Proceedings of the International Conference on Sustainable Computing and Data Communication Systems (ICSCDS-2023)
IEEE Xplore Part Number: CFP23AZ5-ART; ISBN: 978-1-6654-9199-0

Heart Disease Prediction using Ensemble ML

1st Madhvan Bajaj
Computer Science and Engineering
Graphic Era Hill University
Dehradun, India
bajajmadhvan21@gmail.copm

2nd Priyanshu Rawat
Computer Science and Engineering
Graphic Era Hill University
Dehradun, India
priyanshurawat15202@gmail.com

3rd Chandradeep Bhatt
Computer Science and Engineering
Graphic Era Hill University
Dehradun, India
bhattchandradeep@gmail.com

4th Rahul Chauhan
Computer Science and Engineering
Graphic Era Hill University
Dehradun, India
chauhan14853@gmail.com

5th Teekam Singh
Computer Science and Engineering
Graphic Era Deemed to be University
Dehradun, India
teekamsingh.iitr@gmail.com

2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS) | 978-1-6654-9199-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICSCDS56580.2023.10104770

***Abstract* — Worldwide, heart disease is one of the main causes of mortality. Heart disease early identification and prevention can significantly enhance patient outcomes and save healthcare expenditures. This research intends to create a model for forecasting an individual's probability of developing heart disease based on several risk factors. A sizable patient data collection, comprising demographic information, laboratory test results, and lifestyle characteristics, will be used to train the algorithm. Standard measures like accuracy, sensitivity, and specificity will be used to assess the model's performance. This project aims to develop a tool that will help medical practitioners recognize high-risk individuals and carry out early therapies to stop the progression of heart disease.**

Keywords* — *heart disease prediction, ensemble machine learning, sensitivity, specificity

I. INTRODUCTION

With an estimated 17.9 million fatalities each time, heart complaint is one of the main causes of mortality encyclopaedically. It causes a great deal of morbidity and mortality, which is a huge public health problem. According to studies, heart complaints will soon overtake all other causes of mortality in the world [2]. Improving patient outcomes and lowering healthcare costs depend on the early identification and treatment of cardiac complaints. Discovering models that can read a person's chance of developing heart disease based on several threat factors. By employing a sizable collection of patient data to train the model, including demographic information, laboratory test results, and life variables. Determine the main threat variables that may be used to read the development of heart complaints by examining patterns and trends in this data. Standard criteria like accuracy, sensitivity, and specificity are used to assess the models. The idea is to develop a tool that will aid medical interpreters in spotting high-threat individuals and putting early curatives into action to stop the progression of heart complaints. The model might also be used to identify possible troubles and help people in making educated health-related decisions. This action has the implicit to significantly impact the battle against heart complaints and enhance the lives of individuals tormented by this illness by offering insightful information on heart complaints' causes and threat factors. The usage of conventional techniques for gathering data from patients in hospitals is currently gradually declining.

In the meantime, there is a noticeable rise in the use of computer technology for medical diagnosis and treatment. Additionally, data sizes are progressively growing. The challenge dealing with these vast volumes of data is difficult. A typical database is built using Boolean logic, where data is either entirely true or entirely false. As a result, it is now understood that the development of artificial intelligence techniques is crucial for tackling complicated problems.

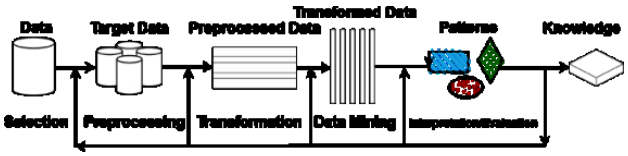

Fig. 1. Steps in Knowledge Discovery Process

Due to the complexities of the human mind and body, which result in a limited understanding of their activities, medical diagnosis is a particularly effective design.

