Exploiting Machine Learning Models for Identification of Heart Diseases

Inam Ullah[1], Azhar Imran[1, *] Areesa Ashfaq[1] and Muhammad Adnan[1]

[1]Department of Creative Technologies, Air University, Islamabad, 44230, Pakistan.
*Corresponding Author: Azhar Imran. Email: azhar.imran@mail.au.edu.pk.

_____

Abstract: Cardiovascular illnesses have surpassed disease as the leading cause of death in industrialized, emerging, and underprivileged countries in recent decades. The mortality rate can be lowered through early detection and effective care of cardiac diseases. However, reliable detection of heart disorders in all conditions and serving for doctor's 24-hour medical consultations are not possible since they require more intelligence, time, and talent to train the computer for a specific duty. Artificial intelligence and machine learning algorithms may have a substantial impact on the lives of those who are afflicted with chronic diseases. In this work, we used artificial intelligence to create a hybrid model for the prediction of cardiac disease and flags for prevention. The goal of this research was to develop a prediction model that can identify and characterize a patient's illness. We used data of 68975 patients from the University of California, Irvine's repository and trained several algorithms for cardiovascular disease prediction. In our simulation maximum classifiers have achieved the greatest accuracy, which is a huge accomplishment for our research team. Our proposed hybrid Machine Learning algorithm framework improved heart disease prediction performance when compared to previously offered prediction models. The proposed methods are decision tree and random forest models, which have a 99.980% percent accuracy for training, followed by support vector machine, which has a 99.30 percent accuracy followed by other classifiers which yield higher results in comparison. Existing Cardiovascular Diseases history was considered the most important factor determining the prediction model's accuracy. Our findings reveal that our Cardiovascular Diseases prediction model based on machine learning techniques developed for health screening datasets is simple to apply and more accurate. Our proposed classifiers have the best accuracy rate.

Keywords: Cardiovascular Diseases; Prediction; Heart diseases Prediction; Machine learning.

1. Introduction

Cardiovascular diseases (CD) are a severe health issue that affects individuals all over the world. CD kills about 647,000 Americans each year, according to the Centers for Disease Control and Prevention (CDC) [1]. Metabolic abnormalities like IGT, dieting carbohydrates, or chronic diseases, are tracked by two of the three main disorders: overweight, Body - mass in males as well in females, hyperglycemia, triacyl glycerol or HDL cholesterol threshold causes heart disease, furthermore, in the American Heart Association (AHA) estimates that 130 million individuals in the United States will have some kind of cardiovascular disease by 2035 [1, 26]. According to the Centers for Disease Control and Prevention (CDC) and the American Heart Association (AHA), Each year, the United States spends $219 billion, with spending an estimated $1.1 trillion by 2035 [2]. From the World Health Organization (WHO), 17.9 million people worldwide die each year from heart disease [3]. The most significant behavioral risk factors for CD and stroke include age, sex, poor eating, sedentary lifestyle, cigarette smoking, and problematic alcohol use [5]. As a

result of behavioral risk factors, people may have increased blood pressure, blood glucose, or blood lipids [4]. Chronic diabetic syndrome affects 64 out of 1.32 billion people in the United States over the age of 20.

The metabolic syndrome raises the risk of cardiovascular disease (CD), Yet, in a later joint report, the American Heart Association (aha and the Europe Association for the Study of Diabetes questioned its therapeutic efficacy First, this remark highlights that metabolic syndrome is an ill-defined condition with no established utility as a model for estimating potential Cardiovascular. Second, it raises a worry about the possibility of clinicians being misled in the management of those with one or deuce CD risk factors. [5, 26] Finally, it accepts that this condition is useful in predicting diabetes, yet disputes whether it has predictive ability over glucose intolerance. Metabolic syndrome has been associated with a higher risk of heart disease, particularly in men and women over 45. The metabolic syndrome, in addition to glucose intolerance, predicts diabetes. Machine learning approaches enable the enrichment of knowledge-based decision support that extracts hidden patterns in medical domain data [6, 25].

Institutions gather a large quantity of data regarding disease diagnostic testing, patients, and many other concerns nowadays. Machine learning is a set of techniques for detecting hidden patterns or correlations in data. As a result, and machine learning algorithm is being suggested within that study for the such invention of a prediction of heart disease, which has been validated and uses fixed automat diagnostic samples. Machine learning approaches help healthcare professionals diagnose diseases more accurately and quickly in the early stages. Initiated an algorithm to predict the presence of heart disease that uses classification methods of learning algorithms along with a chosen heart disease data set classification [7]. In [8, 23] mainly classifiers for the prediction of HD (heart diseases) have been used. where authors proposed to develop an application that can forecast a risk of heart disease given basic symptoms like age, sex, and pulse rate, for which neural networks showed the most accurate and reliable results. The focus of this research is to see if the patient is at risk for cardiovascular heart disease based on their medical characteristics such as glucose, diastolic blood pressure, BMI, and sugar level. The data set includes the medical records of a patient. Researchers are using this information and determine whether or not the patient may suffer from heart disease. To do so, we categorize and analyze data on 14 features to see if he is at risk for heart disease. The hybrid Classifier chain is used to train on these medical features.

The development of good medical diagnosis processes for a range of illnesses takes a lot of time and work. Logistic regression, KNN, and Random Forest Classifier are all covered in [9]. The highest level of accuracy was 87.5 percent. The development of good medical diagnosis processes for a range of illnesses takes a lot of time and work. deployed an Intelligent Heart Disease Prediction System in the year 2020 that includes eight different classifiers to predict heart disease, we identified several limitations since they used basic classifiers, and the majority of them were based on decision trees [10]. Predict cardiac illness in the year 2020 using four different classifiers, none of which had an accuracy higher than 95% [11]. The research methodologies authors used in this study have the potential to save lives from the CDC.

To improve prediction accuracy and explore events related to heart disease, a predictive analytic framework was designed using a Random Forest classifier, and experimental results showed that classification using the Classification algorithm can be successfully used in predicting events and risk factors related to HD [6,12]. To identify cardiac sickness a decision Tree employs clustering methods (K-Means) [13]. The inlier approach involving clusters produced a classification accuracy of 83.9 % within the diagnosis of heart disease. In his article, used data mining approaches to predict cardiac disease using an intrinsic neural network [14]. In detection, he evaluated main risk indicators such as age, diabetes, intake of alcohol, obesity, or physical handicap, among others. This solution achieved an accuracy result of 89%. Suggested a data mining-based technique for classifying heart disease. The study found that Decision Tree has 95.29 percent sensitivity with 98.01 % accuracy when employing sensitivity and accuracy markers for evaluation [16, 17]. Bayesian classifiers, on the other hand, scored 87.10 percent as well as 91.30 %, accordingly [15]. The results of the previous study revealed that the Decision Tree outperformed the alternative classifier.

2. Materials and Methods
2.1 Methodology
The most sophisticated models from the most popular models are built and tweaked (optimized) for the Comparison of the best models for each category in Figure 1. We start by importing libraries and then follow them for checking in Figure 1, which shows basic processes through the proposed approach.

