm operates by constructing multiple decision tress during training phase. The decision of most of the trees is chosen by the random forest as final decision. The “forest” it

builds, is an ensemble of Decision Trees, most of the time trained with the "bagging" method. The general idea of the bagging method is that a combination of learning models increases the overall result. Thus, Random Forest is an ensemble classifier which combines bagging and random selection of features. Here, the algorithm predicts the stage of heart failure of the patient details given as test data through random selection and bagging method of 14 features that are considered in the training data set. For example, a patient may be observed to have certain symptoms. It is used to compute the voting that a proposed diagnosis is correct, given that observation. RF technique recognizes the characteristics of patients with heart disease. It shows the possibility of each 14-input attribute for the predictable state. The main goal of this system is to predict heart disease using data mining technique such as Random Forest Algorithm. Raw hospital data set is used and then preprocessed and transformed the data set. Then apply the data mining technique such as Random Forest algorithm on the transformed data set. After applying the data mining algorithm, heart disease is predicted and then accuracy is calculated.

Advantages

- Random Forest can perform both Classification and Regression tasks.
- It is capable of handling large datasets with high dimensionality.
- It enhances the accuracy of the model and prevents the overfitting issue.

4. System Specification

Hardware Requirements

- System: Pentium IV 2.4 GHz.
- Hard Disk: 40 GB.
- Monitor: 15 VGA Color
- Mouse: Logitech
- RAM: 512 Mb

Software Requirements

- Operating system: Windows XP.
- Coding Language: Java

References

[1] J. L. Scully, ''What is a disease?'' EMBO Rep., vol. 5, no. 7, pp. 650–653, 2004.

[2] Reddy, K. S., Patel, V., Jha, P., Paul, V. K., Kumar, A. S., Dandona, L., & Lancet India Group for Universal Healthcare. (2011). Towards achievement of universal health care in India by 2020: a call to action. The Lancet, 377(9767), 760-768.

[3] Steen, V. D., & Medsger, T. A. (2007). Changes in causes of death in systemic sclerosis, 1972– 2002. Annals of the rheumatic diseases, 66(7), 940-944.

[4] Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur Heart J. 2016;37(27):2129–200.

[5] A.Hussain , Aljaaf AJ, Al-Jumeily D, , Dawson T, Fergus P, AlJumaily M. Predicting the likelihood of heart failure with a multi level risk assessment using decision tree. In: 2015 Third International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE); 2015. p. 101–106. IEEE.

[6] P. Croft, D. G. Altman, and J. J. Deeks, ''The science of clinical practice: Disease diagnosis or patient prognosis? Evidence about 'what is likely to happen' should shape clinical practice,'' BMC Med., vol. 13, no. 1, p. 20, 2015.

[7] E. Choi, M. T. Bahadori, A. Schuetz, W. F. Stewart, and J. Sun, ''Doctor ai: Predicting clinical events via recurrent neural networks,'' in Proc. Mach. Learn. Healthcare Conf., 2016, pp. 301–318.

[8] Srinivas, K., Rani, B. K., & Govrdhan, A. (2010). Applications of data mining techniques in healthcare and prediction of heart attacks. International Journal on Computer Science and Engineering (IJCSE), 2(02), 250- 255.

[9] Kaur, S., Singla, J., Nkenyereye, L., Jha, S., Prashar, D., Joshi, G. P., & Islam, S. R. (2020). Medical diagnostic systems using artificial intelligence (ai) algorithms: Principles and perspectives. IEEE Access, 8, 228049- 228069.

[10] Ntiloudi, D., Giannakoulas, G., Parcharidou, D., Panagiotidis, T., Gatzoulis, M. A., & Karvounis, H. (2016). Adult congenital heart disease: a paradigm of epidemiological change. International journal of cardiology, 218, 269-274.

[11] Yahaya, L., Oye, N. D., & Garba, E. J. (2020). A comprehensive review on heart disease prediction using data mining and machine learning techniques. American Journal of Artificial Intelligence, 4(1), 20-29.

[12] Olaronke, I., & Oluwaseun, O. (2016, December). Big data in healthcare: Prospects, challenges and resolutions. In 2016 Future technologies conference (FTC) (pp. 1152-1157). IEEE.

[13] Sakthivel M., Dr. VenKatesan M., Router Group selection using BAT Algorithm with Bayesian Filter (BABF): An Efficient Trajectory Error Detection. Asian Journal of Research in Social Sciences and Humanities, Year: 2017, Volume : 7, Issue : 1.