Machine learning can be said to be a method of teaching computers to learn using data, without being specifically programmed for a certain task. Many different models get encompassed in ML. the first one is known as supervised model. A model is trained on labeled data in supervised learning, a sort of learning in machine environments where the desired output is already known. The objective of supervised learning is to develop a function that can translate inputs to output such that the model can anticipate the right result when it is presented with fresh, illustrative data using the patterns it has discovered from the training set. An input-output pair-containing dataset which is labeled is used to train the model. The model is then put to the test on brand-new datasets to forecast the results based on the patterns it has discovered. Linear regression and support vector machines [9] are a few types of supervised learning. The second common learning is unsupervised and this learning don't provide the model access to labeled data. In its place, the model is given unlabeled data and is required to discover patterns or associations on its own. The model is taught to generate predictions based on the patterns it discovers from the labeled data in supervised learning, in contrast when the model is given labeled data. Anomaly detection, dimensionality reduction, and grouping are common unsupervised learning methods.

Machine learning and supervised data mining techniques are categories. This also covers his two actions. The first is learning, which involves creating and honing a model using labelled data (the training set). The second is a classification (test), where class labels are generated for the given data using the model. By forecasting, determine the classifier model's accuracy (test data). Finding associative rules or patterns in data is done using the data mining approach

known as associative rule mining. In association rule mining, a pattern is found based on how one item is related to other things purchased simultaneously. Utilizing predetermined support and confidence levels, it identifies common item groupings within the data. The association rule approach is used to diagnose heart disease to determine the link between various characteristics used for analysis and sort the patient with all of the risk factors needed for illness prediction.

Machine learning techniques called clustering separate related data points into clusters. To find patterns or structures in the data that are not immediately apparent by just looking at the raw data, clustering is used. To group data points that are similar to one another, clustering algorithms first define a similarity metric for each data point. Clustering algorithms come in a wide variety of forms, including k-means, hierarchical clustering, and density-based clustering. With the help of the popular clustering technique K-means, a dataset is divided into k number of clusters, and each observation is assigned to the cluster with the closest. The technique known as "hierarchical clustering" produces a hierarchical representation of the items being clustered, and each level of the hierarchy represents a grouping of the data at a particular level of resolution. A technique, density-based clustering marks clusters as dense areas in the data space. Market segmentation, picture segmentation, and anomaly detection are just a few examples of the many applications that clustering, an unsupervised learning technique, is utilized [3].

A tree-like graph is used in a decision tree or decision model as a decision support tool [12]. This tool takes an input as an object or record defined by a collection of characteristics and returns a decision containing the expected output value for the input. Both discrete and continuous input quality is possible. After a series of tests are run, decisions are made through a decision tree. Each non-leaf node branch of the decision tree is labeled with a possible outcome of the test, and each node corresponds to a test for the corresponding attribute value. The decision delivered when a leaf node in the tree is reached is specified by each random forest logistic tree and leaf or end node. The phrase "naive bias" in machine learning refers to a particular kind of bias that happens when a model makes unsupported assumptions about the data. Because it can produce overfitting, where the model performs well on training data but badly on untested data, this method can be problematic to work with. A lack of generalization, when the model performs badly on a wide variety of inputs, might also result from it. To make sure the model is not overfitting or underfitting, it is crucial to carefully evaluate the assumptions it makes about the data and to test it using a range of inputs. Machine learning methods with nonlinear data processing capabilities include artificial neural networks. A mathematical or computer model based on a biological neural network is termed an artificial neural network, or simply a neural network. It mimics the biological brain system, in other words. Input, output, and often several hidden layers are all parts of a neural network. They are great tools for seeing intricate patterns in data, and they keep becoming better as they learn from prior mistakes.

A population-based optimization method called a genetic algorithm is one that draws inspiration from natural selection. It may be used to locate the best answers to a variety of optimization and search issues. To identify the optimal answer to a particular problem, the basic concept is to mimic the process of natural selection. The initialization of the population is the first stage of a genetic algorithm.
The genetic algorithm starts the evolution after the population has been established. The population's members are assessed using a fitness function that gives each member of the population a value depending on how effectively it solves the challenge. To choose next-generation parents, the fittest people are chosen. Following that, the crossover is employed to produce kids using the chosen parents. In the crossover, two parents' genetic material is combined to produce a new person. This procedure mimics how reproduction works during natural selection.