Within, data pre-processing is used to remove incomplete information and handle irregularities from patient data obtained from a reliable source. The feature selection approach is used on the data set to assist the classification by choosing suitable as well as meaningful characteristics. Finally, these selected features are used to predict one of the types of common heart diseases. The authors go through the details from each step

The data set is then divided into two sections Training and testing data set in training we train our models initially to learn from patterns from the data set and then check the result of the training by cross-validation. The authors continuously checked the accuracy of the training and testing for each classifier while changing the learning rate and folds where the classifiers required.
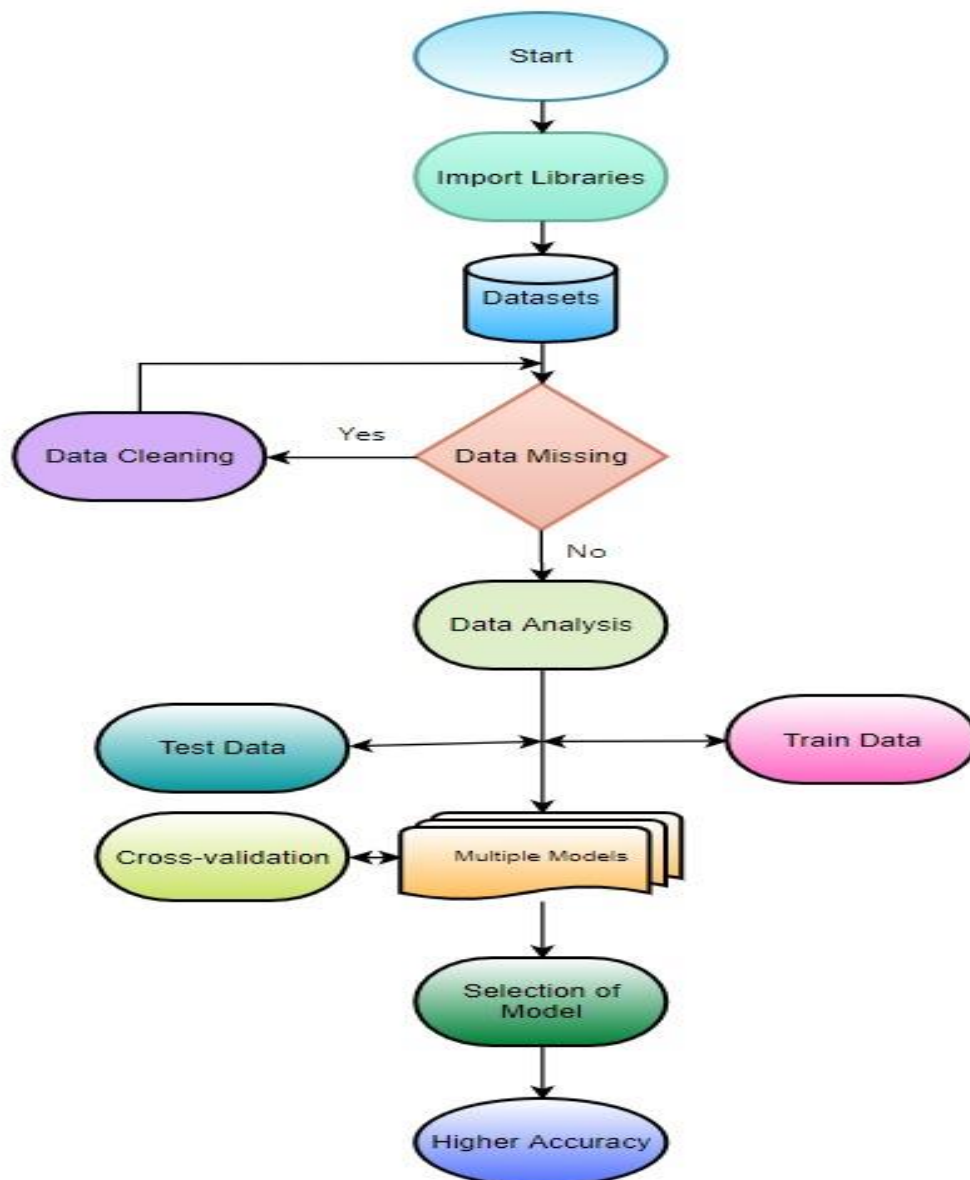

Figure 1. Flow Diagram of Proposed Model

2.1.1 Data Source

This Heart Disease data set comes from the University of California, Irvine's repository. There are 68983 patients and 14 characteristics in the whole data set. The data is divided into two parts a training set (70%, N = 55180) and a test set (30%, N = 13795).

Table 1. Detail of Data set

Features:	14	Boolean:	4	Categorical:	3
Numeric:	7	Patients:	68975	Missing:	8

The ratio between insufficient and sufficient responders was preserved identically to the ratio in the entire data set using a stratified split. Table 1, contains the attributes and other characteristics of the data collection

2.1.2 Pre-Processing

Because the data in the data set had many scales, the Min-Max normalization strategy was used to scale the large continuous data. Consists of removing the minimum and dividing the data throughout the whole data range to linearly convert the data. The data is normalized and mapped to a 0 to 1 range, which helps the machine learning model develop a clearer trend between data and normalize the effect of numerous parameters as a result of Eq. (1)

$$Y = \frac{X - min}{max - min} \qquad (1)$$

Data prepossessing given in Figure 1, is a step in the data mining and analysis process that converts raw data into a format that computers and machine learning can comprehend and analyze. Machines don't understand text, pictures, or video. Noises, missing data, and other variables are all aspects to consider. in real-world data prevent ML models from being employed directly. As a result, following given Figure 1, data pre Treatment is necessary to clean the data and make it acceptable for an ML model, increasing the model's accuracy and efficiency.

2.1.3 Quality Assessment of Patients Data set

Assessment of Data, Locating and Compiling Data that is missing, inconsistent, or duplicate, Mismatched of Data, Mixed Data Values or irregularities Outliers in the data set

2.1.4 Transformation of patients' Data set

Aggregation of patient Data set, Normalization of patient Data set and Feature Selection or Sampling from patient Data set.

2.1.5 Feature Encoding

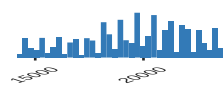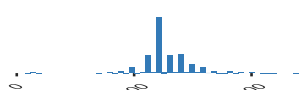Labeling of the patient data set and One-Hot Encoding of the patient data set features.
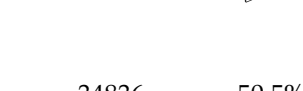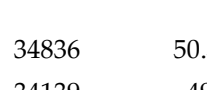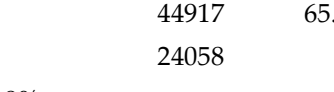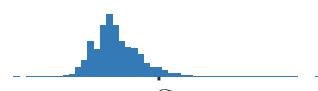
2.1.6 Dimensionality Reduction

Eliminate multi-collinearity to improve the interpretation of the machine learning model's features. Such observations become clearer to visualize.

2.1.7 Feature Scaling of the Data set

Standardization of Data set, normalization of Data set. Now data set has no irregularity and is ready for modeling, ready to train a model to forecast the needed outcome. There are more than 60 predictive modeling algorithms from which to pick. However, researchers reduce and select a few models to analyze, first iden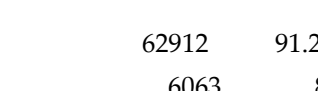tifying the type of problem and solution requirements. The issue is categorization and regression. We want to see if there's a link between output (H-Disease affected or not) and other variables or characteristics (Gender, Age, Alcoholic.). We're using supervised learning, which is a type of machine learning where we trained our model using a set of data. For all that we understand our data set.

