JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Novel approach for cardiac disease detection using data mining technique

[1]Ms.Deepali R. Jawale, [2]Mrs. Rashmi S. Despande, [3]Mrs. Pooja A. Mishra
[1]Assistant Professor, [2]Assistant Professor, [3]Assistant Professor
[123] Computer Engineering ,
[123]DYPIEMR, Pune, India

Abstract : Data mining is an advanced technology, which is the process of discovering actionable information from large set of data is a tool for analysing massive amounts of data and extracting patterns that can be transformed As a result, many are hesitant to seek effective treatment early in the course of an illness. to knowledge that is beneficial. Medical data mining has a great potential for exploring the hidden patterns in the data sets of medical domain. Clinical diagnosis can be done using these patterns.. These data need to be collected in a standardized form. From the medical profiles fourteen attributes are extracted such as age, sex, blood pressure and blood sugar etc. It can predict the likelihood of patient getting heart disease. In this paper these attributes are fed in to K-means algorithms, MAFIA algorithm and Decision tree classification in heart disease prediction, When the data mining technique is applied to heart disease therapy, it can produce results that are as reliable as those obtained when diagnosing heart disease. This would allow medical industries to provide better patient diagnosis and treatment, resulting in higher service quality. The following are the paper's key benefits: early identification of cardiac disease, timely diagnosis, and treatment at a reasonable cost. This concludes that in this paper the focus is on using different algorithms in data mining and sequence of several attributes for effective heart disease prediction and its diagnosis. Decision Tree has tremendous efficiency using fourteen attributes, after applying genetic algorithm to reduce the actual data size to get the optimal subset of attribute acceptable for heart disease prediction.

IndexTerms **- Component,formatting,style,styling,insert.**

I. INTRODUCTION

Early detection of heart diseases can prevent the death rate, people are not aware about the detection of heart disease earlier due to lack of knowledge. The health-care industry is attempting to detect the sickness at an early stage. The majority of the time, it is only discovered in the last stages of an illness or after death. The costof treatment for heart disease is very expensive. The treatment cost is not affordable for everyone. As a result, many are hesitant to seek effective treatment early in the course of an illness. Our project's goal is to diagnose cancer at an early stage for a reasonable price. We can detect disease at an early stage utilising data mining techniques, and we can entirely cure the condition with accurate diagnosis. Health care industry collect huge amount of data, which are not mined to discover hidden information. The data mining technique provides a solution to this challenge. The process of examining a big amount of data and distilling it into valuable information is known as data mining. Now a day's many people suffer from heart disease, so it's necessary to identify disease symptoms earlier. By motivating thisconcept we propose a novel that earlier diagnosis.

II. LITERATURE SURVEY:

Ischemic Heart Disease Detection and Localization Using Machine Learning Methods Using Magneto Cardiographs Rong Tao, Shulin Zhang et al, [1] The goal of this research was to create a system for detecting and localizing ischemic heart disease that was both quick and accurate. Methods: From averaged MCG recordings, the T wave was segmented, and 164 characteristics were recovered. Time domain characteristics, frequency domain features, and information theory features were divided into three categories. Following that, we compared various machine learning classifiers such as KNN, DT, SVM, and XGBoost. We chose three classifiers with the best performance and used model ensemble to average the results to detect IHD cases. This stage made use of all 164 features. We divided IHD patients into three groups based on the location of stenosis: left anterior descending (LAD), left circumflex artery (LCX), and right coronary artery (RCA) (RCA). We utilised the XGBoost classifier and 18 time domain characteristics for this challenge. The SVM-XGBoost model had the best results for IHD identification, with accuracy of 94.03 percent, precision of 86.56 percent, recall of 97.78 percent, F- score of 92.79 percent, AUC of 0.98, and average precision of 0.98. The XGBoost model had accuracy of 0.74, 0.68, and 0.65 for ischemia localisation in the LAD, LCX, and RCA, respectively. Conclusion: We have created an IHD identification and localization system that is fully automated. 1. T wave repolarization synchronisation is an essential component in distinguishing IHD from normal patients, according to our findings. 2. The location of the stenosis is linked to the magnetic field pattern. Importance: The suggested machine learning technology allows clinicians to make a quick and accurate diagnosis.

The flexible analytic wavelet transform was used on ECG signals to characterize coronary artery disease. U. Rajendra Acharyab, et al,[2] The current study proposes an electrocardiogram (ECG)-based automated diagnosis of Coronary Artery Disease (CAD). First, the ECG signals of 40 healthy people and 7 people with coronary artery disease are split into beats. This study included 137,587 ECG beats from healthy people and 44,426 ECG beats from CAD patients. The ECG beats are decomposed using the Flexible Analytic Wavelet Transform (FAWT) approach. The real values of detail coefficients of FAWTbased decomposition are used to calculate the Cross Information Potential (CIP) parameter. In comparison to normal participants, the mean value of the CIP parameter is observed to be higher in CAD patients.