A procedure known as the mutation is used to add random differences in the genetic makeup of the people once the new generation has been produced. Mutation simulates the process, genetic mutation by natural selection. The process of selection, crossover, and mutation is then repeated after evaluating the new generation of people using the fitness function. This procedure keeps on until an ideal answer is discovered or a halting requirement is satisfied. Numerous optimization and search issues, such as function optimization, ML, and artificial intelligence, may be solved using genetic algorithms. They are especially helpful for issues where there is a wide search space and the issue is poorly understood.

II. LITERATURE REVIEW

Numerous studies have been conducted in the area of heart disease prediction, using various machine learning techniques such as decision trees, support vector machines, and artificial neural networks [7-8]. In a study by H. Jindal et al. (2021), a decision tree algorithm was used to predict heart disease based on demographic and clinical risk factors. The results showed an accuracy of 78% in predicting heart disease. In another study in 2021, an SVM was used to predict heart disease based on ECG data. The results showed an accuracy of 85% in predicting heart disease. Utilizing deep learning methods, such as convolutional neural networks (CNNs) and recurrent neural networks, to predict cardiac disease is the state of the art now (RNNs). With accuracy rates over 90%, these algorithms have demonstrated encouraging results in cardiac illness based on ECG data. A CNN was used to predict cardiac illness based on ECG data and it reached an accuracy of 92% [3].

R. Indrakumari proposed a method based on exploratory data analysis for predicting the heart disease. In that work, the dataset is analyzed using the visualization tool tableau and K means clustering; the work include 209 sample with 7 independent attributes for prediction of heart disease. The attributes include ECG at rest, blood glucose level, heart rate, age, blood pressure, the type of chest pain, and four different types of chest pain; Pre-processing the heart disease dataset efficiently involves removing irrelevant items and providing values for any missing tuples. By considering key variables and the four different forms of chest discomfort, heart illnesses are predicted (Atypical angina, Typical angina, Asymptomatic, Non-Anginal pain). The outcomes of the data

analysis to find the requisite hidden patterns for heart disease prediction are shown [1]. Age, the kind of chest pain, blood pressure, blood sugar level, resting ECG, heart rate, and the four different forms of chest pain and exercise angina are factors that are taken into consideration when predicting heart disease. Pre-processing the heart disease dataset efficiently involves removing irrelevant entries and assigning values to tuples that are missing. The K-means technique is used to construct the pre-processed heart disease data set. The topics of asymptomatic pain, atypical angina pain, non-anginal pain, and non-anginal pain are covered. The outcomes are calculated utilizing all four forms of chest discomfort along with other determining factors.

Krittanawong C. et al. explained logistic regression, and random forest ML methods for disease prediction. It can aid in the diagnosis of disease with less medical tests, so that they can be treated, by identifying anybody who is experiencing any heart diseases symptoms, such as high blood pressure. Discussions & Conclusions These results show that the majority of researchers use a number of algorithms, including SVC, Decision Trees for Heart Disease Patients Identification, KNN, and Random Forest Classifier. It is visible from the work that the models of Logistic Regression and KNN are more accurate in the diagnosis of a cardiac condition, which is higher or almost like accuracies reported from prior studies [5]. By cleaning the dataset and using logistic regression and KNN, this project aids in the prediction of patients who would be diagnosed with cardiac illnesses. This is an improvement over the prior models' accuracy of 85%. Accuracy is 87.5%, which is higher than the accuracy of the prior system, which only employed one data mining approach and possibilities.

A work done by Armin Yazdani and others published in 2021, This work succeeded to make a substantial contribution to computing with significant predictors in heart disease prediction. Weighted Association Rule Mining is one of the data mining techniques used to find the connections between attributes and identify the mining rules that provide certain predictions. This study helped WARM get the best confidence score for predicting heart disease with relevant factors. To predict heart disease, a collection of relevant characteristics with varying weights to reflect the strength of each feature was utilized. Only the most widely utilized machine learning approaches for heart disease prediction research were employed in the feature selection phase of this study [4]. The technique we are going to use in Ensemble Learning. Multiple base models are trained on the same data using the ensemble learning technique, which then combines the predictions from each model to create a more reliable and accurate final prediction. Using several models, ensemble learning aims to overcome the weaknesses of individual models, such as overfitting or strong bias [6]. There are several kinds of ensembles, including A technique known as bagging (Bootstrapped Aggregating) involves training several models on a selection of training data. Boosting is the process of successively training many models, with each model focused on the flaws of the one before it. Stacking is the process of training a meta-model to generate predictions utilizing those of many base models. For heft technique, S. Gupta proposed an efficient prioritization method and processor selection schemes. It was a makespan optimizer which can help to schedule the task in cloud end. The next section describes the dataset and methodology of the work.

