

HYBRID MACHINE LEARNING CLASSIFICATION TECHNIQUE TO PREDICT THE ACCURACY OF ACCURACY OF HEART DISEASE

¹CH. Murali Krishna Yadav, ²D. Usha, ³J. Vandana, ⁴G. Neelima, ⁵B. Yuva Teja,

¹Assistant Professor, ^{2,3,4,5}Students
Department of ECE.

N.B.K. R Institute of Science and Technology.
Tirupati (Dist.), Andhra Pradesh.

Abstract- Researchers have shown a lot of interest towards the field of medical science. A significant number of researchers have uncovered various reasons for premature human death. According to the literature that pertains to the subject, various reasons are responsible for the development of diseases, and one such cause is conditions affecting the heart. Many researchers proposed idiosyncratic methods to preserve human life and help health care experts to recognize, prevent and manage heart disease. The expert's decision can be supported by certain convenient methods, but every effective plan also has its limitation. The proposed approach robustly analyses an act of Decision Tree, Random Forest, XGBoost and Hybrid Model. After analysing the procedure, the intended method smartly builds. Initially, the intention is to select the most appropriate method and analysing the act of available schemes executed with different features for examining the statistics.

Keywords — Machine learning, Classification Technique, Decision Tree, Random Forest Stacking Classifier, supervised machine learning.

INTRODUCTION:

Some of the related works represent various convenient methods with the implication but none of the methods aid professionals under different characteristics. Therefore, the design and implementation of these methods pave the way for further research. Additionally, the presented work indicates that the utilization of the data mining method works better than other approaches. With a discussion of research objectives, motivation, and key findings this chapter describes the contribution towards the direction to improve the QoS of the system. Selection and formation are the most appropriate features instead of employing a complete list of features that are associated with the selected dataset. Data mining involves uncovering valuable and anticipatory insights from vast quantities of data. It is an innovative tool with great potential to help companies and mainly focus on the most essential information in their data warehouses. The term frequently used to refer to data mining is Knowledge Discovery in Databases (KDD). KDD is the important process of identifying valid, new, potentially useful, and finally understandable patterns in data. Knowledge discovery process has iterative sequential steps of processes and data mining is one of the KDD processes.

Many researchers have been keenly interested for performing research work using ensemble learning techniques during the last de Significant improvement in performances have been noted using ensemble approaches by several authors. These ensemble approaches have expanded their scope in variety of areas including health care, finance, insurance, automobile, manufacturing, bioinformatics, aerospace and many more. An ensemble of classifiers is a combination of individual classifiers that helps in the classification of new test samples by joining and assessing the individual techniques separately in some manner. In supervised learning, ensemble learning has become one of the profound areas of study among machine learning researchers. Research shows that this composite model using ensemble classification has mostly outperform the capability of one single model. Numerous experimental study and analysis have depicted that the generalized error has been reduced by combining the outputs of multiple classifiers. The key objective of using ensemble approaches has to build a model of ensembles that incorporates a combination of diversified individual classifier techniques with good accuracy. In this work, we have focused our interest in the field of Bioinformatics due to its huge popularity in the research domain. The various ensemble learning methods used in this paper are described as follows. Bagging is one of the commonly used ensembles learning technique based on Bootstrap sampling technique introduced by Breiman. In this samples called bags are created in order to construct the individual classifier of the same algorithm or to manipulate the selection of training data. Here the data point is selected randomly along with the replacement strategy i.e. some data is taken from random sample and some is missed from the original dataset. The separate classifier is trained from each bag. Bagging combines all the classifiers constructed in previous phase and test sample is predicted by giving votes to the classifiers individually

LITERATURE SURVEY:

Comparative analysis of several data mining classification techniques on the parameters accuracy, execution times, types of data sets and applications by Ritu.Sharma, Mr Shiva Kumar, Mr Rohit Maheswari[1]. Prediction model which not only predict the occupational incidents but also provides rules for explaining accident scenarios like near miss, property damage ,or injury cases. Classification and regression tree (CART) is used for prediction purpose by Shoban Sarkar, Atul Patel, Sardhak Madaan, Jareswar Maiti[2]. Novel ensemble learning approach “BBS method” which stands for bagging boosting and stacking with appropriate base classifiers for the classification of the UCI Data sets taken from the field of Bioinformatics by

Aayushi Verma, Shikha Mehta [3]. Demonstrate that Havrda Entropy and Harris Corner Detector based melanoma analysis approach accomplish greater affectability by K.C. Giri, and D. Gautam [4]. Crime forecasting is a way of mining out and decreasing the upcoming crimes by forecasting future crime that will occur by Ayisheshhim AlMaw, Kalyankadam [5]. The Data mining classification techniques, namely Support Vector Machine (SVM) and Random Forest (RF) are Analysed on Diabetes, Kidney and Liver disease Data base. The performance of these techniques is compared, based on precision, Recall, Accuracy, f-measure as well of time by Shakuntala Jatav and Vivek Sharma [6].

PROPOSED SYSTEM:

After reviewing all the existing techniques, some of the researchers signifying the various advantages of each suggested technique and elaborated several restraints that are still associated with obtainable methods and highly affect the working behaviour of the techniques. Among several associated issues, some of the key restraints such as inflexibility time consuming for building a model, alternative parameters, and inaccurate verdicts [7].