The discrimination ability of the retrieved characteristics is next tested using the Student's t-test method and the Kruskal– Wallis statistical test. The features are then sent into the Least Squares- Support Vector Machine (LS-SVM) for classification. Starting with the first level of decomposition, classification accuracy is computed at each level of decomposition. Up to the fourth level of breakdown, we noticed a considerable improvement in categorization accuracy. When comparing the fourth and fifth levels of breakdown, classification accuracy does not improve considerably. As a result, we examined ECG beats up to the fifth level of decomposition. In compared to the Radial Basis Function (RBF) kernel, the Morlet wavelet kernel has a greater classification accuracy (99.60%). (99.56 percent). The methodology that has been developed can be employed in large cardiac screenings.

U. Rajendra Acharya, and et al proposed system for Application of higher-order spectra for the characterization of Coronary artery disease utilising ECG signals, [3] Ischemic Heart Disease (IHD), Myocardial Infarction (MI) or Heart Attack (HA), and Heart Failure are all caused by Coronary Artery Disease (CAD), which is the core cause of a chain of catastrophic heart disorders (HF). Early detection and treatment of this CAD illness is critical, as it may help prevent it from worsening. However, manually interpreting electrocardiogram (ECG) signals to identify CAD faster and more accurately is a difficult task. As a result, computer-aided approaches are required for automatic CAD condition characterization. As a result, the use of Higher-Order Statistics and Spectra (HOS) for automated categorization of normal and CAD states using ECG data is proposed in this paper. 182013 ECG beats (137587 normal beats and 44426 beats with CAD) were used in this study. Each ECG beat is analysed for HOS bispectrum and cumulant characteristics. The collected features are used in the dimension reduction approach of Principal Component Analysis (PCA). Then, using the Bhattacharyya method, entropy, fuzzy Max-Relevancy and Min-Relativity (mRMR), Receiver Operating Characteristics (ROC), t-test, and Wilcoxon ranking methods, the PCA coefficients are ranked. To achieve the best classification performance, all ranked features are exposed to k- Nearest Neighbors (KNN) and Decision Tree (DT) classifiers. The proposed methodology has achieved 98.17% accuracy, 94.57% sensitivity, and 99.34% specificity, using KNN classifier using 13 bispectrum features. Similarly, we have obtained 98.99% average accuracy, 97.75% sensitivity, and 99.39% specificity using DT classifier with 31 cumulant features. In addition, we have formulated and developed an integrated index called Coronary Artery Disease Index (CADI) for automated characterization of normal and ECG signals with CAD condition using a single number. This proposed CADI works efficiently to discriminate normal and CAD ECG classes for the any dataset with priory knowledge of the database. Automated identification of myocardial infarction from ECG signals using a deep convolutional neural network U. Rajendra Acharya, Hamido Fujita, ShuLih Oh, Yuki Hagiwara, Jen Hong Tan, Muhammad Adam, U. Rajendra Acharya, Hamido Fujita, ShuLih Oh, Yuki Hagiwara, Jen Hong Tan, Muhammad Adam [4] The electrocardiogram (ECG) is a useful diagnostic technique for determining the cause of a heart attack.

Myocardial infarction is one of many cardiovascular disorders (CVDs) (MI). The ECG records the electrical activity of the heart, and these signals can be used to detect aberrant heart function. However, because to their limited amplitude and duration, visual interpretation of ECG signals is difficult. As a result, we offer a unique method for detecting MI utilising ECG signals automatically. In this study, we used a convolutional neural network (CNN) method to recognise normal and MI ECG beats (with noise) automatically. Using ECG beats with and without noise removal, we attained an average accuracy of 93.53 percent and 95.22 percent, respectively. In addition, no feature extraction or selection is done in this study. As a result, even in the presence of noise, our suggested method can accurately recognise unfamiliar ECG signals. As a result, this technology could be used in clinical settings to help clinicians diagnose MI.

