Survey on heart disease prediction system based on data mining techniques

M. Thiyagaraj, G. Suseendran

[1]Ph.D. Research Scholar, [2]Assistant Professor
Department of Information and Technology, School of Computing Sciences, Vels University, Chennai, India
mthyagaraj09@gmail.com, suseendar_1234@yahoo.co.in

Abstract

Objectives: To be familiar with the kinds of coronary illness, and information mining procedures to fight them.
Methods/Statistical analysis: To handle this, data mining concepts and techniques used were discussed to discover hidden patterns from medical domain.
Findings: The purpose of predictions in data mining is to discover trends in patient data through patterns generation to improve the health strategy. The algorithms presented here are with a specific end goal to anticipate the coronary illness which includes some constraint.
Keywords: Data mining, CVD diseases, Disease prediction.

1. Introduction

Epidemiological registers a key catalyst for observing and investigating wellbeing and social conditions may even be used for research, assessment, arranging and different purposes to examine and foresee the wellbeing technique of a people.

Cardiovascular maladies are turning into a main source of death rates in India. A fourth of all mortality brings about CVD. Ischemic coronary illness and stroke are the significant causes and are in charge of more than 80 percent of passing. The Global Burden of Disease study evaluates that the age-institutionalized CVD demise rate of 272 for every 100 000 populace in India which is higher than the worldwide normal of 235 for each 100 000 in populace [1]. A few parts of the CVD pestilence in India are specific reasons for concern, including its quickened development, the early time of sickness onset in the populace, and the high case casualty rate. What's more, people from lower financial foundations may not every now and again get ideal treatment, along these lines prompting poor results. Advancement procedures, for example, the detailing in view of compelling execution of confirmation based strategy, fortification of wellbeing frameworks, anticipation, early location, and treatment with the utilization of both ordinary and inventive methods must be completed. A few continuous research identified with group based reviews are trying these methodologies all in all [2].

Figure 1. Report of coronary artery disease

Cardio Vascular Disease (CVD) incorporates coronary heart, cerebrovascular (stroke), hypertensive heart, congenital heart, peripheral artery, rheumatic heart, inflammatory heart disease. Change in life styles is one of the major causes for CVD. An attempt to exploit knowledge, experience and clinical screening of patients to diagnose or recognize heart attacks in health sectors helps to predict diseases through data mining techniques. The goal of this research is to formulate the predictive system using the patterns generated using medical databases. The patient database involves the attributes such as Patient Identification Number (filled with dummy values), Gender, Cardiogram, Age, Chest Pain, Blood Pressure Level, Heart Rate, Cholesterol, Smoking, levels of alcohol and Blood Sugar. The purpose of predictions in data mining is to discover trends in patient data through patterns generation to improve the health strategy. Figure 1 represents the report of coronary artery disease.

The rest of this paper is presented as follows: Section 2 describes the literature about existing survey of heart diseases and hybrid intelligent techniques. Section 3 describes types of heart diseases. Section 4 and 5 describes various data mining and used for the prediction of heart disease and Section 6 involves conclusion.

2. Methodology

An algorithm is being presented to lessen the quantity of principles with a specific end goal to anticipate the coronary illness includes some constraint. The calculation scans for affiliation administers in a preparation set and approves on any testing set [3]. The affiliation rules focuses on heart perfusion estimations and hazard components for each of the four particular supply routes. Affiliation guidelines were produced on a genuine informational index with patient history in regards to coronary illness to yield high exactness rate. Be that as it may, it delivers huge number of tenets for two diverse dataset and approves rules. To channel such guidelines on the test information, Bolster and Certainty have been utilized. Certainty assumes an indispensable part to approve rules. Fundamental standards created for medicinal area stay legitimate in a few individual preparing and testing procedure. The proposed calculation handles the issue of vast number of principles and appropriate approval of guidelines backings as well.