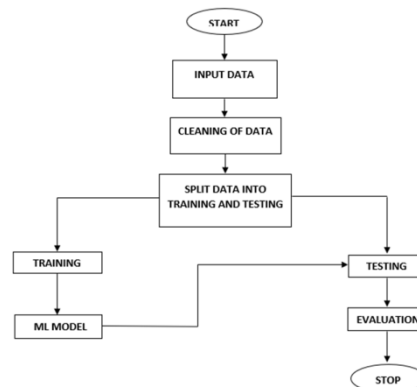


Fig 1: Flow chart

In the proposed work, there are four different classification algorithms were selected along with the two feature compressing methods as CFS with best-first search and Gain ratio with ranker mechanism. As described in the literature survey each algorithm is designed with an obtainable process in an optimized form, such a selected process may not be utilized to build a more competent method. The proposed method investigate and analyze four chosen method such as Hidden Markov Model (HMM), Artificial Neural Network (ANN), Support Vector Machine (SVM) and Decision Tree (J48) along with two other feature compressing methods. [2, 17] After analyzing these feature compressing methods, combine them with the linear models. And if any data is mismatched then reexamine with the other employed technique to improve the QoS.

If you have the necessary rights, you can upload, attach, and manage files to content using the File module. This module is in charge of managing submitted files and evaluating file content. Additionally, it offers choices for showing file content. The Dataset class and other subclasses are defined in the dataset module and are used to manage datasets. Users can use both pre-built and custom datasets. (see the Getting Started page for examples). There are now three built-in datasets available Molecules, knowledge graphs, social networks (including Twitter, Mastodon, and any citation networks linking publications and authors), and other objects are examples of graphs. (such dictionaries, websites with linkages between pages, and UML diagrams), any 3D mesh, sentences described as their syntactic trees, and more! As a result, it is not exaggerating to claim that graphs are ubiquitous. [7].

A graph's (or network's) nodes (or vertices) are its constituent parts, and its edges are the connections between them. (or links). For instance, nodes in a social network are users who form the edges of connections, whereas nodes in a molecule are atoms who form the edges of a chemical bond.

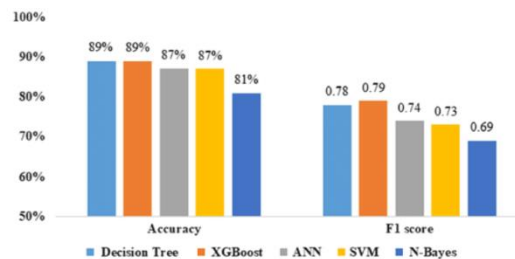
ALGORITHMS:

A supervised learning method called a decision tree can be used to solve classification and regression problems, but it is typically favoured for doing so. It is a tree-structured classifier, where each leaf node represents the classification outcome and inside nodes represent the features of a dataset.

The two nodes in a decision tree are the Decision Node and Leaf Node. While Leaf nodes are the results of decisions and do not have any more branches, Decision nodes are used to create decisions and have numerous branches [8]. Popular machine learning algorithm Random Forest is a part of the supervised learning methodology. It can be applied to ML issues involving both classification and regression. It is built on the idea of ensemble learning, which is a method of integrating various classifiers to address a challenging issue. XGBoost is a distributed gradient boosting library that has been optimised for quick and scalable machine learning model training. A number of weak models' predictions are combined using this ensemble learning technique to get a stronger prediction. Extreme Gradient Boosting, or XGBoost, is one of the most well-known and widely used machine learning algorithms because it can handle large datasets and perform at the cutting edge in many machine learning tasks like classification and regression.

RESULT:

The output of this project is to determine the heart condition of the patient before it occurs any heart disease. This will find with maximum accuracy compared to previous ones.

**CONCLUSION:**

The work done for this inquiry is to increase effectiveness, suitability, and QoS. The characteristics and limitations of existing methods were discussed in the literature survey to build a more efficient method. The proposed work investigates four different algorithms such as the Random Forest, XGBoost and a form of Decision Tree (J48). The proposed method robustly analyses these four methods to exploit statistics and opts for the pair of the finest algorithm that utilizes a linear model based on the feature selection process with best-first search and Gain ratio along with the Ranker method. Several simulations have been carried out to demonstrate the efficiency of the proposed approach. Each comparison has indicated that the proposed approach effectively improves the issues of traditional as well as modern algorithms.

REFERENCES:

1. Ritu Sharma, Mr Shiva Kumar, Mr Rohit Maheshwari (Comparative Analysis of Classification technique in Data mining Using Different Data sets).
2. Shoban Sarkar, Atul Patel, Sardhak Madaan, Jareswar Maiti (Prediction of occupational accidents Using decision tree approach).
3. Aayushi Verma, Shikha Mehta A (Comparative study of ensemble learning methods for classification in Bioinformatics.)
4. K.C. Giri, and D. Gautam A novel paradigm of melanoma diagnosis using machine learning and information theory.
5. Ayisheshhim Almaw, Kalyankadam (Survey paper on Crime prediction using ensemble approach)
6. Shakuntala Jatav and Vivek Sharma (An algorithm for predictive data mining approach in medical diagnosis).
7. Hybrid Machine Learning Classification technique for improve accuracy of heart disease.