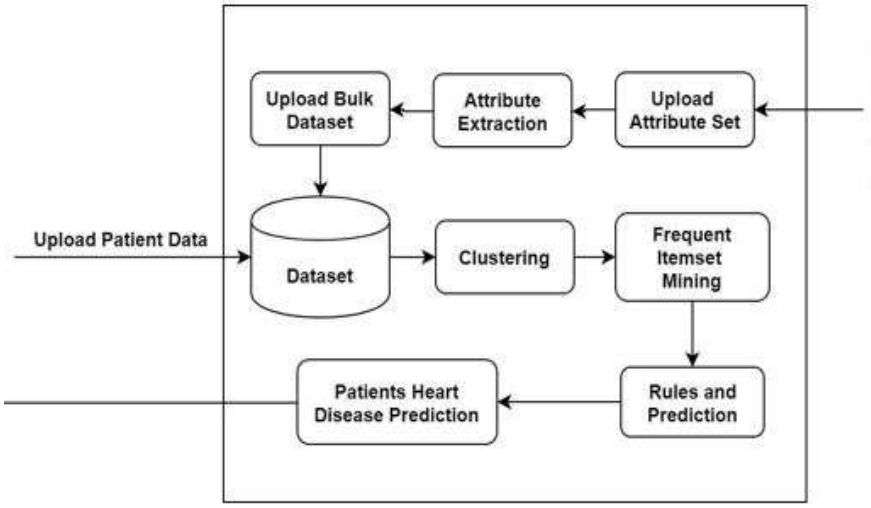
Wavelet Packet Decomposition And Common Spatial Pattern Techniques, Automated Identification Of Coronary Artery Disease From Short-Term 12 Lead Electrocardiogram Signals, Muhammad Adam, Tan Jen Hong, Oh Shu Lih [5] Coronary artery disease (CAD) is caused by the obstruction of coronary arteries, which reduces the normal blood flow to the heart muscles, resulting in irreparable myocardial damage or death (myocardial infarction). In clinical practise, an electrocardiogram (ECG) is used to capture these cardiac activity and determine the presence of CAD. The adoption of computer-aided approaches, on the other hand, can reduce the visual strain and manual time required for analysing complex ECG signals in order to distinguish CAD patients from healthy controls. As a result, an innovative approach was used in this research.

As a result, a novel computer-aided technique for identifying CAD patients is proposed in this work, which uses 2 s of 12-lead ECG signals. To acquire various coefficients, each of the 2 s 12 lead ECG signal beats (3791 normal and 12308 CAD ECG signal beats) is implemented with four degrees of wavelet packet decomposition (WPD). New 2 s. ECG signal beats are reconstructed using the fourth-level coefficients acquired for each lead ECG signal beat. The reconstructed signals are then divided into two data sets, one for generating the common spatial pattern (CSP) filter and the other for obtaining the features vector (vice versa).

For automated classification, the acquired characteristics are input one by one into a k-nearest neighbours (KNN) classifier. Using ten features, the suggested system achieved maximum average classification results of 99.65% accuracy, 99.64 percent sensitivity, and 99.7% specificity. Our suggested algorithm is highly efficient and can be used by clinicians as a diagnostic tool in CAD diagnosis, allowing for quicker treatment and preventing CAD progression.the heart disease by using machine learningtechniques in data mining.

III. PRAPOSED SYSTEM

The proposed method uses data mining techniques to carry out an early diagnosis of heart disease. There is a vast amount of healthcare data that isn't being mined to find hidden information for better decision-making. As a result, we created an application that allows physicians to upload patient data and receive responses from the system. Admin submits attribute sets, which are extracted and used to construct a bulk dataset for disease prediction. The system clusters the dataset using K-means, which leads to frequent item set mining, which is then used to generate predicted heart disease.



K means Clustering :

One of the most basic unsupervised learning strategies for solving the well-known clustering problem is the k-means clustering algorithm. The method employs a straightforward approach to clustering a given data set using a predetermined number of clusters. The main concept is to assign k centres to each cluster. Because different locations produce varied clustering results, these centres should be strategically placed.

As a result, the preferable option is to put them as far apart as feasible. The following step is to associate each point in the given data set with the nearest centre. When there are no outstanding points, the first step is accomplished, and an early group age is completed. We must now recalculate k new centroids as the bary centres of the clusters created in the previous phase. After we've created these k new centroids, we'll need to rebind the identical data set points to the nearest new centre. A loop has been created, and as a result of this loop, the k centres' locations are changing step by step until no further modifications are made, or until the centres stop moving. Finally, this technique seeks to minimise a squared error objective function.

The number of clusters, k, and a database with n objects are provided as inputs.The squared-error criterion is minimised by a set of k clusters.