Genetic Algorithms Support Vector Machines (GA-SVM) which are consolidated to create a half breed approach were proposed in light of a wrapper approach. Hereditary calculation scans for the best property in the dataset and SVM characterizes the examples from the decreased dataset frames the wrapper approach gave 5 set of information have been taken from UCI vault to assess the proposed calculation. The unwavering quality of proposed GA-SVM crossover calculation enhances the normal wellness of a chromosome populace and the substitutes of weaker chromosomes in view of the connection measure to give better order result. This examination decides GA-SVM cross breed as a decent classifier and precision rate of 76% roughly. The heartiness of the GA-SVM half and half in the multi-class area indicates precision of around 84.07% [4].

Approach have been proposed to finding of heart ailments the three distinctive control mining calculations, for example, Apriori, Predictive Apriori and Tertius were utilized to accentuate run extraction on coronary illness database. Cleveland dataset has been utilized for this reason to maintain a strategic distance from protection issue in regards to restorative database. Since determination is costly and tedious with likely blunders, the patient information indicates that the females have less possibility of coronary illness than guys. Coronary illness for both men and ladies was existed because of the nearness of activity initiated angina and components identified with trunk torment. The ECG of men and ladies was distinctive. The hazard variables of ECG for ladies appear to be typical/hyper and even incline. The individual lead communicating for ECG is being hyper for men where slant being up, number of hued vessels being zero and old pinnacle being not exactly or equivalent to 0.56 contrasting and the solid status of men[5].

Kernel F-score Feature Selection is introduced to perform include determination as a pre-preparing venture in the characterization of therapeutic database. The proposed KFFS technique includes two stages. One to change the components of medicinal datasets to bit space stage by methods for Linear or RBF capacities [6]. Utilizing F-score equation, the therapeutic datasets have been ascertained by changing piece capacities from non-directly detachable medicinal dataset to a straightly distinguishable element space. To approve the execution of KFFS technique the UCI database were utilized. The ROC bend qualities and AUC values acquired from Least Square Support Vector Machine and Artificial Neural Network classifiers without KFFS strategy on the order of coronary illness. Correlation made on master framework in light of the arrangement utilized as a part of therapeutic database [7].

Grouping based preprocessing is completed with Principal Component Analysis alongside differential advancement classifier to determination and foresee coronary illness utilizing Electronic Medical Record. Coronary illness information would have statistic properties, clinical side effects, and its discoveries, lab test comes about

identified with particular Electro-Cardio Graphy indicating angina and coronary localized necrosis. Results were assessed for four diverse heart dataset and furthermore with joined dataset indicates high order exactness rate even without preprocessing. The outcome speaks to that predisposing of information before arrangement may not backings to expand information dimensionality, but rather gives better precision. The individual dataset outflank with roughly 94.5% exactness and consolidated datasets indicates 82% precision when contrasted and alternate classifiers separately [8].

Restorative choice emotionally supportive network has been actualized with different instruments and systems to build up a compelling finding for coronary illness to enhance the capacity of the doctors. Proposed technique utilizes Statically Analysis System construct programming situated in light of neural systems gathering model empowers an expansion in speculation execution by consolidating a few individual neural systems prepare on a similar stage. SAS base programming 9.1.3 backings all assignments inside in a solitary, coordinated arrangement in this way giving adaptability to advance compelling joint efforts. Exploratory outcome gets 89.01% exactness, 80.95% affectability and 95.91% [9].

A genuine coded GA based framework could choose the basic clinical components critical to conclusion heart infections for five noteworthy heart ailments that are hypertension, coronary illness, rheumatic, valvular coronary illness, incessant pulmonale and innate coronary illness. It utilizes coronary illness information with 352 cases and for each case 40 symptomatic elements were recorded. Among 352 instances of coronary illness datasets 24 basic indicative elements have been recognized and their comparing conclusion weights for supporting or precluding the determination from claiming every coronary illness have been resolved with high exactness rate [10].

2.1. Heart attack prediction using intelligent techniques

In this segment the part of astute procedure, for example, neural system for coronary illness expectation is clarified beneath:

Wise Heart Disease Prediction System have been created utilizing CANFIS and Genetic Algorithm. Versatile based fluffy sources of info are adjusted with a secluded neural system to quickly and precisely rough complex capacities. The CANFIS show joins with the neural system versatile abilities and fluffy rationale quantitative approach then coordinated with hereditary calculation to analysis the impact of sickness. Coactive neuro-fluffy derivation framework demonstrates displays great preparing execution as far as precision. Coronary illness Dataset was taken from UCI Repository. Proposed Coactive Neuro-fluffy model is reliable and vigorous in recognizing nonlinear relationship and mapping between the diverse traits [11].

