Improvised Cardiovascular Disease Prediction using Classifiers and Ensembling Technique

Aashik Bharath Egnya Varahan
R00182866
Cork Institute of Technology
Cork, Ireland
aashik.egnya-varahan@mycit.ie

ABSTRACT

With big data growth in biomedical and healthcare communities, accurate analysis of medical data benefits early disease prediction, patient care, and community services. In this research, the utilization of Machine Learning algorithm for accurate heart disease prediction with the prevailing dataset is explored. The advancements in Machine Learning and easy availability of digitalized healthcare datasets has paved approach for reliable analysis of benefitting the early disease prediction. The initial step for any data-driven decision analysis is Data Mining, which provides a way to get the right information from the dataset and to understand the hidden patterns. This paper investigates a method termed ensemble classification, which is used for improving the accuracy of weak algorithms by combining multiple classifiers and analyses various Feature Selection methods to extract a set of important features. Experiments with this technique were performed using a heart disease dataset. A comparative analytical approach was done to determine how the ensemble technique can be applied for improving prediction accuracy in heart disease. The focus of this paper is not only on increasing the accuracy of weak classification algorithms, but also on the implementation of the algorithm with a medical dataset, to show its utility to predict the cardiovascular disease at an early stage. The results of the study indicate that ensemble techniques, such as majority voting, are effective in improving the prediction accuracy of weak classifiers and exhibit satisfactory performance in identifying risk of heart disease. A maximum increase of 8% accuracy for weak classifiers was achieved with the help of ensemble classification. The performance of the process was further enhanced with a feature selection implementation, and the results showed significant improvement in prediction accuracy.

Keywords— Heart Disease, Machine Learning, Ensemble Classifiers, Predictive Analytics

I. INTRODUCTION

One of the prominent diseases that affect many people during middle or old age is heart disease, and in many cases eventually leads to fatal complications. According to statistics from WHO [1], it has been estimated that 24% deaths due to non-communicable diseases in India are caused by heart ailments. One-third of all global deaths are due to heart disease. Half of the deaths in the United States and in other developed countries are due to heart ailments. Around seventeen million people die due to cardiovascular disease (CVD) every year worldwide, and the disease is highly prevalent in Asia [2], [3].

The Cleveland Heart Disease Database (CHDD) [4] is considered the de facto database for heart disease research. The dataset consists of 14 attributes and 303 instances. There are 8 categorical attributes and 6 numeric attributes. The description of the dataset is shown in Table 1.

Table 1: Feature Information of Cleveland Heart Dataset

	Description	Variable
age	age in years	continuous
sex	1 = male, 0 = female	categorial
cp	chest pain type: 1: typical angina, 2: atypical angina, 3: non-anginal pain, 4: asymptomatic	categorial
trestbps	resting blood pressure in mm Hg	continuous
chol	serum cholestoral in mg/dl	continuous
fbs	fasting blood sugar > 120 mg/dl: 1 = true, 0 = false	categorial
restecg	0: normal, 1: having ST-T wave abnormality, 2: left ventricular hypertrophy	categorial
thalach	maximum heart rate achieved	continuous
exang	exercise induced angina (1 = yes; 0 = no)	categorial
oldpeak	ST depression induced by exercise relative to rest	continuous
slope	the slope of the peak exercise ST segment: 1: upsloping, 2: flat, 3: downsloping	categorial
ca	number of major vessels: (0-3) colored by flourosopy	continuous
thal	3: normal, 6: fixed defect, 7: reversable defect	categorial
target	diagnosis of heart disease: (0 = false, 1 = true	categorial

Patients from the age group between 29 and 79 have been selected in this dataset. The demographic characteristics such as age, sex have been recorded along with other blood sample report metrics. The "target" feature is the predictor variable which classifies whether a person has heart disease or not.

Several existing research on the prediction of heart diseases focuses mainly on classification algorithms. Some researchers have prescribed using ensemble methods to improve classification accuracy in heart disease prediction [5], [6].

Y. Alp Aslandogan, et. al., worked on three different classifiers called K-nearest Neighbour (KNN), Decision Tree, Naïve Bayesian and used Dempster's rule for these three viewpoints to appear as one concluding decision. This classification based on the combined idea showed increased accuracy [7].

Franck Le Duff worked on creating Decision tree quickly with clinical data of the physician or service. He suggested few data mining techniques which can help cardiologists in the predication survival of patients. The main drawback of the system was that the user needs to have knowledge of the techniques and we should collect sufficient data for creating an suitable model [8].

M. Guddadhe et. al., functioned for forecasting of heart disease, Blood Stress and Sugar by the aid of neural systems. Hearings were accepted out on example best ever of patients. The neural system is verified with 13 types, as blood pressure, period, angiography etc. [9]

In this research, our study will be focused on improving the weakness of weak classification algorithms by combining them with other classification algorithms. A research on using ensemble techniques such as bagging, boosting and majority voting is done and the results are evaluated. The results are further enhanced by applying feature selection. The results are a measure to indicate how these classifiers can effectively be used in the medical field.