Table 2. Exploratory Data Analysis

Features	Feature details	Data Type	Toggle Details		
Age: (In days)	Maximum: 10798 Unique: 11.7% Minimum: 10798	Type= Numeric Zeros: 0%			
Active: Physical activity:	Distinct count: 2 Unique: <0.1%	Type = Boolean Zeros: 0%	1 0	65277 13569	80.3% 19.7%
ap_hi: Systolic blood pressure:	Maximum: 240 Unique: 0.1% Minimum: 7	Type = Numeric Zeros: 0%			
ap_lo: Diastolic blood pressure	Distinct count: 94 Maximum: 190 Unique: 2% Minimum: 7%	Type = Numeric Zeros: < 0.1% Mean: 81.34			
Bmi: Body Mass Index:	Distinct count: 3782 Maximum: 298.66 Unique: 5.5% Minimum: 3.4717	Type = Numeric Zeros: < 0.1% Mean: 81.335			
Cardio: Patients	Distinct count: 2 Unique: < 0.1 %	Type = Boolean Zeros: 0%	1 0	34836 34139	50.5% 49.5%
Alco: Alcohol intake	Distinct count: 2 Unique: < 0.1 %	Type = Boolean Zeros: 0%	1 0	55406 1356	80.3% 19.7%
Cholesterol level	Distinct count: 3 Unique: < 0.1 %	Type = Boolean Zeros: 0%	1 2 3	51734 9430 7901	75.0% 13.5% 11.5%
Id: Patient's identification number	Distinct count: 68967 Maximum: 68975 Unique: 100.0%	Type= Numeric Mean: 34991.12 Zeros: <0.1%			
SEX: Male or female	Distinct count: 2 Maximum: 68975 Unique: <1.0%	Type= Categorical Zeros: 0%	1 0 34.9%	44917 24058	65.1%
Gluc: Glucose level	Distinct count: 3 Unique: <1.0%	Type=Categorical Zeros: 0%	1 2 3	58639 9430 5248	85.0% 7.4% 7.6%
Height: In (cent meter)	Distinct count: 108 Maximum: 68975 Unique: 0.2%	Type=Categorical Mean: 164.36 Minimum: 55			
Smoke: Smoking addicted	Distinct count: 2 Unique: < 0.1%	Type = Boolean	0 1	62912 6063	91.2% 8.8%
Weight: In(kg)	Distinct count: 281 Maximum: 200 Unique: 0.4%	Type= Numeric Minimum: 11			

3. Exploratory Data Analysis

The first thing we do after importing a data set is to learn as much as we can about it. This library creates a full report for our data set, which contains the information listed in statistics Table 2, that are descriptive (mean, average, etc.) Toggle Details is a type of statistic that is used to tell us about how our data is distributed

3.1. Correlation Analysis

Correlation is a measure of how closely qualities are connected. Furthermore, to offer a clearer view of the feature correlations between each of the features, a heat map illustrating the correlations between all qualities is provided by Pearson's R correlation:

$$r = \frac{\sum_{i=1}^{n}(x - \bar{x})}{\sqrt{\sum_{i=1}^{n}(x - \bar{x})^2}} \frac{(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{2}$$
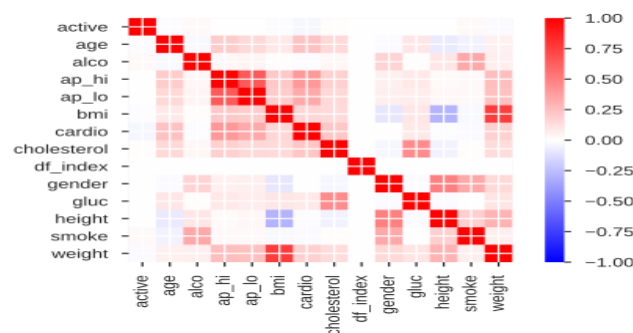

Figure 2. Pearson's correlation of Features

Pearson's correlation from coefficient (r) is a measure of several parameters' linear correlation. The rating ranges from -1 to +1, where -1 represents in Figure 2. Net negative linear correlation, "0" represents none linear correlation, and 1 expresses positive linear correlation. Moreover, from Eq (2). r seems unaffected by changes in the location and scale of the two variables, meaning that the angle to the x-axis has no impact on r since a function. Figure 3, of the correlation matrix with all data, tells that Age, Gender, glucose, smoke, physical activity, weight, Alcohol intake, gender, BMI, and all others in the matrix are the most related features of heart diseases. The most common detrimental health effect is smoking.
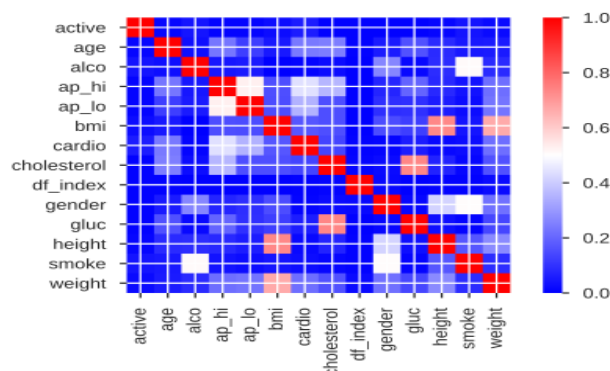

Figure 3. Correlation Heatmap of all Features

$$\emptyset_c = \sqrt{\frac{X^2}{N(k-1)}} \qquad (3)$$

Wherefrom Eq (3), ϕc stands for Cramer's V, x^2 is the Pearson chi-square census from the forenamed test N is the total number of persons who had included in the test k is the variable that has the fewer categories or is variable. In figure 4, glucose and cholesterol are positively correlated.
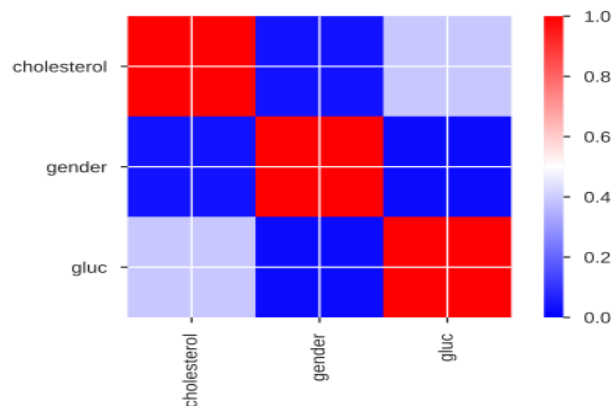

Figure 4. Cramer's correlation of CD

4. Machine Learning Algorithms

To validate the model's performance, we deployed machine learning methods to the UCI Machine Learning Repository Data set. Only the classifier that has the high accuracy is taken into consideration for performance assessment among those, five algorithms with more accuracy (i.e., more than 90%). Each of the used algorithms is briefly discussed below Logistic classifier is a Machine Learning technique for resolving categorization problems [18]. It's a predictive analytic approach based on probability. For feature creation and completion of targets, our team utilized to assess our assumptions and results, and find the coefficient of the features in the decision function from Eq. (4) Positive coefficients increase the log-odds of the response and therefore increase the probability, Negative coefficients reduce the response's log-odds and hence decrease the probability. The function of cost is

$$-\log\big(h\theta(x)\big)\ if \qquad y = 1 \qquad (4)$$
$$-\log\big(1 - h\theta(x)\big)\ if\ y = 0$$

The main scoring coefficient is glucose, height as a lower bound, and active at the upper bound.

4.1 Decision Tree Classifiers

One of the most significant and widely applied approaches in both classification and forecasting is a decision tree. A decision tree is indeed a flowchart-like branching structure that contains a node and identifies a trial, every branch presents its algorithm's finding, and every leaf node (terminating node) holds its target class [19]. The tree classifier terminology is as follows:

1. **Root Node:** It represents the complete folk or sample, which is then broken down into two further correlative groups.

2. **Decision Node:** It starts when a sub-node is categorized into several integral nodes.