Coronary illness Prediction System has been instituted utilizing information mining strategies to extricate shrouded data to misuse canny therapeutic finding utilizing recorded heart database. To build up the framework, properties, for example, sex, circulatory strain, cholesterol, and so on like other 13 input traits are utilized. To get prescient outcomes two more qualities, for example, heftiness and smoking must be considered. A Multi-layer Perceptron Neural Networks makes sense of an arrangement of information with an arrangement of suitable one. It comprises of 3 layers like information layer, concealed layer and yield layer yet there is an association between each layer and weights are relegated to every association. The essential capacity of neurons of information layer is to separation enter into neurons in concealed layer. The dataset comprises of absolutely 573 records of coronary illness database. The aggregate records are partitioned into two datasets for both preparing and testing. Preparing dataset comprises of 303 records and testing with 270 records. The Replace Missing Values channel filters all records and replaces missing qualities with mean mode technique known as Data Preprocessing. After pre-preparing the information, Neural Network is utilized for arrangement [12].

The Cerebrovascular illness has been characterized utilizing simulated neural system with back proliferation mistake technique. It empowers sustain forward nets with at least one layers of hubs between the information and yield hubs [14]. These extra layers contain concealed units or hubs however not specifically associated with information and yield hubs. The neural system was prepared utilizing back proliferation calculation alongside sigmoid capacity on one concealed layer for 16 input properties. Prescient models were utilized as a part of assortment of utilizations for analysis purposes. Dataset (100 records = 60 guys and 40 females) were gathered from government medicinal fields. The info values acquired from the records includes input factors in the information layer with each of the 16 hubs. The neural system weights were introduced haphazardly and scope of weights was observed to be -0.5 and 0.5 and the learning rate was set to be 0.1 and 0.9 to gauge the preparation, approval and speculation exactness [13].

3. Types of heart diseases

Figure 2. Types of Heart Diseases
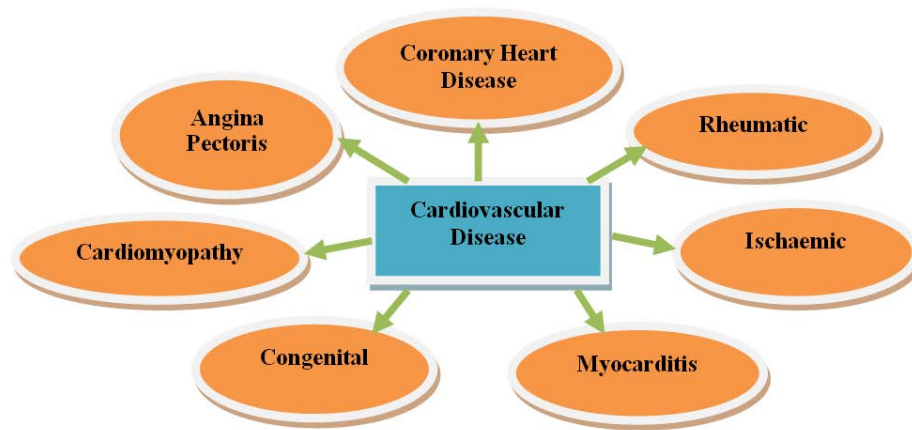
Different sorts of coronary illness influences diverse segments of the heart regularly named as cardio. Figure 2 indicates the various types of cardiovascular maladies in India. All sort of heart maladies goes under the classification of cardiovascular illnesses. A few sorts of Heart infections are as per the following.