III. DATASET AND METHODOLOGY

The dataset contains 303 patient-level characteristics, such as whether or not they have heart disease. Age and sex are examples of features (0: Female 1: Male), Resting Blood Pressure, Cholesterol (Serum Cholesterol in mg/dl), Fasting Blood Sugar, Chest Pain Type (0: Typical Angina, 1: Atypical Angina, 2: Non-Anginal Pain, 3: Asymptomatic), Resting Blood Pressure, Less than 120 mg/ml (0: Less Than, more than 120 mg/ml), Electrocardiogram taken while resting (0: Normal, 1: ST-T Wave Abnormality, 2: Left Ventricular Hypertrophy), Max Heart Rate Attained, Exercise Induced Angina (1: Yes, 0: No).

- SVM Classifier: For classification and regression analysis, Support Vector Machines (SVM), a supervised learning technique, is used. It works by finding the hyperplane in high-dimensional space that separates the classes in the most efficient way [13]. SVM has proven useful in several real-world applications due to its skill with non-linear boundaries and large-dimensional data.

- K-Neighbors Classifier:
  A supervised learning approach for classification and regression analysis is K-Nearest Neighbors (KNN). It operates by identifying the K nearest data points to a new sample and predicting the class label of the new sample based on the majority of these K nearest data points. Both linear and non-linear boundaries may be handled using the straightforward, quick, and efficient KNN method [14, 17, 18]. It works well with short datasets and low-dimensional feature spaces, but when dealing with huge datasets or high-dimensional feature spaces, it can become sluggish and ineffective.

- Decision Tree: A well-liked ML learning technique in a supervised method for tasks like regression analysis is the decision tree and for classification problems. It builds a model of choices and potential outcomes that resembles a tree. Each leaf node provides a class label or prediction, whereas each interior node represents a judgment based on a characteristic. Up until a halting requirement is satisfied, the algorithm continually splits the data depending on the characteristic that yields the greatest information gain [7]. By traversing the final tree based on the feature values of a fresh sample, predictions may be made.

- AdaBoost Classifier: An ensemble learning approach for classification and regression analysis is called AdaBoost (Adaptive Boosting). It functions by fusing many "weak" learners to create a strong learner. A model that outperforms random guessing but still makes mistakes is said to be weakly learner. AdaBoost trains a new weak learner to rectify the mistakes made by the previous weak learner by concentrating on the samples that were misclassified in each iteration. Votes that are weighted

according to how well each weak learner anticipated the outcome are merged with all of the weak learners' predictions. A quick and efficient technique that avoids overfitting is called AdaBoost.

- Gaussian NB: A probabilistic supervised learning technique used for categorization analysis is called Gaussian Naive Bayes (GNB). It employs the Bayes theorem to determine the likelihood of a class label given the feature values of a fresh sample under the assumption that the features in the data are independent and have a Gaussian distribution. The forecast is then made using the class label with the highest likelihood. GNB is a straightforward and quick technique that performs well for text classification and high-dimensional data. However, because it ignores the links between characteristics, its independence assumption between features might be a serious restriction.

- Gradient Boosting Algorithm: An ensemble learning approach for classification and regression analysis is called gradient boosting. It functions by sequentially integrating several weak learners, with each weak learner attempting to fix the mistakes committed by the preceding one. Gradient Boosting trains the weak learner to minimize this difference by updating the weights of the data in each iteration using a loss function to assess the difference between the actual and predicted output. The combined guesses of all weak learners result in the final prediction. Gradient Boosting is a strong and flexible technique that typically produces state-of-the-art results in a variety of machine learning applications and can handle complicated non-linear boundaries, but it may also be vulnerable to overfitting if not properly calibrated.

