EFFICIENT CLASSIFICATION OF HEART DISEASE USING MACHINE LEARNING ALGORITHM

Dr. K.P.PORKODI [1], AISHWARYA A [2], DIVYA R [3], INDHUJA K [4], MANASA S [5]

[1] Associate Professor, Department of Computer Science and Engineering, Vivekanandha College of Technology for Women, Thiruchengode, Namakkal Dist.

,[2345] Final Year CSE, Vivekanandha College of Technology for Women, Thiruchengode, Namakkal Dist.,

ABSTRACT:

Heart disease is the one of the major disease and many human beings suffered without any symptoms. In healthcare, especially in finding of heart disease in particular time plays a crucial role in cardiology area. In this paper, we proposed an effective and perfect system to predict heart disease system based on machine learning systems. This system is organized by using various classification algorithms such as LR,KNN,SVM,DT,NB and RF. The proposed algorithmic technique also solve the problem of feature selection and increase the accuracy of classification. In addition with that, the proposed algorithmic technique could use non-invasive clinical data for the heart Disease diagnosis and assessing its severity. The implementation of novel hybrid method helps to improve the accuracy of the EDA diagnosis. The proposed novel hybrid method result shows high accuracy of data is compared to previously proposed techniques. In addition to that, the proposed system is easily be adapted with the existing technology.in this proposed system technology more the 300 instance are collected and the results are compared with existing technology. The results prove that it has more accuracy and it can be easily implemented to identify CAD disease in healthcare field.

INTRODUCTION:

The work proposed in this paper is primarily concerned with the data mining techniques in the prediction of heart disease.Heart disease is a serious health problem that has affected millions of people all over the world.The heart is a muscularcomposed primarily of cardiac muscle. By contraction and dilation in a rhythmic pattern, the heart circulates throughout the body.Heart diseases include coronary heart disease, cardiomyopathy, and cardiovascular disease. Cardiovascular diseases are the conditions that affect the heart and blood vessels. In today's world cardiovascular disease has become life threatening disease. Cardiovascular disease is difficult to diagnose because it is influenced by various factors.High blood pressure, excessive lipids in the blood, stress, metabolic syndrome, diabetes, and other factors are all major causes of cardiovascular disease. A family of history is also a contributing factor.The best way to avoid heart-related illnesses is to live a healthier lifestyle and identify them early on. Before the cardiovascular disease diagnosis, several tests are performed, including electrocardiogram (ECG), blood pressure, cholesterol, blood sugar and auscultation.In the existing diagnostic approaches are ineffective to detect the heart disease at early stages because of various reasons for instance execution time and accuracy. When the medical professionals and modern technology are unavailable, diagnosing and treating the heart disease is incredibly difficult. Many people lives can be saved with the proper diagnosis and treatment.The World Health Organization (WHO) lists cardiovascular diseases as the leading cause of death globally with 17.9 million people dying every year. The majority of the people in the United States are affected by heart disease. Half the deaths in the United States and other developed countries are due to cardiovascular diseases. Only a healthy lifestyle and early detection can prevent heart-related illnesses.Aphysician's examination of the patient's medical history, physical examination report, and interpretation of concerned symptoms is traditionally used to diagnose HD. However, the findings of this

method of diagnosis are insufficient in finding HD patients. Furthermore, it is both costly and computationally challenging to investigate. To address these problems, a noninvasive diagnostic system based on machine learning (ML) classifiers is being developed. An expert decision system based on machine learning classifiers and artificial fuzzy logic is used to diagnose heart disease effectively. Practitioners have improved their use of computer systems to enhance decision-making support in recent years. Machine learning (ML) which is subfield of data mining is adept at handling large, well formatted datasets. In order to find hidden discrete patterns and analyze the given data, machine learning plays a critical role.Following data analysis, machinelearning methods aid in the prediction and early diagnosis of heart disease.The primary aim of this study is to give physicians a tool to diagnose heart disease at an early stage. As a result, it will be easier to treat patients effectively and prevent serious consequences. This research intends to pinpoint the most relevant /risk factors of heart disease. As well as predict the heart disease using various algorithms namely Logistic Regression (LR), K-Nearest Neighbor(KNN), Naive Bayes(NB), Decision Tree(DT), Support Vector Machine(SVM) andrandom forest(RF) algorithm.