3.1. Coronary heart disease (CAD)

Real hazard variables of Coronary Heart Disease are hypertension, high blood cholesterol, expending tobacco, lopsided eating routine, no physical action, diabetes, propelling age, legacy air. Other hazard components include neediness, uneducation, wretchedness, and so forth. To deal with Coronary Heart Disease, three ascription strategies to be specific Artificial Neural Network with Rough Set Theory, k-Nearest Neighbor and Concept Most Common Attribute Value Filling were connected on UCI coronary illness datasets. Missing characteristics are overseen by looking at the three ascription techniques and principles are created from the three datasets utilizing Rough Set Theory. Bolster separating was connected on three arrangements of produced standards where UCI coronary illness datasets ANNRST could be considered as the best technique [14].

Molecule Swarm Optimization is additionally used to determination of CHD utilizing Cleveland and Hungarian Heart Disease datasets [15]. It incorporates four phases, for example, attribution of missing information, choice tree acceptance and manage extraction from ascribed dataset, utilizing fluffy enrollment works, the fresh guidelines were changed into fluffy standards lastly fluffy participation capacities were tuned by PSO. The produced Fuzzy Expert System based standards create translation for analysis of coronary illness. The proposed approach can translate the choices made by the FES by giving 93.27% precision when contrasted with other existing methodology.

Half breed motor has been exhibited to speak to hazard evaluation issue in coronary illness which was outlined utilizing proof and fluffy set hypotheses [16]. Additionally, indistinct character and instability demonstrating issues were grouped under two stage. The issues information were spoken to with fluffy sets and fluffy standards in first stage and finishes with fluffy impedance rules. The frameworks issue was ascertained with the assistance of conviction and sensible capacity from existing stage result which gives conviction interim as definite yield. It gives clear data on hazy character and vulnerability issues in CHD analysis chance evaluation and data combination indicates precision rate of 91.58%.

An essential hazard element of coronary illness is of sort 2 diabetes [18]. It is accepted that diabetes hereditary defenselessness variations brings about high CHD hazard where the individual and combined impact of 38 across the board hereditary variations announced before to be aligned with sort 2 diabetes on danger of episode CHD in 20,467 members of the European Prospective Investigation into Cancer and Nutrition (EPIC) with 2,190 of 20,467 result participants had an occurrence CHD occasion amid a mean record of 10.7 years [17].

3.2. Angina pectoris

It identifies with trunk torment happens because of poor supply of blood to the heart known as angina. It gives a notice flag of heart assault and results in trunk torment at normal interims draw out for few moments or minutes.

4. Congestive heart failure

Coronary illness expectation technique has been produced utilizing information mining and machine learning approach. Tolerant with or without a particular ailment was grouped utilizing characterization trees [18]. For the benefit of constrained precision in order trees, information mining and machine learning have built up an option arrangement which involve supported trees, sacked trees and arbitrary backwoods. The execution was contrasted and distinctive sorts of heart disappointments, for example, saved discharge division and lessened launch portion. Decreased launch portion depended on countless trials and protected discharge was substantially littler and the general guess is by all accounts comparative. The two distinct sorts of heart disappointments were analyzed from the consequences of echocardiography which was regularly accomplished for heart tolerant. In high asset district and treatment choice need before echocardiogram information were accessible and forecast in view of order of heart disappointment was given by adaptable tree based strategy when contrasted and customary characterization and relapse trees. Furthermore, it demonstrates better execution over traditional order strategies.

Basic leadership display has been instituted to gauge the basic elements and information related with Congestive Heart Failure (CHF). The exact finding of coronary illness qualities was very troublesome for crisis tolerant. The precise conclusion of coronary illness is impacted by Rough Sets and choice tree calculations. RS-based model and Logistic Regression were two subsets basic to separate CHF patients with hazard variables were established among 72 research facilities. Approval is performed utilizing vali10-crease cross-approval technique and the consequence of RS-based model was reliably superior to anything LR-based model [19].

An examination convention was made to plan self saw signs and care needs of patients with extreme to exceptionally serious Chronic Obstructive Pulmonary Disease, Congestive Heart Failure (CHF) or Chronic Renal Failure and its results for their nearest relatives. Cross sectional similar and imminent longitudinal review were made on patients with end arrange COPD, CHF or CRF. This review improves the information about the signs, mind needs, parental figure load, end-of-life care treatment inclinations and correspondence needs from the perspectives of patients, their families and their treating doctor. The information acquired here is fundamental to improve painkilling watch over patients with COPD, CHF or CRF. Consequently, the investigation of convention includes its qualities, shortcoming and clinical outcomes.