II. RESEARCH

A. Feature Selection

Feature Selection is one of the core concepts in machine learning which hugely impacts the performance of your model. The data features that you use to train your machine learning models have a huge influence on the performance you can achieve.

Irrelevant or partially relevant features can negatively impact model performance. The following are the different methods used to extract the important features from a large set.

a. Univariate Feature Selection

Univariate Feature Selection works by selecting the best features based on Univariate statistical tests. Each feature is compared to the target variable, to see whether there is any statistically significant relationship between them. It is also called as Analysis of Variance (ANOVA). The relationship between each feature and the target feature is analysed. After the analysis, each feature is valuated with a score and the test scores of all features will be compared later. The features with highest score will be selected for Modelling.

There are three types of tests available under Univariate feature selection method which can be used according to the requirements.

- Chi-Square test
- F-test [Classification & Regression]
- mutual_info_classif test

In this research, we have used f-test for classification for the feature's extraction.

b. Tree-based Feature Selection

Decision trees work by reducing the level of uncertainty or impurity in the data and can use one of a range of metrics to do this (gini, entropy, etc.). In this method, the feature is selected on the criteria that which provides the largest reduction in overall uncertainty. Also, for each feature we can quantify (measure) the reduction in uncertainty.

Tree based feature selection builds many trees and calculates the average reduction in uncertainty achieved by each feature across all the trees and uses this as a means of feature ranking. The features with the largest reduction have the highest impact (the most important features).

c. Recursive Feature Elimination

The Recursive Feature Elimination (or RFE) works by recursively removing attributes and building a model on those attributes that remain.

It uses the model accuracy to identify which attributes (and combination of attributes) contribute the most to predicting the target attribute.

B. Ensemble Techniques

Ensemble is a strategy that can be used to improve the accuracy of a classifier. It is an effective meta classification technique that combines weak learners with strong learners to improve the efficacy of the weak learner. In this paper, the ensemble technique is used to improve accuracy of various algorithms for heart disease prediction. The aim of combining multiple classifiers is to obtain better performance as compared with an individual classifier. The procedure for ensemble is shown in Figure 1.

a. Boosting

Boosting is an algorithm used for ensemble. In boosting, the original dataset is divided into various subsets. The classifier is trained with the subset to produce a series of models of moderate performance. New subsets are created based on the elements that are not correctly classified by the previous model. Then, the ensemble process boosts their performance by using combining technique to the weak models together using

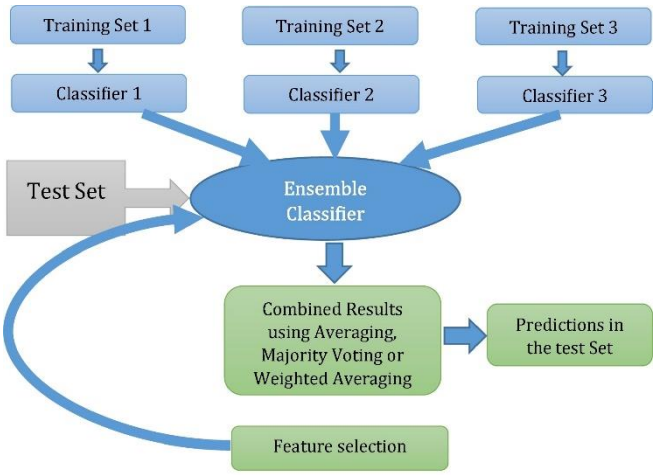

Figure 1: Ensemble Process

a cost functions. The algorithm for boosting is shown in Figure 2.

Let D={d_1,d_2,d_3, … d_n} be the given dataset
E = {}, the set of ensemble classifiers
C = {c_1, c_2, c_3, …c_n}, the set of classifiers
X = the training set, X ∈ D
Y = the test set, Y ∈ D
L = n(D)
Let init = 1
S(init)=A random subset of X; S(init)⊂X
M(0) = { }
fori =1 to L do
if i>1
s(i) = Set of incorrectly classified instances of M(i-1) + S(i)
M(i) = Model trained using C(i) on S(i)
E = E∪C(i)
end if
next i
for i = 1 to L
R(i) = Y classified by E(i)
next i
Result = max(R(i): i=1,2, …, n)

Figure 2: Boosting Algorithm

b. Bagging

Bagging is also known as *bootstrap aggregation*. Bagging randomly selects some patterns from the training set with replacement. The newly created training set will have the same number of patterns as the original training set with a few omissions and repetitions. The new training set is known as *Bootstrap replicate*. In bagging, bootstrap samples are fetched from the data and the classifier is trained with each sample. The voting from each classifier is combined, and the classification result is selected based on majority voting or averaging. Research shows that bagging can be used to increase the performance of a weak classifier optimally. The algorithm for bagging is shown in Figure 3.