- MLP Classifier: A sort of artificial neural network called a multilayer perceptron (MLP) is utilized for supervised learning in classification and regression analysis. It functions by transferring the input to the output through a non-linear activation function and has three levels: an input layer, one or more hidden layers, and an output layer. Stochastic gradient descent and other optimization methods are used throughout the training stage to update the weights connecting neurons. While MLP is a strong and adaptable algorithm that can handle challenging non-linear bounds, it may also be time-consuming to train and prone to overfitting if improperly regularized. It is frequently done to pre-process the data before training an MLP classifier since it is sensitive to the scale of the input characteristics.

- Random Forest Classifier: A classification and regression analysis technique called Random Forest is applied. It functions by creating a forest of trees by mixing several decision trees. As each tree votes to determine the class label of a new sample, the algorithm randomly chooses a subset of the characteristics to divide on at each node. The average of all predictions from all trees (for regression) or the majority vote (for classification) are used to get the final forecast. As a result of the unpredictability in feature selection and the use of several trees in combination, Random Forest is a quick and efficient technique that resists overfitting.

IV. COMPARATIVE ANALYSIS

This work has several steps and these steps shown in the figure 2. On the data in which the work dealt with was analyzed and to predict possible values using various techniques.

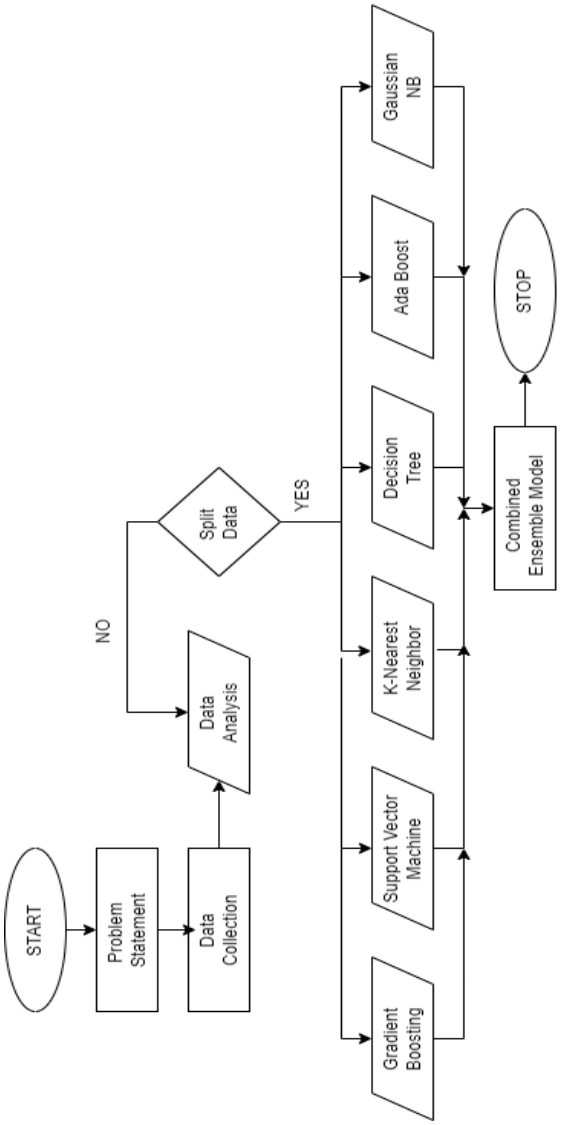

Fig. 2. Flowchart of the proposed work.

- Problem definition and data collection: The first stage is to precisely describe the issue that machine learning (ML) will be used to address, after which you should gather the relevant data to train and test the model. This stage also entails figuring out the model's performance KPIs and the business needs.

- Data preparation and cleaning: In this stage, the gathered data are cleaned, normalized, and divided into training and test sets to be used in the model. This stage also involves preparing the data, which includes feature scaling, addressing missing values, and eliminating outliers.