3.3. Cardiomyopathy

It debilitates the heart muscle or an adjustment in the structure of the muscle because of insufficient heart pumping. A portion of the basic reasons for cardiomyopathy are hypertension, liquor utilization, viral diseases and hereditary imperfections.

3.4. Congenital heart disease

Studies have been made in identifying the false heart rate input and heart signs in patients with inherent coronary illness and nervousness [20]. It is prescribed that an arrangement of high quality uneasiness and sickness history increment the impression of heart manifestations. Fake heart signs may bring about an improved impression of heart signs of innate coronary illness patients. Execution of two exercise assignments has been made utilizing 36 with inborn coronary illness patients and 44 solid controls. Their result demonstrates that intrinsic coronary illness empowers variety with high characteristic uneasiness brings about a weakness to see heart side effects.

Populace based calculation is being made with deference of a proposed anatomic and clinical sorting of inborn heart absconds got from the global pediatric and intrinsic cardiovascular code [21]. Each individual distortion is coded utilizing six digit codes from the extensive rundown of IPCCC and regrouped all sores into 10 sorts and 23 sub-sorts got from multi dimensional procedure including anatomic, indicative and restorative standard. It absolutely utilizes code numbers from considerable rundown of IPCCC yet can hold up ICD-10 codes. The comprehensiveness, ease and anatomic premise make the characterization profitable for clinical and epidemiologic research.

3.5. Arrhythmias

It identifies with the confusion in the musical development of the pulse. The pulse can be moderate, quick, or unpredictable. These unusual heartbeats are brought on by a short out in the heart's electrical framework.

3.6. Myocarditis

It is a kind of inflammation of the heart muscle caused by viral, fungal, and bacterial infections affecting the heart. It is an uncommon disease with less symptoms like joins pain, swelling leg or fever that cannot be directly related to the heart.

3.7. Ischemic heart disease (IHD)

Plaque develops inside the coronary corridors in Ischaemic Heart Disease. It is ordinarily known as coronary vein sickness. In this condition, the blood supply to the heart is diminished because of narrowing of veins. It is principally brought on because of blockage of the supply routes by the testimony of cholesterol on its dividers. Individuals experiencing incomplete blockage of vein show indications like angina while finish blockage of conduit prompts a heart assault. Now and again, diminished blood and oxygen supply prompts the utilitarian hindrance of the heart tissue. A portion of the papers identified with this kind of illnesses are as per the following,

Highlight choice lessens the quantity of contributions under assessment. The proposed strategy assesses productivity regarding cost, time and exactness. It utilizes an Artificial Neural Network to choose vital components from the info layer of the system and chooses the critical elements from an Ischemic Heart Disease database for 712 patients with the assistance of Multi Layer Perception Neural Network. The underlying number of characteristic and highlight determination of the credits is by all accounts 17 and 12. The lessened 12 include characteristics were given as contribution for neural system. The anticipated exactness in preparing was 89.4% high and in testing 82.2% high. What's more, lessened 12 ascribes subjected to be an intriguing element determination for Ischemic Heart Disease dataset [22].

Improved useful movement of bone marrow infers flowing forebear cells after intra coronary naturally separated bone marrow cells transplantation in patients with IHD [7]. The localized necrosis rebuilding was affected by intracoronary autologous Bone Marrow Cells Transplantation (BMCs-Tx) in patients with interminable myocardial dead tissue and intracoronary autologous are inspected. Segregated bone marrow cells transplantation from the randomized controlled review is made through cardiovascular capacity in light of the utilitarian movement of Bone Marrow-Circulating Progenitor Cells (BM-CPCs). 56 patients with IHD have been randomized to either get as of late detached BMC-Tx or a control gathering that does not get cell treatment. Conceivable result are intracoronary transplantation of autologous as of late isolates BMCs prompt a critical diminishment of infarct size and an increase of worldwide launch division and also infracts divider development speed following 3 and 12 months record contrasted with control gather. At long last the intracoronary transplantation of autologous as of late detached BMCs prompt change of BM-CPCs practical movement in fringe blood with better regenerative adequacy for IHD patients.