d. Majority Vote

The majority voting classifier is a meta classifier that is

Let D={d_1,d_2,d_3, … d_n} be the given dataset
E = {E_1, E_2, E_3, …E_n}, the set of ensemble classifiers
C = {c_1, c_2, c_3, …c_n}, the set of classifiers
X = the training set, X ∈ D
Y = the test set, Y ∈ D
K = meta level classifier
L = n(D)
for i =1 to L do
M(i) = Model trained using E(i) on X
Next i
M=M∪K
Result = Y classified by M

Figure 3: Bagging Algorithm

used to combine any classifier through majority voting. The final class label would be the class label that had been predicted by most of the classifiers. The majority voting algorithm is shown in Figure 4.

Let $c_{i,j}$be the prediction of the i^{th} classifier on a class with j labels

$$\sum_{i=1}^{n} c_{i,j} = max_{j=1,\ldots,m} \sum_{t=1}^{n} c_{i,j}$$

The ensemble classifier's probability for the decision to be better is

$$P_{ens} = \sum_{k=\left(\frac{n}{2}\right)+1}^{n} \binom{n}{k} p^k (1-p)^{n-k}$$

Figure 4: Majority Voting Algorithm

C. Classification Algorithms

Classification is a supervised machine learning procedure that is used for predicting the outcome from existing data. This paper proposes an approach for the diagnosis of heart disease using classification algorithms, and to improve the classification accuracy using an ensemble of classifiers. The dataset has been divided into a training set and a test set, and individual classifiers are trained using the training dataset. The efficiency of the classifiers is tested with the test dataset. The working of the individual classifiers is explained below.

a. *Naive Bayes*

The Naive Bayes classifier or simply, the Bayesian classifier, is based on the Bayes theorem. It is a special case of the Bayesian network, and it is a probability-based classifier. In the Naïve Bayes network, all features are conditionally independent. The changes in one feature therefore does not affect another feature. The Naïve Bayes algorithm is suitable for classifying high dimensional datasets. The classifier algorithm uses conditional independence. Conditional independence assumes that an attribute value is independent of the values of the other attributes in a class.

b. *Random Forests*

Random Forest is a tree-based classification algorithm. As the name indicates, the algorithm creates

a forest with many trees. It is an ensemble algorithm which combines multiple algorithms. It creates a set of decision trees from a random sample of the training set. It repeats the process with multiple random samples and makes a final decision based on majority voting. The Random forest algorithm is effective in handling missing values, but it is prone to overfitting. Appropriate parameter tuning can be applied to avoid overfitting. The algorithm for Random Forest is shown in Figure 5.

Let D be a training set $D = \{(x_1, y_1), \ldots, (x_n, y_n)\}$
Let $h = h_1(x), h_2(x), \ldots, h_k(x)$, *an ensemble of weak classifiers*
If each h_k *is a decision tree, the parameters of the tree are defined as* $\Theta = (\theta_{k1}, \theta_{k2}, \ldots, \theta_{kp})$
Each decision tree k leads to a classifier $h_k(X) = h(X|\Theta_k)$
Final Classification f(x) = Majority of $h_k(X)$

Figure 5: Random Forest Algorithm

In this research, we will explore the above discussed feature selection methods, classification algorithms and ensemble techniques to predict the presence of heart disease in a patient with highest accuracy.

III. METHODOLOGY

In this study, the performance of different classification algorithms and ensemble techniques used to predict the outcome of a cardiovascular disease presence will be evaluated.

Before evaluating the model, the dataset needs to be prepared and processed to understand the relationship between each feature in the dataset with the response feature.

a. Dealing with Missing Values

The very first step in Data PreProcessing is to deal with the missing values present in the dataset. The machine learning algorithm cannot process the missing values and it should either be dealt or the observation which has the missing value must be removed.

Removal of observations with missing values will result in loss of other important feature information in the dataset. Also, the removing the observations may result in reduction of accuracy if the dataset is considerably small (i.e., has less samples).

In our dataset, there were no missing values found in the 303 observations with 14 attributes. The output of this code snippet to check missing values in the dataset in shown in Figure 6.

If there are missing values present in the dataset, it can be treated by certain methods such as Imputation. In this method, the missing values are filled with the mean, median (for numerical features) and mode (for ordinal features).

```
In [4]: # NAs present in each feature columns
        heart_dataset.isnull().sum()

Out[4]: age         0
        sex         0
        cp          0
        trestbps    0
        chol        0
        fbs         0
        restecg     0
        thalach     0
        exang       0
        oldpeak     0
        slope       0
        ca          0
        thal        0
        target      0
        dtype: int64
```

Figure 6: Output of Missing Values

b. Handling Categorical Data

To make the data compliant *for* the algorithms to process accurately, the categorical features in the dataset are converted into integers. Although the numbered categorical features have no specific ordering, the learning algorithm assumes that the observation with larger value has more importance.