- Model selection and training: You choose an acceptable ML model in this stage and train it using the ready data.

To enhance the performance of the model, this stage also entails fine-tuning its hyperparameters.

- Evaluation and validation: To make sure the model is correct and effectively generalizes to new data, it must be assessed and verified once it has been trained. To check for overfitting, this phase involves employing methods like cross-validation and testing the model on hypothetical data.
- Deployment and monitoring: The model may be deployed in a production setting and its performance tracked after it has been examined and verified. This stage entails establishing an API for accessing the model, putting up an infrastructure for deploying the model and keeping track of the model's performance in a real-world situation.
- Maintenance and retraining: To enhance the model's performance and allow it to respond to changes in the issue domain, it should be frequently retrained on fresh data. In this stage, the model's performance is monitored, data drift is looked for, and the model is updated to maintain performance

A. Validation

Validation in machine learning is the process of evaluating the performance of a model on a dataset that was not used during the training process. This allows for an estimate of the model's performance on unseen data. There are several techniques for validation, including:

- Holdout validation: The dataset is split into a training and a test set, and the model is trained on the training set and evaluated on the test set.
- K-fold cross-validation: The dataset is split into k subsets, and the model is trained and evaluated k times, with each subset serving as the test set once and the remaining subsets serving as the training set.
- Leave-one-out cross-validation: Like k-fold cross-validation, but with k equal to the number of observations in the dataset, resulting in a training set with one observation left out each time.
- Bootstrap validation: Random samples of the data with replacement are used to train and evaluate the model multiple times.
- Time-series cross-validation: This is a technique used in Time series forecasting, where the data is divided into windows of time and the model is trained on previous time steps and evaluated on the next time step.
- Validation techniques like K-fold cross-validation and bootstrap validation can help to reduce the variance in the estimate of the model's performance, but they will not talk about the model's performance on unseen data.

The figure 3 shows all confusion matrix using SVC, K-Neighbor, AdaBoost, decision tree, Gaussian NB, Gradient Boosting, Random Forest and MLP.