3.8. Rheumatic heart disease

Rheumatic fever damages the heart muscle and heart valves, caused by streptococcal bacteria.

3.9. Valvular disease

Valvular coronary illness (Disease of valves) is a sort of malady including at least one of the four valves of the heart. These conditions happen to a great extent because of maturing. The vast majority are in their late 50s when analyzed, and more than one in ten individuals more than 75 would experience the ill effects of this illness. Early identification will enhance results from surgery. To enhance the capacity of doctor in finding of valvular coronary illness SAS base programming was utilized to conclusion of valvular coronary illness. Proposed framework includes neural system gathering strategy to make new models by consolidating the back probabilities or the anticipated qualities from numerous antecedent models. The examination comes about shows 97.45% arrangement exactness. and furthermore 100% and 96% affectability and specificity qualities were gotten in valvular coronary illness analysis.

3.10. Data mining concepts in health care

Information Mining goes for finding learning from the accessible information and displaying it in an exact shape that is effectively gotten to by the people. It analyzes a lot of information that are every now and again gathered. Information mining is most helpful in an exploratory examination procedure where no foreordained thoughts are discovered in regards to intriguing result. Information digging looks for new, significant, and nontrivial data from enormous volumes of information. It is a helpful exertion of people and PCs. Best outcomes are accomplished by having information of specialists in depicting issues and objectives with the hunt abilities of PCs. The two essential objectives of information mining have a tendency to be expectation and portrayal. Expectation includes a few factors

or fields in the dataset to anticipate obscure or future estimations of intrigue. On other hand, designs extraction depicts the information that are deciphered by people.

3.11. Predictive data mining

Predictive models can be used to forecast explicit values, based on patterns extraction from known values.

3.12. Descriptive data mining

An unmistakable model underscores designs from existing information in order to make significant subgroups, for example, statistic groups. On the prescient end of the range, the objective of information mining is to deliver a model, as an executable code to perform grouping, expectation, estimation, or other comparative errands. For, illustrative, end of the range, the objective is to comprehend the investigated framework by revealing examples and connections in extensive datasets.

4.   Data Mining techniques used for predictions

The three distinct information mining characterization procedures, for example, Neural Networks, Decision Trees and Naive Bayes calculations have been utilized to break down the coronary illness dataset. Figure 3 outlines the review of the exploration work.

- **Classification**

Order is a machine learning strategy used to foresee aggregate participation for information occasions. So as to anticipate the climate on a specific day, order calculations might be utilized to determine whether it is sunny, stormy or overcast.

- **Supervised learning**

Directed learning is the machine learning undertaking performed on marked preparing information. The preparation information comprise of an arrangement of preparing tests. In administered adapting, each example is a couple comprising of an information protest as a vector and a coveted yield esteem is likewise called the supervisory flag. A regulated learning calculation inspects the preparation information and produces a construed work called classifier. The gathered capacity can foresee the right yield an incentive for any substantial information question through the taking in calculation summed up from the preparation information to inconspicuous circumstances reasonably.

Figure 3. Overview of the research methodology
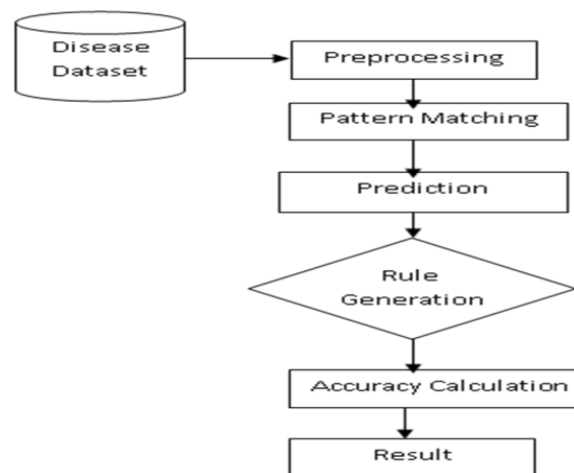
- **Unsupervised learning**

In machine learning calculations the unsupervised learning strategy deals with the issue of discovering concealed structure in unlabeled dataset. Since the specimens gave to the learner are unlabeled, there is no mistake or reward flag to assess a potential arrangement. This condition disengages unsupervised gaining from administered learning.