3. **Parent Node**: It is an integral node that splits into sub-nodes

4. **Child Node:** When **the** parent node is divided, the integral nodes are the children's nodes

5. **Leaf Node**: It can also say that terminal node are nodes that do not divide.

One of the best classifiers so far is the decision tree which yields the following score evaluated on the train and Test data set respectively.

Train Accuracy=99.98%

Test Accuracy=64.61%.

4.2 Support Vector Classifier

Support Vector Classifier is a type of supervised learning model that explores data used for classification, regression, and analysis while non-linear SVC performs classification by finding the hyperplane that amplifies it, hyperplanes are decision borders that aid in the classification process. A range of techniques might be assigned to data points along each side. The hyperplane's dimension is also based on the number of attributes.

Train Accuracy=72.13%

Test Accuracy=72.76%

4.3 Soft Voting Classifier

Soft voting classification the incongruity to majority voting delivers the class as the argmax of all projected probabilities. Each classifier allocates a probabilistic value of each data locus that it links to a given target. The predictions are mean and freighted on the report of the connection of the classifier. The vote is then cast for the target label with the largest sum of weighted probability**.** The Predicted probability method should be included in all basic models. Because it incorporates the predictions of several models, the voting classifier process better overall results than other rooted models.

Train Accuracy=83.19%

Test Accuracy=72.78%

4.4 Hard Voting Classifier

A voting classifier is a meta-based classifier that worked on a majority vote to combine similar or conceptually perceptible machine learning models for prediction [18]. All classifier poll for a class in hard voting (leading voting), so the class that has the majority votes wins. In statistical language, the corps expected target label is the mode of the classification of individually estimated labels. In this classifier, the concluded prediction is made via majority polls, in which the collector picks the class that appears repeatedly among the base models. A classifier like this might be beneficial for a group of models that are all doing well to balance out their particular flaws. The estimated accuracy from different classifiers is given which is the base for the result of the hard voting classifier, Logistic Regression, Random Forest, and AdaBoost Classifier.

Train Accuracy=77.62%

Test Accuracy=73.16%

4.5 XG Boost Classifier

A Gradient Boosting classifier is a decision tree embedded in a packed ML algorithm that uses a gradient booster as a foundation [20]. XG Boost is the most popular way to design accurate models using formal data, often known as tabular or structured. We use it to make our models better. We'll use 10-folded cross-validation to fine-tune the XGB Classifier model's parameters. has a uniform learning rate of eta is "o", "0.5" and "0.005" same for gamma while the maximum depth is 2,12, and child weight=1,9,0.025 and at minimum depth is "10".

Train Accuracy=80.32 %

Test Accuracy=73.09%

4.6 Neural Network-1

NN is more composite and more significant and standard machine learning, linked to deep learning models. For building a NN we used Kera's API. While Kera's is a high-level API of Tensor Flow, structured by Google. Kera's allows you to inaugurate neural networks by connecting blocks. we initialize ANN as sequential and activation function 'relu' with the input of the first layer of units '32'. followed by the second layer of Dense units '64' while rate '0.5' in a training session of ANN epochs up to "100" with **a** batch size of "16" while keeping the learning rate to LR "0.001" according to the data we Train on 44144 samples, validate on 11036 samples accuracy we get for this module is 50.67 which is not a good approximation for that particular data set.

Train Accuracy= 50.67%

Test Accuracy= 49.85%

4.7 Neural Network-2

The working method is the same as that of the Neural Network 1 model is sequential, we used a dense layer of 16 at input followed by 64 then 32 and the outer layer is single. The first 3 layers have an activation function of "relu" while outputted has "sigmoid" it gives a score that is the same as Neural Network 1

Train Accuracy=50.67%

Test Accuracy=49.85%

4.8 K-Nearest Neighbors Algorithm

K-Nearest Neighbors classifie**r** (KNN) we can use for classification and also for regression. Classifications were done by a majority voting of its nearest neighbors, we keep the K of 10-fold of the neighbors, which Calculate the Euclidean distance between K number of neighbors Take the K nearest neighbors as per the calculated Euclidean distance by Eq. (5). Among these k neighbors, count the number of the data points in each category update new points category-wise for whom the number of the neighbor is highest.   Our model Accuracy is:

Train Accuracy=81.52%

Test Accuracy=66.87%

$$d = \sqrt{(x_2 - x_1) + (y_2 - y_1)} \tag{5}$$

While 'd' is the Euclidean distance.

4.9 Naive Bayes

Naive Bayes classifiers are important probabilistic classifiers in machine learning and are created using the Bayes theorem and significant (naive) isolation hypotheses between features. Because the number of parameters required is proportional to the number of variables (features) in a having-to-learn assignment, Naive Bayes classifiers are easily customized. These classifiers are a series of Bayes' Theorem Eq (6), based on classification techniques. It is a group of algorithms that share a similar idea or maximum likelihood estimation, namely that each item of characteristics being categorized is autonomous of the others, as by Bayes' Theorem from Eq (6).

$$P(A|B) = \frac{P(B|A)*P(A)}{P(B)} \tag{6}$$

Where 'A' and 'B' is the event and P(B) is not equal to zero

Naive Bayes classifiers accuracy for our data:

Train Accuracy=70.27%

Test Accuracy=70.26%

4.10 LGBM Classifier

The Light Gradient Boosting Machine approach divides its tree leaf-wise with the greatest fit adopting decision tree approaches, while other boosting techniques split its tree depth-wise or level-wise rather than leaf-wise. As a result, the Light GBM algorithm achieves better accuracy, which is difficult to achieve with any of the current boosting approaches [13]. We're utilizing and fine-tuning the model parameters using cross-validation of the LGBM Classifier model. Number of folds is 10, learning rate = 0.0462, maximum depth is 9, number of leaves =1570, N estimator=470, min child weight =3.725, col sample by tree= 0.935, so the resultant accuracy is:

Train Accuracy=72.71%

Test Accuracy=73.00%

4.11 Multi layers Perceptron

The Perceptron is a supervised learning technique for binary classifier features that determine if a vector of numeric values the input conforms to a particular category or not. It's a classification technique or a type of linear classifier. To make accurate predictions, a linear predictor function is also used to connect a weight value with the vector. The method enables distance learning since it analyses variables in the learning set one at a time. The major components of the multilevel perceptron are shown in Figure 5.
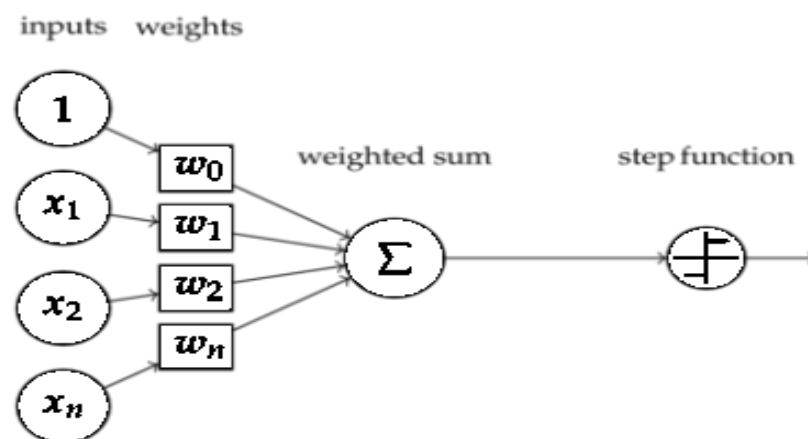
Figure 5. Multi-Layer Perceptron

1. Input

A perceptron uses all of the characteristics as input. The input of a perceptron is represented by [x1, x2, x3,...,xn], where x is the value of a feature while n represents the total number of attributes We also have a unique type of input known as bias. The value of the BIAS is depicted in Figure 5, as w0.