LITERATURE REVIEW:

In literature review various machine learning techniques have been proposed by researcher to diagnosis CAD. This research study explains the important of existing CAD techniques based on machine learning ideas. AlekseiDudchenko.[1] developed state of heart system in machine learning techniques and the performance output based on accuracy is 72%. In another study jalaluddin khan. [2] showed that the support vector machine classifier had the best classification performance on the subset of features chosen using the Minimal Redundancy Maximal Relevance feature selection algorithm. Senthilkumarmohan, chandrasegarthirmala. [3] used SBS feature selection algorithm to improve classification accuracy and shorten the computing time of the predictive system. Gudadhe et al. [4] has developed a diagnostic system for HD classification based on multi-layer Perceptron and support vector machine (SVM) algorithms, with an accuracy of 80.41 %. Humar et al. [5] based on neutral network, created a heart disease classification system with fuzzy logic integration and reached accuracy of 80%. Liu et al [6] designed a system of classification for heart disease with rough set and relief procedure. They have been recorded accuracy of 89%. Mohan et al [7] developed prediction system using hybrid machine learning method for heart disease identification. Fahd SalehAlotaibi. [8] concluded that Decision Tree (DT) accuracy was high compared to other algorithms namely Decision Tree, Logistic Regression, Random forest, Naive Bayes and SVM classification algorithms. Deepika, Seema . [9] proposed data mining method to predict heart disease using various algorithms namely, Naïve Bayes (NB), Support Vector Machine (SVM) and Artificial Neutral Network (ANN). SVM accuracy rate is high compared to all other algorithms. Danger et al [10] suggested a prediction system that relies on hidden patterns and relationships between Naive Bayes, Decision Trees and Neutral Network.Somasundaram [11] suggested a technique that incorporates data mining and pattern recoginition for predicting heart disease.