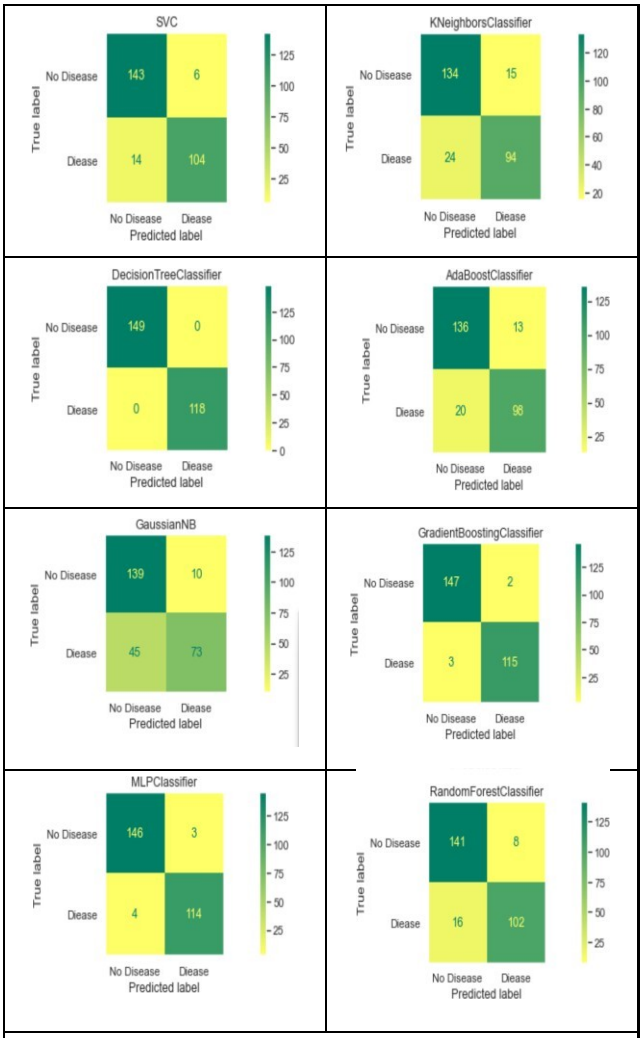

Fig. 3. Confusion matrix attained using different techniques

V. DISCUSSION

The work applied the ensemble learning method to enhance the functionality of our model for detecting heart disease. A machine learning technique called ensemble learning combines many basic models to get a forecast that is more reliable and precise. The ensemble can minimize the volatility and bias of the predictions and enhance performance by pooling the results of many models. To do this, we gathered a dataset of medical records of patients and tested a number of classifiers, such as SVM, AdaBoost, K-Nearest Neighbors (KNN), Decision Trees, Gaussian Naive Bayes, Multilayer Perceptron, Gradient Boosting, and Random Forest.

Accuracy, precision, recall, and F1-score were the metrics employed in the 10-fold cross-validation process used to assess each classifier's performance. The study findings demonstrated the efficiency which was given by ensemble learning methods in detecting cardiac illnesses and the potential it attained to enhance heart disease early diagnosis and therapy. Our findings suggest that Decision Tree Tuning with Support Vector Classifier is an effective classifier for heart disease detection and can be used to improve the early diagnosis and treatment of heart disease. Further studies are needed to validate our results on larger

and more diverse datasets and to explore the use of deep learning models for heart disease detection.

Table 1. Accuracy achieved by different classifier

Model(s)	Accuracy (%)
Gradient Boost Classifier Hyperparameter Tuning with KNN	91.76
Gradient Boost Classifier Hyperparameter Tuning with Decision Tree Classifier	95.13
Gradient Boost Classifier Hyperparameter Tuning with SVC	95.88
Gradient Boost Classifier Hyperparameter Tuning with Random Forest Classifier	95.50
Gradient Boost Classifier Hyperparameter Tuning with Ada Boost Classifier	94.38
Gradient Boost Classifier Hyperparameter Tuning with MLP	95.88
Gradient Boost Classifier Hyperparameter Tuning with Gaussian Naive Base	94.38
Decision Tree Hyperparameter Tuning with SVC	98.12
MLP Hyperparameter Tuning with Gaussian Naive Bias	95.88

Based on an examination of several recent studies that used various data mining and machine learning approaches and algorithms to forecast the development of heart disease. To us, it is different. Weka, MATLAB, and other experimental tools are used to predict cardiac disease using data mining and machine learning approaches. In several trials, distinct patient datasets with heart disease are employed. Most studies use data from the UCI repository's online Cleveland database. The dataset consists of 303 entries with a total of 75 characteristics, 14 of which are critical, and some of which have missing values. On various datasets, fewer experiments have been conducted. According to the study, neural networks with 15 characteristics perform with 100% accuracy in one experiment whereas neural networks with 8 attributes perform with 76.55% accuracy. In most studies with various numbers of characteristics, Naive Bayes also provides excellent accuracy of over (90%). The accuracy of decision lists (J48) in a case increases to 99.62%, which is a very good performance. Therefore, the number of characteristics employed, and the implementation tool used determine the accuracy of the various strategies. We draw the following conclusions from this study, which should be considered in future research for high accuracy and more accurate heart disease detection utilizing intelligent prediction systems. Most trials have trained prediction models using little datasets. So, to train and validate our prediction models, we must use real data from a big number of heart disease patients at reputable medical facilities in our nation. To select the features based on their influence on the patient's health and, if required, add additional crucial heart disease symptoms, we must contact highly qualified cardiology doctors for a more accurate diagnosis.

VI. CONCLUSION

By integrating several data mining and machine learning methodologies, as well as text mining of the unstructured medical data that is easily available in enormous quantities at medical institutions, it is important to develop more complex hybrid models for accurate prediction. Additionally, the employment of genetic algorithms for feature selection and optimization significantly improves the overall performance of intelligent prediction models. In this work, classification approaches were given greater attention than regression and association rules. Therefore, we must take these factors into account to get better comparable results in future studies. The choice of research instruments and techniques directly relates to the accuracy of the research. Therefore, selecting the right experimental instrument for approach implementation is also a crucial consideration.