2. Weights

The values that are calculated during the model's training. Weights are first set to an initial value, and these values are modified with each training mistake. The weights for perceptron are portrayed by [w1, w2, w3,... wn].

3. Bias

In a classifier, its biased neurons enable the boundary to be changed to the left or right. Bias neuron also grants a classifier to change its boundary of the decision in algebraic terms. Its goal is to "move every point in a defined direction at a consistent distance." Bias aids in the quicker and more accurate training of the model.

4. Weighted summation

The total of the values obtained by multiplying each weight [wn] associated with each feature value [xn] is known as weighted summation. For all [i to n], we use wi xi to express the weighted summation.

5. Step/activation function

Activation functions are responsible for making neural networks nonlinear. For example, in linear classification, it is crucial to make the perceptron as linear as feasible.

6. Output

The step/activation function receives the weighted summation, and the value we obtain after computing is our expected output

Train Accuracy=62.3%

Test Accuracy=72.41%

4.12 Stochastic Gradient Descent

SGD (Stochastic Gradient Descent) is just a simple and efficient method for fitting linear classifiers as well as regressors to convex attributes like (linear) Support Vector Machines as well as Logistic Regression.[21] Ignoring the fact that SGD has been in use for a long period in the field of artificial intelligence, this has only received much interest in the case of larger learning. In-text classification and natural language processing, SGD has been used to hold big, dense machine learning issues. Our motivation for using stochastic optimization techniques stems from the fact that while training deep learning models, the optimization algorithm is frequently treated as a summation of a definite range of functions

$$\mathrm{f}(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} f_i(\boldsymbol{x}) \qquad\qquad (7)$$

The loss function f i (x) is derived upon in Eq. (7) that training data point sorted by i. It's worth noting that in gradient descent, the per-iteration processing cost grows linearly with the scale of the data set. As a result, when n is large, the gradient descent per-iteration computing cost seems to be quite large. As a result, stochastic gradient descent is a more minimal approach. Instead of determining the gradient f(x)f(x) at each phase, stochastic gradient descent randomly selects I at homogeneous and evaluates fi(x)fi(x). The point is that stochastic is a beneficial move. Although f(x)f(x) is an unbiased predictor of f(x)f(x), from Eq (8) learning algorithm employs fi(x)fi(x) as an unbiased estimator.

$$\mathrm{E_i} \, \nabla \, \mathrm{f_i(x)} \; = \; \frac{1}{n} \, \sum_{i=1}^{n} \nabla \, \mathrm{f_i(x)} \; = \; \nabla \, \mathrm{f(x)} \qquad\qquad (8)$$

The characteristics of Stochastic Gradient Descent are competence. The simplicity of deployment there are quite a lot of options for quite well the code). Stochastic Gradient Descent has several drawbacks: A lot of parameters are demanded SGD, including the parameter value, and the number of iterations' is affected by features scalability. It gives us its best accuracy because our data set is a huge number of observations and obtained

Train Accuracy=70.01%

Test Accuracy=70.27%

4.13 AdaBoost Classifier

AdaBoost (Adaptive Boosting) is a prominent boosting approach that seeks to construct a strong classifier by merging many weak classifiers. A single classifier may not always be able to effectively forecast an entity's category, so we establish a stronger method that combines multiple weak classifiers, each learning from all the others' incorrectly classified instances. Any of your fundamental classifiers, from decision Trees (typically this same usual) through Logistic Regression, would be the classifier referenced here. AdaBoost may be placed at

the top of any classifier to learn from its deficiencies and recommend a rather more appropriate prediction, rather than just being a model by itself. For this purpose, it's often alluded to as the "highest in classifiers." Stumps aren't a smart method to make decisions on their own. A stump, on either end, can only emphasize one factor. A glance at the AdaBoost algorithm's insides individually, looking at various factors to see if an individual is "healthy" (of healthy life) or whether. AdaBoost is a blessing whenever it comes to order to improve the accuracy of our different classifiers. The accuracy of the classifier for our data set is with a learning rate of 0.0189,

Train Accuracy=71.28%

Test Accuracy=71.53%

4.14 Ridge Classifier

The Ridge classifier is a model optimization technique that can be used to interpret data with co-integration. L2 normalization is obtained by using the methodology. Once there is a concern in co-integration, least-squares are indiscriminate, and deviations are large, the estimated quantities often distant out from absolute measurements. It's most popular regression approach for predicting a solution to such an expression without any different solution is indeed the ridge classifier

$$\text{Min}\left(\left|\left|Y - X(\theta)\right|\right|_2^{\,2} + \lambda\left|\left|\theta\right|\right|_2^{\,2}\right) \tag{9}$$

The alpha parameter is used to indicate this λ in Eq. (8) of The Ridge regression costs equation, we are essentially manipulating the penalty term by simply adjusting the values in alpha. The greater the alpha coefficient, the greater the penalty, and hence the size of the coefficients is lowered.

Train Accuracy=72.34%

Test Accuracy=72.58%

4.15 Random Forests

Random Forest is an essential technique that may be used both for ordered and recurrent problems. It can build or iterative estimate a single tree and also create a large classification tree in a short period by employing classifications. Deny the reality as its algorithms tend to just be overfitting, independent decision trees are nonetheless beneficial for decision trees. The random forest seems to have a lower average over gradient boost trees, denying the reality that this just utilizes huge peripheral decision trees. Irrespective, the sort of data would have an impact on how it was revealed. For our data analysis, we utilize a 300-point predictor and K fold =5, which yields

Train Accuracy=99.98%

Test Accuracy=71.58%

4.16 Gradient Boosting Classifier

Gradient boosting is an artificial intelligence algorithm that overcomes repetition and shifts constraints. It is being used to develop a prospective system by integrating prior models plus decision trees. This then evolves the models in a segmented, distinct style, and concludes the concept by providing modifications on additional operations at every splitting, its attributes are often analyzed by chance, like all other innovative tools. As a consequence, even if the training data and maximum attributes are the same, the best-recognized splitting may be different if such requirements for a few splits specified during the find the best split are similar. To accomplish strategy throughout matching, its randomized value should be maintained. in, Eq (10), To improve Fm, added the estimator's hm(x) to our analyses. Thus

$$F_{m+1}(x) = F_m + h_m(x) = y \tag{10}$$

Maximum depth=9, n estimators=763 the resultant accuracy is:

Train Accuracy=97.52%

Test Accuracy=71.34%

4.17 Logistic Regression

Logistic regression analyzes the relationship among a categorical variable (aspect) or one or even more independent variables through performing simulations using a logistic function, which may be the cumulative logistic distribution (features). The logistic regression model has been based on the likelihood of just a 2-level outcome measure. For such purposes of classification, I'll suppose that each of the progressions has indeed been chosen as that of the event occurrence, then I'll simply refer to the event in the following statement [22]. This venue's opportunities are determined by dividing the likelihood of an event that occurs by the probability of the event not occurring. Gambling often uses possibilities, with "equal actual chances" (odds=1) meaning that the event will happen half of the time. We used Logistic Regression to evaluate our hypotheses and conclusion for features development and execution targets. To accomplish this, get the coefficients of both the attributes therein the decision function. The estimated coefficient improves the response's log-odds (and thus improves its probability), while negative factors lower the response's log-odds (and thus reduce the probability) The highest coefficient is glucose. By absolute value, the greatest numbers are height and smoke so for the Data set

Train Accuracy=72.4%

Test Accuracy=72.85%

4.18 Support Vector Machine

Linear Support Vector Machine (SVM) is indeed a classification of linearly separable data, Figure 6, suggests that a data set can be divided into two classes using only a single solid line, and thus the classifier is named Linear SVM.