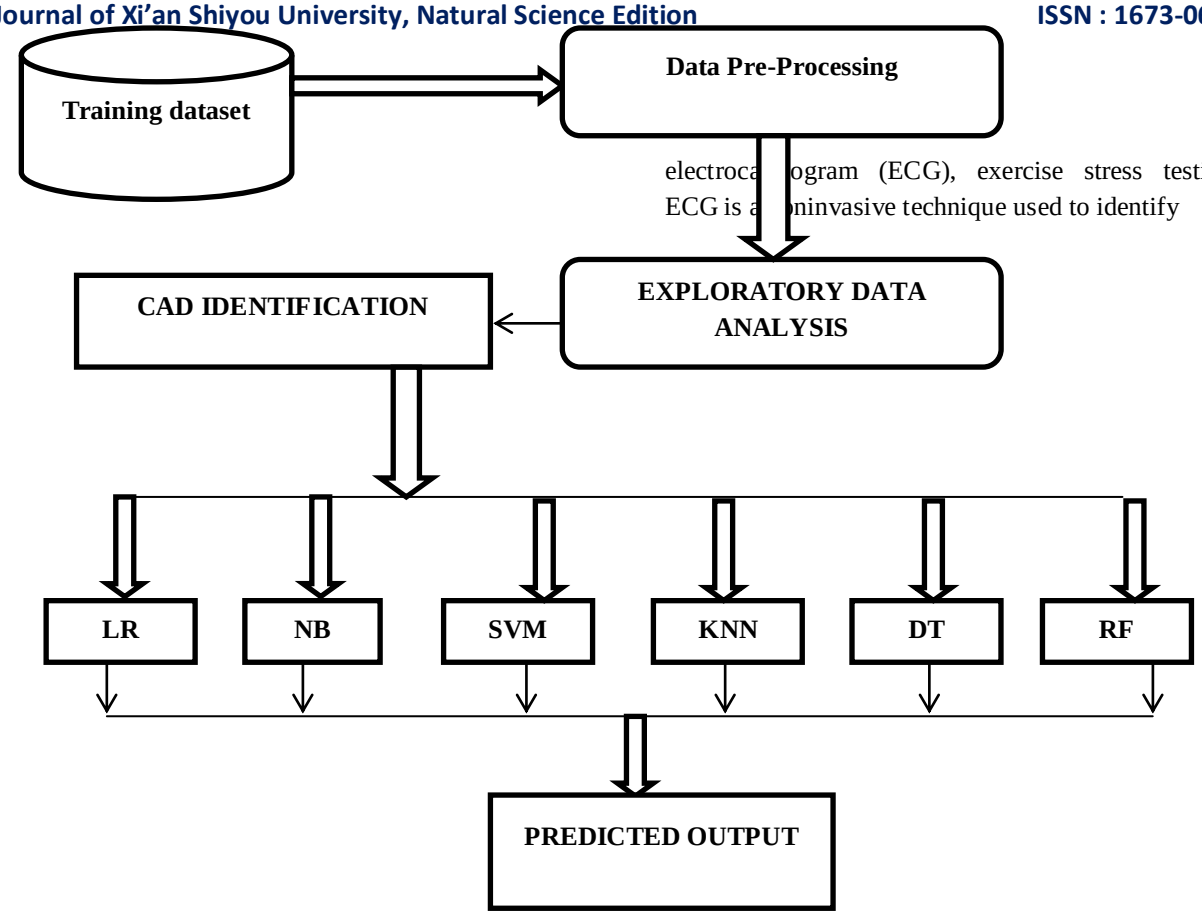
EXISTING SYSTEM:

Existing medical systems, such as hospital administration and decision-making systems, place a heavy emphasis on gathering and processing all medical data.The entire patient records are loaded and all factors are considered. Medical data is difficult to analyze because generating a probabilistic rating requires factors such as test results, current epidemics, medical history, external climate conditions, and a variety of other factors that may or may not be present in the report .A data mining technique was built based on this new metric to investigate the causal association between medications and the risk of coronary artery disease. The exclusive causal-leverage was employed to rank the potential causal associations between each of the selected drugs.Algorithm might successfully place known ADRs at the top of the database's

symptoms list.Existing systems have failed to utilize and understand the importance of misdiagnosis. A important characteristic that connects and addresses all of these difficulties. Mining the misdiagnosis attribute is the key because the first diagnosis by the users would have already covered all the underlying variables like patients medical history,Climatic conditions, neighborhood, and various other factors, allowing the user to just concentrate on either missed variables like hidden symptoms, prevailing conditions, complications, etc., or heart Diseases that are similar to the one already diagnosed.

PROPOSED SYSTEM:

In the proposed work user will search for the heart Disease diagnosis (heart Disease and treatment related information) by giving symptoms as a query in the search engine. These symptoms are preprocessed to make the subsequent process of finding the symptoms keyword, which aids in the rapid identification of heart disease, easier.The symptoms of a keyword are matched against a database of stored medical input to identify the various heart diseases associated with that keyword. Multipleheart Diseases is identified, it will make the pattern matching about the multiple heart Diseases and also find the probability of heart Diseases. Then the heart Disease will make a differential diagnosis to find the heart Disease accuracy. The keyword which is a preprocessed symptom is matched with the heart Diseases stored in the local database to identify the corresponding heart Disease related to those symptoms given by the user. In the proposed work user will search for the heart Disease diagnosis (heart Disease and treatment related information) by giving symptoms as a query in the search engine. These symptoms are preprocessed to make the further process easier to find the symptoms keyword which helps to identify the heart Disease quickly. The symptoms which keyword is matched with the stored medical input database to identify the multiple heart Diseases related to that keyword. Multiple heart Diseases is identified, it will make the pattern matching about the multiple heart Diseases and also find the probability of heart Diseases. Then the heart Disease will make a differential diagnosis to find the heart Disease accuracy. In pattern recognition, exploratory data analysis (EDA), and confirmatory data analysis (CDA) with Feature Selection algorithm is a method for classifying objects based on closest training examples in the feature space. Exploratory data analysis (EDA)are a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification. This feature has been identified as the most suitable for the present system. Individual data instances have structures that make extracting signatures simple. Just collect the symptoms that enough to scaling samples. Using a range of inquiries, it is possible to estimate the degree and degree of cardiac disease. The probability of vocabulary gap between diverse health seekers makes the data more consistent compared to other formats of health data. It is possible heart Diseases ranked according to the number of symptoms matched in the database. Can assign different weights in case of different heart Diseases like common symptoms using LR, NB, SVM, KNN, DT and RF Algorithms.



electrocardiogram (ECG), exercise stress testing, ECG is a noninvasive technique used to identify

Fig 1: Overall Proposed Architecture Diagram

PROBLEM DEFINITION:

Cardiovascular heart Diseases (CVD) are caused by disorders of the heart and blood vessels and result in coronary heart Disease, heart failure, cardiac arrest, ventricular arrhythmias and sudden cardiac death, ischemic stroke, transient ischemic attack, subarachnoid and intra cerebral hemorrhage, rheumatic heart Disease, abdominal aortic aneurysm, peripheral artery heart Disease and congenital heart Disease. According to World Health Organization (WHO), 17.5 million people died from CVD in 2012 amounting to 31 % of all global deaths. CAD is a type of CVD in which presence of atherosclerotic plaques in coronary arteries, leads to myocardial infarction or sudden cardiac death. In order to diagnose positive sign of heart Disease and to assess the level of damage of heart muscles, certain tests may be prescribed by a medical practitioner including nuclear scan, angiography, echocardiogram, CAD case, though it could lead to undiagnosed symptoms of CAD. This limitation leads to angiography which is an invasive diagnosis to confirm CAD cases and is considered as the gold standard for heart Disease detection and severity analysis. However, it is costly and requires high level of technical expertise. Researchers are, therefore, seeking less expensive and effective alternatives, say, using data mining for predicting CAD cases. During the past few decades, image processing, signal processing, statistical and machine learning techniques have been increasingly applied to assist medical diagnosis using ECG and echocardiogram. ECG and echocardiogram are specialized processes conducted by trained practioners. Sometimes ECG is not able to confirm CAD cases. This process is complex, costly, involves lot of time and effort. To overcome these limitations many researchers used

other risk factors excluding angiography to predict CAD cases. These methods are noninvasive, less complex, low cost, reproducible and objective diagnoses, can do automated detection of heart Disease and can be used for screening large number of patients based on clinical data easily obtained at hospitals.