Method:

1) Choose k items at random as the first cluster centres;
2) Repetition
3) based on the mean value of the objects in the cluster, assign each item to the cluster to which it is most comparable;
4) update the cluster means, i.e. calculate the mean value of the objects for each cluster;
5) until no change;

Decision Tree :

A decision tree is a flowchart-like tree structure in which each internal node represents an attribute test, each branch reflects the test's conclusion, and leaf nodes indicate classes or class distributions. The root node is the topmost node in a tree. Rectangles represent internal nodes, while ovals represent leaf nodes. The sample attribute values are checked across the decision tree to classify an unknown sample. Decision trees may be readily turned to classification rules by drawing a path from the root to a leaf node that carries the class prediction for that sample[12].

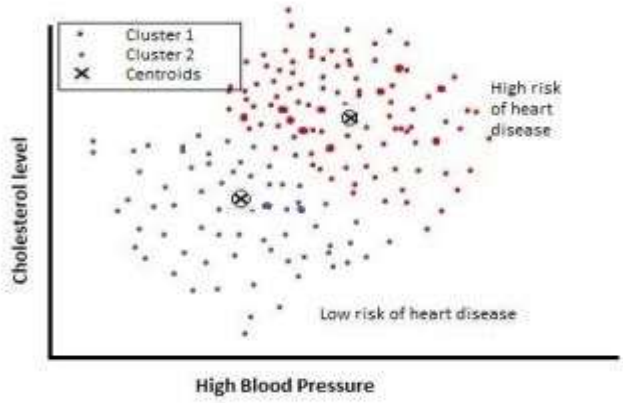


Fig. Cardiac Disease Detection

IV. RESULTS AND CONCLUSION :

As a result, the focus of this study is on the use of different methods in data mining and the sequence of several attributes for effective heart health. Prediction and diagnosis of disease After applying a genetic algorithm to minimise the actual data quantity to acquire the ideal subset of attributes acceptable for heart disease prediction, the Decision Tree has great efficiency employing fourteen attributes.

REFERENCES

[1] M. Kurnar, R. B. Pachori, and U. R. Acharya, "Characterization of coronary artery disease using flexible analytic wavelet transform applied on ECG signals," Biomedical Signal Processing and Control,vol. 31, pp. 301-308, Jan 2017.

[2] U. R. Acharya, V. K. Sudarshan, J. E. W. Koh, R. J. Martis, J. H. Tan, S. L. Oh, et al., "Application of higher-order spectra for the characterization of Coronary artery disease using electrocardiogram signals," Biomedical Signal Processing and Control, vol. 31, pp. 31-43, Jan 2017.

[3] A. Gallino, "The dilemma of detecting silent myocardial ischemia," International Journal of Cardiology, vol. 244, pp. 86-86, Oct 1 2017.

[4] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, and M. Adam, "Application of deep convolutional neural network for automated detection of myocardial infarction using ECGsignals," Information Sciences, vol. 415, pp. 190- 198, Nov 2017.

[5] O. S. Lih, M. Adam, T. J. Hong, H. Yuki, V. K. Sudarshan, K. J. E. Wei, et al., "Automated identification of coronary artery disease from short-term 12 lead electrocardiogram signals by usingdecomposition and common spatial pattern techniques, Journal of Mechanics in Medicine and Biology, vol. 17, Nov 2017.

[6] M. Anbarasi, E. Anupriya, N.Ch.S.N.Iyengar, "Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm",International Journal of Engineering Science and Technology Vol. 2(10), 2010, 5370-5376.

[7] M.Akhil jabbar, Dr.Priti Chandra, Dr.B.L Deekshatulu " Heart Disease Prediction System using Associative Classification and Genetic Algorithm", ICECIT, 2012.

[8] R. Sethukkarasi and Kannan, "An IntelegentSystem for Mining Temporal rules in Clinical database using Fuzzy neural network", European Journal of Scientific Research, ISSN 1450-216, Vol70(3), pp 386-395, 2012.

[9] Mohammed Abdul Khaleel, Sateesh Kumar Pradhan, " Finding Locally Frequent Diseases Using Modified Apriori Algorithm", International Journal of Advanced Research in Computer andCommunication Engineering Vol. 2, Issue 10, October 2013.

[10] Boshra Bahrami, Mirsaeid Hosseini Shirvani, "Prediction and Diagnosis of Heart Disease by Data Mining Techniques", Journal of Multidisciplinary Engineering Science and Technology (JMEST) ISSN:3159-0040 Vol. 2 Issue 2, February–2015.

[11] Deepika .N, "Association Rules for Classifiaction of Heart Attack Patients", IJAEST, Vol 11(2), pp 253-257, 2011.

[12] Sarath Babu, "Heart Disease Diagnosis Using Data Mining Technique" IEEE ICECA 2017, 750-753