SVM is a supervised learning model that uses a classification method to solve 2 different classification and regression problems. After being provided sets of labeled training data for each category, SVM models may classify new text.

A support vector machine utilizes those sets of data and develops the hyperplane (which is effectively a vector in two dimensions) that best distinguishes the tags [23,24]. In Figure 7, the line represents the decision boundary.
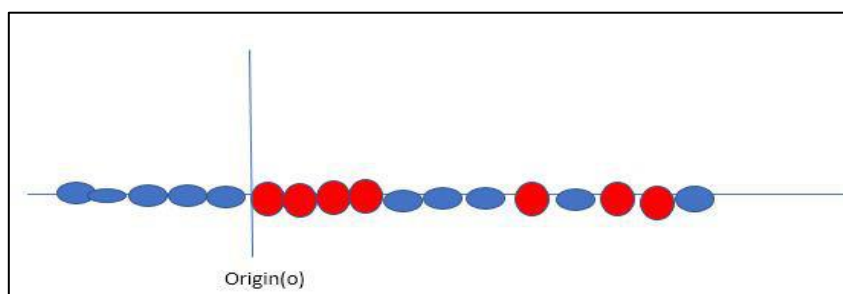

Figure 6. Linear SVM

After that, the hyperplane may be utilized to make predictions. is the h hypotheses function: After that, the hyperplane can be utilized for production by using Eq. (11) just the hypotheses assumption

$$h(xi) = \{+1 \ \ if \ W\char`^T + b \geq 0$$
$$h(xi) = \{-1 \ \ if \ W\char`^T + b \leq 0 \hspace{3cm} (11)$$
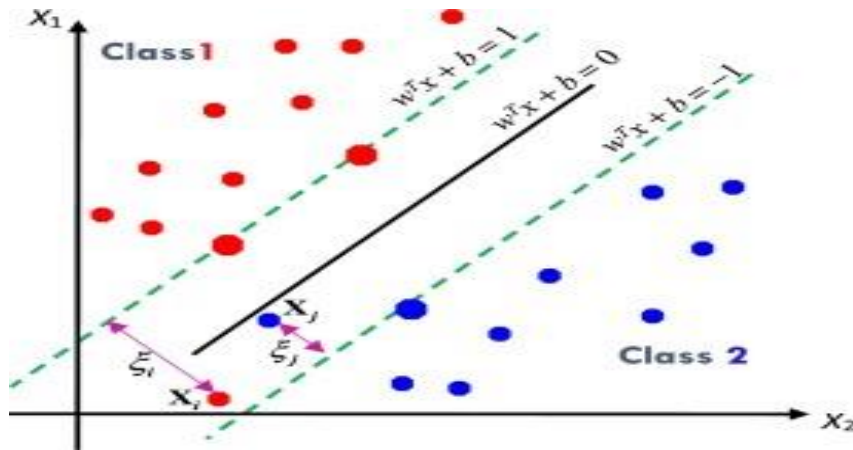$$h(xi) = \{0 \ \ if \ W^T + b = 0$$

Figure 7. Hyper plane of SVM

Length of a vector: The standard deviation of a vector x, represented by ||x|| in Eq (12), is the length of the vector. Use the Euclidean norm formula. X= (x1, x2……)

$$||X|| = \sqrt{x_1^2 + x_2^2 \dots \dots} \hspace{3cm} (12)$$

The direction of a vector: From Eq. (13 & 14) W is the direction of a vector X= (x1, x2),

$$W = (\frac{x_1}{||X||}, \frac{x_2}{||X||} \hspace{3cm} (13)$$

$$\cos(\theta) = \frac{x_1}{||X||} \hspace{1cm} , \cos(\alpha) = \frac{x_2}{||X||} \hspace{3cm} (14)$$

SVC is an approach that is similar to SVM. It, too, SVM is based on kernel functions, but it's designed for unsupervised learning. Experimental accuracy of tests and training comes out:

Train Accuracy=99.31%

Test Accuracy=57.86%

4.19 Extra Tree Classifier

Extra Trees Classifier is an adaptive supervised classifier based on decision trees.   Extra Trees Classifier, like Random Forest, makes some data choices and subgroups randomized to avoid over-learning and overfitting [18, 22]. Extra Trees is related to Random Forest in that it yields multiple trees and subdivides nodes using random parameters, although there are multiple major differences: it doesn't bootstrap findings (indicating it measures without partial substitute) and endpoints are divided using random splits rather than optimal splits. Bootstrap is the default value for numerous tree buildings, suggesting that it samples while replacing. Random splits among a random subset of the attributes offered at each node are being used to separate nodes. The randomization of Extra Trees occurs via random divisions of all findings, rather than data set bootstrapping. The

word "Extra Trees" refers to "additional trees". The default values for factors that influence tree size (e.g., max depth, min samples leaf, etc.). For our data set, we used K-Fold Cross-validation is 10, estimators=146 maximum depths=3, as a result of the extra tree classifier:

Train Accuracy=71.99%

Test Accuracy=72.42%

4.20 Bagging Classifier

Heterogeneous machine learning is categorized into two types: bagging and boosting. The bagged approach helps for both regression and statistical classification. Bagging is a mechanism that's used in decision trees to improve model consistency by increasing accuracy and reducing variance. Boosting is a homogeneous weak learner model that differs from Bagging in how it differs. Learners in this approach acquire sequentially as well as adaptively to enhance a learning algorithm's prediction accuracy. Thus, for us, overfitting is no longer a problem.

Train Accuracy=98.02 %

Test Accuracy=69.11%

5. Results and Analysis

As seen in Figure 10, the correlation matrix and cholesterol have such a significant impact, even though the correlation with the target class isn't especially strong. Unless we evaluate the average BMI for healthy persons to the same average BMI of ill people in Figure 9, we can see that the latter is higher. From Eq. (15) normal BMI range is stated to be between 18.5 and 25. Depending upon their BMI,