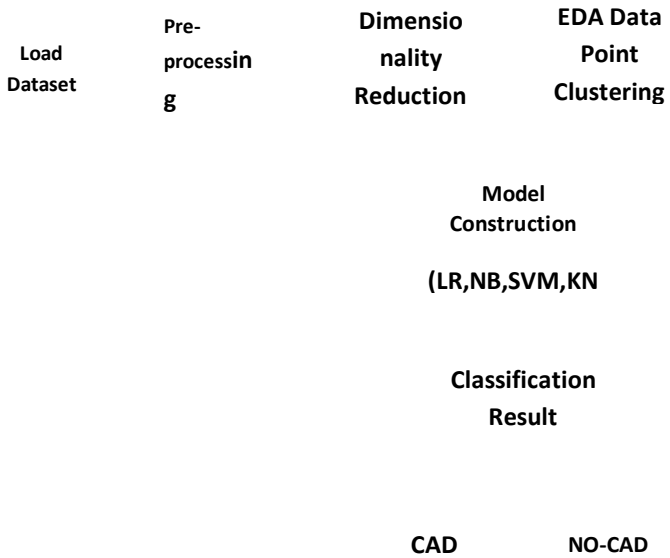


FIG 2: SYSTEM FLOW DIAGRAM

ALGORITHM	ACCURACY
Logistic Regression	85.25
Naive Bayes	85.25
Support Vector Machine	81.97
K-Nearest Neighbors	67.21
Decision Tree	81.97
Random Forest	95.08

GRAPHICAL REPRESENTATION:

A chart is a graphical representation of data, in which "the data is represented by symbols, such as bars in a bar chart, lines in a line chart, or slices in a pie chart ". A chart can represent tabular numeric data, functions or some kinds of qualitative structure and provides different info. The term "chart" as a graphical representation of data has multiple meanings: A data chart is a type of diagram or graph that organizes and represents a set of numerical or qualitative data. Maps that are adorned with extra information (map surrounded) for a specific purpose are often known as charts, such as a nautical chart or aeronautical chart, typically spread over several map sheets. Other domain specific constructs are sometimes called charts, such as the chord chart in music notation or a record chart for album popularity.Then the data are clustered using ,NB,SVM,KNN,DT and RF using all the features of CA. Charts are often used to ease understanding of large quantities of data and the relationships between parts of the data. Charts can usually be read more quickly than the raw data. They are used in a wide variety of fields, and can be created by hand (often on

graph paper) or by computer using a charting applications.

Fig 2: Graphical Performance Analysis

Table 1: Accuracy Comparison Of Algorithm

CONCLUSION:

Clinical finding is a significant region of exploration which assists with recognizing the event of a coronary illness. The framework, utilizing different methods referenced, will thus uncovered the root coronary illness alongside the arrangement of most plausible heart Diseases which have comparative side effects. The information base utilized is a portrayal data set so to decrease the dataset tokenization, separating and stemming is finished. The venture presents a novel mixture model to recognize and affirm CAD cases requiring little to no effort by utilizing clinical information that can be effectively gathered at clinics. Intricacy of the framework is diminished by decreasing the dimensionality of the informational collection with EDA. It gives reproducible and target finding, and subsequently can be a significant extra device in clinical practices.

FUTURE ENHANCEMENT :

Furthermore, we know that irrelevant features also degrade the performance of the diagnosis system and increased computation time. Thus another innovative touch of our study to used features selection algorithms to selects the appropriate features that improve the classification accuracy as well as reduce the processing time of the diagnosis system. In the future, we will use other features selection algorithms, optimization methods to further increase the performance of a predictive system for HD diagnosis. The controlling and treatment of disease is significance after diagnosis, therefore, i will work on treatment and recovery of diseases in future also for critical disease such as heart, breast, Parkinson, diabetes.

REFERENCES:

1. S. I. Ansarullah and P. Kumar, ''A systematic literature review on cardiovascular disorder identification using knowledge mining and machine learning method,'' Int. J. Recent Technol. Eng., vol. 7, no. 6S, pp. 1009–1015, 2019

2. A. U. Haq, J. P. Li, J. Khan, M. H. Memon, S. Nazir, S. Ahmad, G. A. Khan, and A. Ali, ''Intelligent machine learning approach for effective recognition of diabetes in E-healthcare using clinical data,'' Sensors, vol. 20, no. 9, p. 2649, May 2020

3. A. U. Haq, J. Li, M. H. Memon, M. H. Memon, J. Khan, and S. M. Marium, ''Heart disease prediction system using model of machine learning and sequential backward selection algorithm for features selection,'' in Proc. IEEE 5th Int. Conf. Converg. Technol. (ICT), Mar. 2019, pp. 1–4

4. U. Haq, J. Li, M. H. Memon, J. Khan, and S. U. Din, ''A novel integrated diagnosis method for

breast cancer detection,'' J. Intell. Fuzzy Syst., vol. 38, no. 2, pp. 2383–2398, 2020.