- **Neural Networks**

Neural system is additionally alluded as an Artificial Neural Network (ANN) by the innovator of the primary neuro PCs, Dr. Robert Hecht-Nielsen. Neural system is a sort of registering framework made up of various straightforward, very interconnected preparing components, to process data in view of their dynamic state reaction to outer sources of info.

ANNs are handling gadgets that are approximately displayed after the neuronal structure of the mammalian cerebral cortex on substantially littler scales. A vast ANN has hundreds or even a huge number of processor units, though a mammalian cerebrum has billions of neurons with a relating increment in extent of their general collaboration with rising conduct. In spite of the fact that ANN specialists are not concerned whether the systems precisely look like natural frameworks or not.

- **Decision Trees**

Decision tree is a flowchart-like structure in which each inner hub speaks to a test on a property, each branch speaks to the aftereffect of the test and each leaf hub speaks to a class mark. The way from root to leaf decides the order rules. For choice investigation, a choice tree and the firmly related impact outline are utilized as a visual and scientific choice bolster tool.   A decision tree comprises of 3 sorts of hubs, for example, Decision hubs Chance hubs and End hubs. Choice hubs looks like a square Chance hubs takes after a circle and End hubs takes after a triangle. Decision trees are normally utilized as a part of operations research and operations administration. On the off chance that the choices are made through online with no review under deficient learning, a choice tree ought to be paralleled by a likelihood show as a best decision display or online choice model calculation. Decision trees may likewise ascertain contingent probabilities.

- **Naive Bayes**

It is one of the characterization procedures in view of Bayes Theorem with a suspicion of freedom among indicators. Just, a Naive Bayes classifier accepts the nearness of a specific component in a class that is inconsequential to the nearness of some other element. For instance, an organic product might be thought to be an apple on the off chance that it is red, round, and around 3 creeps in breadth. Regardless of the possibility that these elements rely on upon each other or identify with the presence of alternate elements, these properties autonomously add to the likelihood that the organic product is an apple. Guileless Bayes model is basic and even utilized for huge datasets. In the event of straightforwardness, Naive Bayes is known to beat even very advanced order strategies. Bayes hypothesis gives a technique to ascertain back likelihood $P(c|x)$ from $P(c)$, $P(x)$ and $P(x|c)$.

- **Clustering**

Grouping is one of the information mining strategies that structures important/helpful bunch of items that have comparable trademark utilizing programmed strategy. Contrasted with grouping, bunching strategy characterizes the classes and incorporates questions inside it, while in order articles are set into predefined classes. In expectation of coronary illness, by applying grouping strategy, bunches are framed that determines the rundown of patients who may have same hazard figure and another bunch demonstrates the rundown of patients with high glucose alongside related hazard elements[23].

5. Conclusion

The database has the information which is diametrically opposed to specialist's supposition and experience. The doctors can enhance the quality of treatment for the patients as predictive algorithms work as the supportive network.

6. References

1. Kimio Tanaka, Natarajan Gajendran, Hideki Asaoku, TaichiKyo, NanaoKamada**.** Higher involvement of subtelomere regions for chromosome rearrangements in leukemia and lymphoma and in irradiated leukemic cell line*. Indian Journal of Science and Technology**.** 2012; 5(1), 1801-1811.
2. S. Akila, S. Chandramathi. A hybrid method for coronary heart disease risk prediction using decision tree and multi layer perceptron. *Indian Journal of Science and Technology*. 2015; 8(34), 1-7.

3. Carlos Ordonez. Association rule discover with the train and test approach for the heart disease prediction. *IEEE Transactions on Information Technology in Biomedicine*. 2006; 10(2), 334-343.

4. Chang-Sik Son, Yoon-Nyun Kim. Decision-making model for early diagnosis of congestive heart failure using rough set and decision tree approaches. *Journal of Biomedical Informatics*.2012; 45, 999-1008.