$$\text{BMI} = \frac{Mass_{kg}}{heigh_m^2} \qquad (15)$$

From Figure 8, alcoholic women have a greater HD risk than drinking males. As can be seen, age and cholesterol have a substantial influence, although the correlation to target class isn't particularly strong. Patients with CD have increased cholesterol and blood glucose levels in Figure 11, as can be shown. In general, they are less active. People above the age of fifty are more vulnerable to CD given in Figure 11, Patients having Cardiovascular had significantly increased cholesterol and blood glucose levels. In general, they are less active. On average, men drink alcohol greater frequently than women. It has been found in Figure 12. Patients above the age of 55 are always more subject to CD.
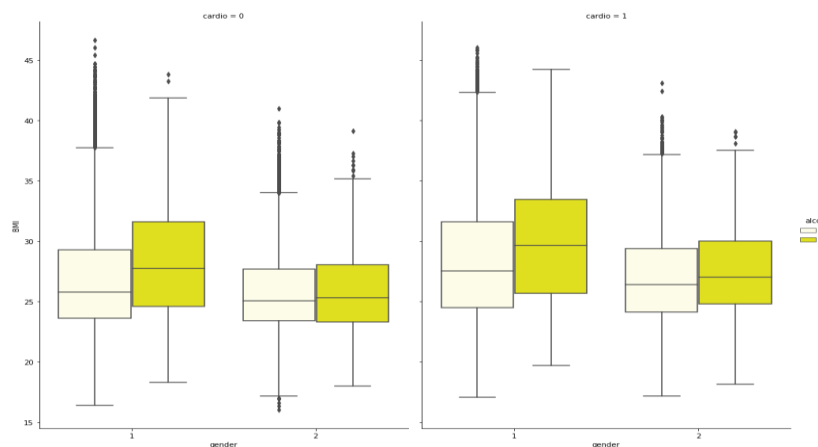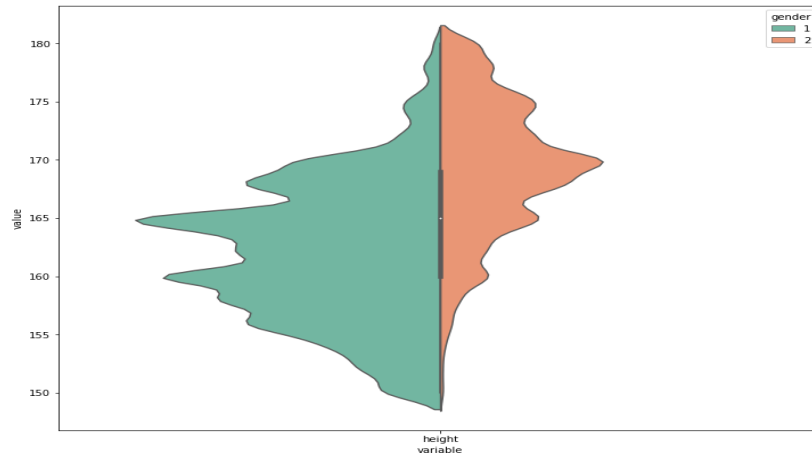

Figure 8. Alcoholic females vs. men
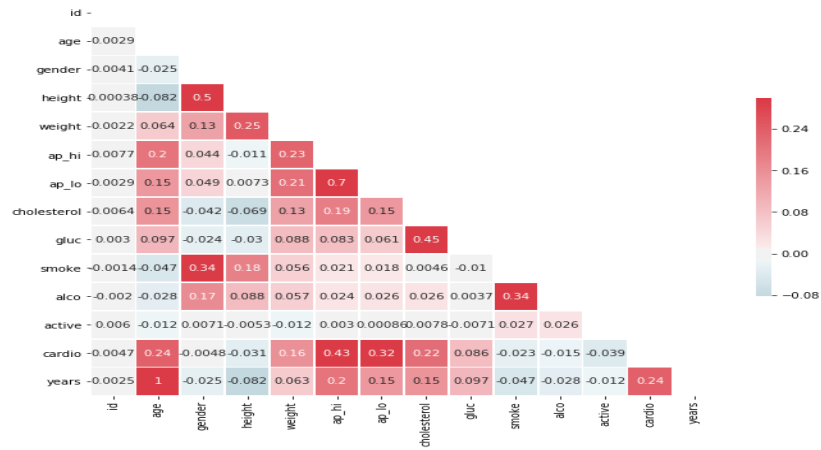
Figure 9. BMI of Females vs. Men

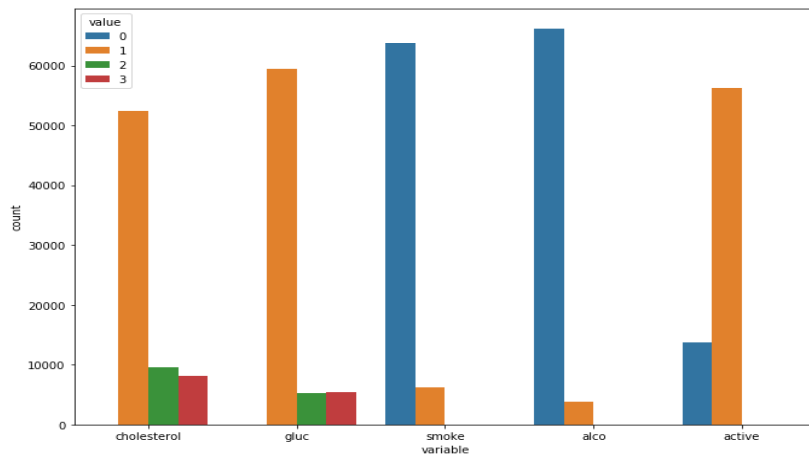**Figure 10.** Weighted Correlation Matrix of a data set

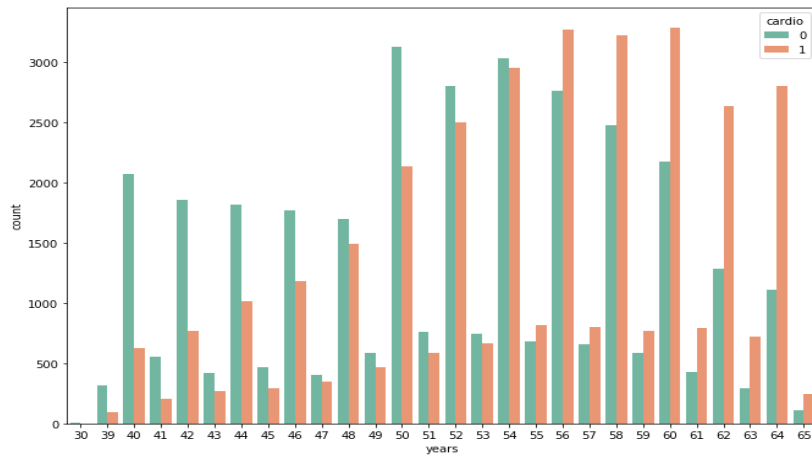**Figure 11.** Correlation of Feature

Figure 12. Heart Diseases effect on age

6. Conclusion and Future Work

The most common cause of death from heart disease is indeed a lag in diagnosis. To help minimize this, our findings of this article recommended a composite heart disease decision support system. The study's key contribution is to offer an improved decision support system for identifying heart disease that is more accurate than previous methods. The authors have discovered the finest algorithms through simulation and employed them when developing the hybrid recommendation system, whether that is the stage or data point, feature extraction, or classifier choice. The authors used the Cleveland data set in a simulated environment created in Google Collab and tested and compare the proposed approach.
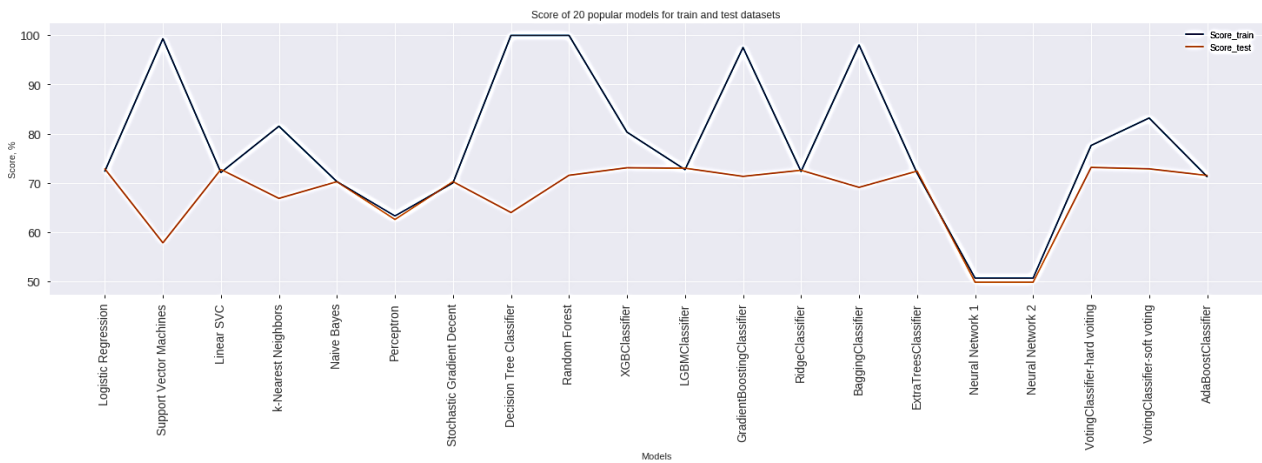