5. S. Mohan, C. Thirumalai, and G. Srivastava, ''Effective heart disease prediction using hybrid machine learning techniques,'' IEEE Access, vol. 7, pp. 81542–81554, 2019.

6. G. G. N. Geweid and M. A. Abdallah, ''A new automatic identification method of heart failure using improved support vector machine based on duality optimization technique,'' IEEE Access, vol. 7, pp. 149595–149611, 2019.

7. A. UlHaq, J. Li, Z. Ali, J. Khan, M. H. Memon, M. Abbas, and S. Nazir, ''Recognition of the Parkinson's disease using a hybrid feature selection approach,'' J. Intell. Fuzzy Syst., vol. 39, pp. 1–21, May 2020, doi: 10.3233/JIFS-200075
8. R. Sivaranjani, V. S. Naresh, and N. V. Murthy, ''4 coronary heart disease prediction using genetic algorithm based decision tree,'' Intell. Decis. Support Syst., Appl. Signal Process., vol. 4, p. 71, Oct. 2019

9. A. U. Haq, J. Li, J. Khan, M. H. Memon, S. Parveen, M. F. Raji, W. Akbar, T. Ahmad, S. Ullah, L. Shoista, and H. N. Monday, ''Identifying the predictive capability of machine learning classifiers for designing heart disease detection system,'' in Proc. 16th Int. Comput. Conf. Wavelet Act. Media Technol. Inf. Process., Dec. 2019, pp. 130–138.

10. A. U. Haq, J. P. Li, M. H. Memon, J. Khan, A. Malik, T. Ahmad, A. Ali, S. Nazir, I. Ahad, and M. Shahid, ''Feature selection based on L1-norm support vector machine and effective recognition system for Parkinson's disease using voice recordings,'' IEEE Access, vol. 7, pp. 37718–37734, 2019.

11. 11. Masethe HD, Masathe MA. Prediction of heart disease using classification algorithm. Wcess 2014.
12. Amir M. Early diagnosis of heart disease using Classification and regression trees. University Caglian Italy 2013; 14.
13. Sonwane J. Prediction of heart disease using multilayer perceptron neural network. ICICES 2014; 16.

14. Jabbar MA, Deekshatulu BL, Priti C. Heart disease classification using nearest neighbour classifier with feature subset selection. Annals Computer Science 2013.

15. Carlos Ordonez, "Improving Heart Disease Prediction Using Constrained Association Rules," Seminar Presentation at University of Tokyo, 2004.

16. Franck Le Duff, CristianMunteanb, Marc Cuggiaa, Philippe Mabob, "Predicting Survival Causes After Out of Hospital Cardiac Arrest using Data Mining Method", Studies in health technology and informatics, Vol. 107, No. Pt 2, pp. 12569, 2004.
17. Kiyong Noh, HeonGyu Lee, HoSunShon, Bum Ju Lee, and KeunHoRyu, "Associative Classification Approach for Diagnosing Cardiovascular Disease", Springer, Vol:345, pp: 721 727, 2006.

18. Boleslaw Szymanski, Long Han, Mark Embrechts, Alexander Ross, KarstenSternickel, Lijuan Zhu, "Using Efficient Supanova Kernel for Heart Disease Diagnosis", proc. ANNIE 06,intelligent engineering systems through artificial neural networks, vol. 16, pp:305310, 2006.
19. SellappanPalaniappan, RafiahAwang, ―Intelligent Heart Disease Prediction System Using Data Mining Techniques‖; 9781424419685/08/$25.00©2008 IEEE.
20. Chaitrali S. Dangareet. al., "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", (IJCA) (0975 – 8887), Vol. 47, No. 10, June 2012, page no. 4448.