5. DorairajPrabhakaran, Panniyammakal, Jeemon, Ambuj Roy. Cardiovascular diseases in India. *Current Epidemiology and Future Directions*. 2016; 133(16), 1605-1619.

6. P. Gayathri, N. Jaisankar. Comprehensive study of heart disease diagnosis using data mining and soft computing techniques. *International Journal of Engineering and Technology*.2013; 5(3), 2947-2958.

7. Goekmen R. Turan, I. Bozdag. Improved functional activity of bone marrow derived circulating progenitor cells after intra coronary freshly isolated bone marrow cells transplantation in patients with ischemic heart disease.*Journal of stem cell review and report*.2011; 7, 646-656.

8. Hongmei Yan, Jun Zheng. Selecting critical clinical features for heart diseases diagnosis with a real-coded genetic algorithm. *Journal of Applied Soft Computing*. 2008; 8(2), 1105-1111.

9. JesminNahar, Tasadduq Imam. Association rule mining to detect factors which contribute to heart disease in males and females. *Journal of Expert Systems with Applications*. 2013; 40(4), 1086–1093.

10. Kemal Polat, Salih Gunes. A new feature selection method on classification of medical datasets: Kernel F-score feature selection. *Journal of Expert Systems with Applications*. 2009; 36(7), 10367–10373.

11. Latha Parthiban, R. Subramanian. Intelligent heart disease prediction system using CANFIS and genetic algorithm. *International Journal of Biological and Life Science*.2007; 15, 157- 160.

12. Lucile Houyel, Babak Khoshnood. Population-based evaluation of a suggested anatomic and clinical classification of congenital heart defects based on the International Paediatric and Congenital Cardiac Code. *International Journal of rare diseases*. 2011; 64(6).

13. S. Muthukaruppan, M.J. Er. A hybrid particle swarm optimization based fuzzy expert system for the diagnosis of coronary artery disease. *Journal of Expert Systems with Applications*.2012; 39(14), 11657–11665.

14. OlatubosunOlabode, Bola Titilayo Olabode. Cerebrovascular accident attack classification using multilayer feed forward artificial neural network with back propagation error. *Journal of Computer Science*.2012; 8(1), 18-25.

15. PasiLuukka, Jouni Lampinen. A classification method based on principal component analysis and differential evolution algorithm applied for prediction diagnosis from clinical EMR heart data sets. *Journal of computer intelligence in optimization adaption, learning and optimization*.2010; 7, 263-283.

16. Peter C. Austin, Jack V. Tu. Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes. *Journal of Clinical Epidemiology*. 2013; 66(4), 398-407.

17. Petra A. Karsdorp, MerelKindt. False heart rate feedback and the perception of heart symptoms in patients with congenital heart disease and anxiety. *International Journal of behavioral Medicine*.2009; 16(1), 81-88.

18. R. Pfister, D. Barnes, R.N. Luben, K.T. Khaw, N.J. Wareham, C. Langenberg. Individual and cumulative effect of type 2 diabetes genetic susceptibility variants on risk of coronary heart disease. *Journal of Diabetologia.*2011; 54(9), 2283-2287.

19. K. Rajeswari, V. Vaithiyanathan. Feature selection in ischemic heart disease identification using feed forward neural networks. *International Symposium on Robotics and Intelligent Sensors*. 2012; 41, 1818–1823.

20. Resul Das, Ibrahim Turkoglu. Diagnosis of valvular heart disease through neural networks ensembles. *Journal of Computer Methods and Programs in Biomedicine*.2009; 93, 185-191.

21. N.A. Setiawan. A comparative study of imputation methods to predict missing attributes values in coronary heart disease data set.*Journal in Department of Electrical and Electronic Engineering*. 2008; 21, 266–269.

22. J.Senthil Kumar, S.Appavu. The personalized disease prediction care from harm using big data analytics in healthcare. *Indian Journal of Science and Technology*. 2016; 9(8), 1-6.

23. VahidKhatibi, Gholam Ali Montazer. A fuzzy-evidential hybrid inference engine for coronary heart disease risk assessment. *Journal of Expert Systems with Applications*. 2010; 37(12), 8536–8542.