Figure 13. Scores of Classifiers

Used EDA (Exploratory data analysis) a technique for gaining insights into data who might just compile a summary of the numerical data presented in Table 2. In comparison to previous hybrid decision support systems, it has performed better, Hard Voting Classifier and XGB Classifiers in Figure 13, had the highest Score while accuracy is 77.62 % for Training Accuracy and 73.16 % for Testing Accuracy, followed by the XGB classifier given in Figure 13, with 80.32 % for Train Efficiency and 73.09 % for Test, The suggested technology is not a substitute for a doctor, but it may be employed in isolated and rural locations where there are no cardiac doctors and other modern healthcare centers. Furthermore, it might help the doctor make rapid judgments. There are several drawbacks to the suggested system as well. It can only determine whether or not such a patient has cardiovascular problems. This approach is unable to determine the degree of cardiac disease at this earlier

movement. To enhance the system's performance, the Publishers want to test innovative ways of extracting features in the future, including digitization of the ECG by employing exploratory data analysis extraction. The authors also propose to use recurrent neural networks and model a system for diagnosing cardiac problems. In addition, the authors want to apply the presented technique to the diagnosis of additional chronic illnesses such as renal disease, hypertension, and malignancy. The publishers would like to put their algorithm under the supervision of a doctor so order to assess its functionality using real-time data from heart disease patients. Furthermore, the proposed system may be expanded by employing an android smartphone to collect data in real-time from rural parts and interact with a doctor on the other side connected with an app to save people's lives in crucial situations in remote areas. Using Infrared sensors for capturing clinical data such as pulse rate, oxygen level, and body temperature.

References

1. McGurnaghan, S., Blackbourn, L. A. K., Mocevic, E., Haagen Panton, U., McCrimmon, R. J., Sattar, N., ... & Colhoun, H. M. (2019). Cardiovascular disease prevalence and risk factor prevalence in Type 2 diabetes: a contemporary analysis. Diabetic Medicine, 36(6), 718-725.
2. Mackay, J., Mensah, G. A., & Greenlund, K. (2004). The atlas of heart disease and stroke. World Health Organization.
3. Expert Panel on Detection, E. (2001). Executive summary of the third report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (adult treatment panel III). Jama, 285(19), 2486-2497
4. Lorenzo, C., Williams, K., Hunt, K. J., & Haffner, S. M. (2007). The National Cholesterol Education Program–Adult Treatment Panel III, International Diabetes Federation, and World Health Organization definitions of the metabolic syndrome as predictors of incident cardiovascular disease and diabetes. Diabetes care, 30(1), 8-13
5. Goodman, G., & Bourbakis, N. (2020, October). Predicting the change in state of the human heart based on synthetic heart chamber volume data. In 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE) (pp. 655-661). IEEE.
6. Mishra, R., & Monica (2019). Determinants of cardiovascular disease and sequential decision-making for treatment among women: A Heckman's approach. SSM - population health, 7, 100365.
7. Kim, J. O. R., Jeong, Y. S., Kim, J. H., Lee, J. W., Park, D., & Kim, H. S. (2021). Machine Learning- Based Cardiovascular Disease Prediction Model: A Cohort Study on the Korean National Health Insurance Service Health Screening Database. Diagnostics, 11(6), 943
8. Schantz, P. M., & Tsang, V. C. (2003). The US Centers for Disease Control and Prevention (CDC) and research and control of cysticercosis. Acta tropica, 87(1), 161-16
9. Jindal, H., Agrawal, S., Khera, R., Jain, R., & Nagrath, P. (2021). Heart disease prediction using machine learning algorithms. In IOP Conference Series: Materials Science and Engineering (Vol. 1022, No. 1, p. 012072). IOP Publishing.
10. Baak, M., Koopman, R., Snoek, H., & Klous, S. (2020). A new correlation coefficient between categorical, ordinal and interval variables with Pearson characteristics. Computational Statistics & Data Analysis, 152, 107043.
11. Almustafa, K. M. (2020). Prediction of heart disease and classifiers' sensitivity analysis. BMC bioinformatics, 21(1), 1-18.
12. Haq, A. U., Li, J. P., Memon, M. H., Nazir, S., & Sun, R. (2018). A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. Mobile Information Systems, 2018.
13. Nikam, A., Bhandari, S., Mhaske, A., & Mantri, S. (2020, December). Cardiovascular Disease Prediction Using Machine Learning Models. In 2020 IEEE Pune Section International Conference (PuneCon) (pp. 22-27). IEEE.
14. Amin, M. S., Chiam, Y. K., & Varathan, K. D. (2019). Identification of significant features and data mining techniques in predicting heart disease. Telematics and Informatics, 36, 82-93.
15. Atallah, R., & Al-Mousa, A. (2019, October). Heart disease detection using machine learning majority voting ensemble method. In 2019 2nd international conference on new trends in computing sciences (ictcs) (pp. 1-6). IEEE.
16. Kumari, S., Kumar, D., & Mittal, M. (2021). An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier. International Journal of Cognitive Computing in Engineering, 2, 40-46.
17. Khanna, D., Sahu, R., Baths, V., & Deshpande, B. (2015). Comparative study of classification techniques (SVM, logistic regression and neural networks) to predict the prevalence of heart disease. International Journal of Machine Learning and Computing, 5(5), 414.
18. Hasan, K. Z., Datta, S., Hasan, M. Z., & Zahan, N. (2019, February). Automated prediction of heart disease patients using sparse discriminant analysis. In 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE) (pp. 1-6). IEEE.
19. Maheswari, S., & Pitchai, R. (2019). Heart disease prediction system using decision tree and naive bayes algorithm. Current Medical Imaging, 15(8), 712-717.
20. Jabbar, M. A., Deekshatulu, B. L., & Chndra, P. (2014, November). Alternating decision trees for early diagnosis of heart disease. In International Conference on Circuits, Communication, Control and Computing (pp. 322-328). IEEE.
21. Singh, Y. K., Sinha, N., & Singh, S. K. (2016, November). Heart disease prediction system using random forest. In International Conference on Advances in Computing and Data Sciences (pp. 613-623). Springer, Singapore.
22. Shouman, M., Turner, T., & Stocker, R. (2012). Integrating decision tree and k-means clustering with different initial centroid selection methods in the diagnosis of heart disease patients. In Proceedings of the International Conference on Data Science (ICDATA) (p. 1). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).
23. Amin, S. U., Agarwal, K., & Beg, R. (2013, April). Genetic neural network based data mining in prediction of heart disease using risk factors. In 2013 IEEE Conference on Information & Communication Technologies (pp. 1227-1231). IEEE.
24. Sudha, A., Gayathri, P., & Jaisankar, N. (2012). Effective analysis and predictive model of stroke disease using classification methods. International Journal of Computer Applications, 43(14), 26-31.

25.   World Health Organization. (1999). Definition, diagnosis and classification of diabetes mellitus and its complications: report of a WHO consultation. Part 1, Diagnosis and classification of diabetes mellitus (No. WHO/NCD/NCS/99.2). World Health Organization