REFERENCES

[1] R. Indrakumari, T. Poongodi, and S. Jena, "Heart Disease Prediction using Exploratory Data Analysis," Procedia Computer Science, 173, 130-139, 2020.

[2] C. W. Tsao, A. W. Aday, Z. I. Almarzooq, A. Alonso, A. Z. Beaton, M. S. Bittencourt, and American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee, "Heart disease and stroke statistics—2022 update: a report from the American Heart Association,"145(8), e153-e639, 2022.

[3] H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, "Heart disease prediction using machine learning algorithms," In IOP conference series: materials science and engineering ,Vol. 1022, No. 1, 2021.

[4] A. Yazdani, K. D. Varathan, Y. K. Chiam, A. W. Malik, and W Ahmad, "A novel approach for heart disease prediction using strength scores with significant predictors" BMC medical informatics and decision making, 21(1), 194, 2021.

[5] N. Nandal, L. Goel, and R. Tanwar, "Machine learning-based heart attack prediction: A symptomatic heart attack prediction method and exploratory analysis. F1000Research," 11(1126), 1126, 2022.

[6] M. K. Hossen, "Heart Disease Prediction Using Machine Learning Techniques," American Journal of Computer Science and Technology, 5(3), 146-154, 2022.

[7] M. M. Ahsan and Z. Siddique, "Machine learning-based heart disease diagnosis: A systematic literature review," Artificial Intelligence in Medicine, 102289, 2022.

[8] T. Chithambaram, K. N. Logesh, and M. Gowsalya, "Heart Disease Detection Using Machine Learning," Version 1, 2020.

[9] I. Kumar, J. Virmani, H. S. Bhadauria, and M. K. Panda, "Classificationof breast density patterns using PNN, NFC, and SVM classifiers," In Soft Computing Based Medical Image Analysis, pp. 223-243, 2018.

[10] S. Gupta, S. Iyer, G. Agarwal, P. Manoharan, A. D. Algarni, G. Aldehim, and K. Raahemifar, "Efficient prioritization and processor selection schemes for heft algorithm: A makespan optimizer for task scheduling in cloud environment" Electronics, 11(16), 2557, 2022.

[11] C. Krittanawong, H. U. Virk, Z. Wang, K. W. Johnson, R. Pinotti, and W. W. Tang, "Machine learning prediction in cardiovascular diseases: a meta-analysis" Scientific reports, 10(1), 16057, 2020.

[12] K. K. Lee, D. Doudesis, M. Anwar, F. Astengo, C. Chenevier-Gobeaux, Y. E. Claessens, and N. L. Mills, "Development and validation of a decision support tool for the diagnosis of acute heart failure: systematic review, meta-analysis, and modelling study" bmj, 377, 2022.

[13] L. Riyaz, M. A. Butt, M. Zaman, and O. Ayob, "Heart disease predictionusing machine learning techniques: a quantitative review." In International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021, Vol 3, pp. 81-94, 2021.

[14] R. Yunus, U. Ulfa, and M. D. Safitri, "Application of the K-Nearest Neighbors (K-NN) algorithm for classification of heart failure," Journal of Applied Intelligent System, 6(1), 1-9, 2021.

[15] N. Mohd, J. Sharma, and D. Upadhyay, "A survey: heart disease prediction using machine learning techniques," Webology, 18(4), 2426-2439, 2021.

[16] I. V. Derevitskii, D. A. Savitskaya, A. Y. Babenko, and S. V. Kovalchuk, "Hybrid predictive modelling: Thyrotoxic atrial fibrillation case," Journal of Computational Science, 51, 101365, 2021.

[17] I. Kumar, J. Virmani, H. S. Bhadauria, and M. K. Panda, "Classification of breast density patterns using PNN, NFC, and SVM classifiers," In Soft Computing Based Medical Image Analysis, pp. 223-243, Academic Press, (2018).

[18] I. Kumar, A. Kumar, V. D. A. Kumar, R. Kannan, V. Vimal, K. U. Singh, and M. Mahmud, "Dense tissue pattern characterization using deep neural network," Cognitive computation, 14, 1728-1751, 2022.