In order to avoid this misinterpretation, a technique referred to as 'one-hot encoding'. This technique will create new dummy features for each unique value in the nominal feature column.

After implementing one-hot encoding the features such as sex, cp, fbs will be converted as sex_0, sex_1, cp_0, cp_1, cp_2, cp_3, fbs_0 and fbs_1 creating new dummies. The dummies will take values as shown in Figure 7.

sex_0	sex_1	cp_0	cp_1
0	1	0	0
0	1	0	0
1	0	0	1
0	1	0	1
1	0	1	0

Figure 7: Dummy Features

The disadvantage of one-hot encoding is the addition of very large number of features leads to long training time or underperformance of few models due to the rapid increase in the dimensionality.

One approach to mitigate this issue is feature selection or dimensionality reduction which allows to dramatically reduce the overall number of features.

c. Scaling Data

Feature scaling is another important step in Data PreProcessing for machine learning. The machine learning and optimization algorithms behave better when all the features are on the same scale.

The two most common data transformation techniques used are:

- Normalising Data
- Standardizing Data

Normalization is the rescaling of features into the range between 0 and 1. It improves the performance of algorithms which assigns a weight to features such as linear regression and algorithms that utilize geometric distance such as KNNs. The formula for normalization is shown in Figure 8.

$$X_{new} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Figure 8: Formula for Normalization

Standardization, on the other hand, facilitates the transformation of features to a standard Gaussian distribution with a mean of 0 and a standard deviation of 1.

This scaling happens independently on each individual feature by computing the relevant statistics on the samples in the training set.

Standardization is a common requirement for dealing with dataset features with different ranges and also for many machine learning estimators such as Logistic Regression, SVMs. It is important to note that standardization is less sensitive to outliers than normalization. The formula for standardization is shown in Figure 9.

$$z = \frac{x_i - \mu}{\sigma}$$

Figure 9: Formula for Standardization

where z is the standardized value, x is the input feature value, μ is the mean and σ is the standardisation.

In this research, both Normalization and Standardization is used for scaling the dataset. Scaling is applied to numerical variables such as age, trestbps, chol, thalach and oldpeak. The output of the normalized and standardized dataset is show in Figure 10.a and Figure 10.b.

```
In [26]: heart_dataset_standardized
Out[26]:
```

	age	trestbps	chol	thalach	oldpeak
0	0.952197	0.763956	-0.256334	0.015443	1.087338
1	-1.915313	-0.092738	0.072199	1.633471	2.122573
2	-1.474158	-0.092738	-0.816773	0.977514	0.310912
3	0.180175	-0.663867	-0.198357	1.239897	-0.206705
4	0.290464	-0.663867	2.082050	0.583939	-0.379244

Figure 10.a: Normalized Dataset

```
In [13]: heart_dataset_normalized
Out[13]:
```

	age	trestbps	chol	thalach	oldpeak
0	0.708333	0.481132	0.244292	0.603053	0.370968
1	0.166667	0.339623	0.283105	0.885496	0.564516
2	0.250000	0.339623	0.178082	0.770992	0.225806
3	0.562500	0.245283	0.251142	0.816794	0.129032
4	0.583333	0.245283	0.520548	0.702290	0.096774

Figure 10.b: Standardized Dataset

e. Handling Imbalance

Class Imbalance in a classification problem is a common problem where a disproportionate ratio of observations in each class. The machine learning algorithm works best when the number of samples in each class are about equal to give maximum accuracy and reduced error.

The most common approaches to deal with the imbalanced data are:

- Under-Sampling
- Over-Sampling

Under-Sampling technique attempts to rebalance the data by removing data from the majority class until the classes are balanced.

Over-Sampling technique attempts to rebalance the data by gradually increasing the size of the majority class by duplicating instances from the majority class.

The drawback of Under-Sampling is loss of information whereas Over-Sampling many result in overfitting.

In this research, the dataset is balanced by default and there is no need of implementing any techniques to deal with class imbalance. The proportionate ratio of the classes in the response variable is shown in Figure 11.

f. Data Split

It is an important part of evaluating the model to be built. The complete dataset is separated into a separate training set and test set with a defined ratio. In this research, we have split the dataset in a ratio of 80:20 i.e., the training dataset will have 80% of records from the main dataset and the test dataset will have the remaining 20% of the data.

g. Feature Selection

Feature Selection is the process where the features in the dataset are automatically selected based on the weight of their contribution to the predictor variable. Presence of irrelevant features in the dataset can decrease the accuracy of the model and makes the model to learn based on irrelevant features.

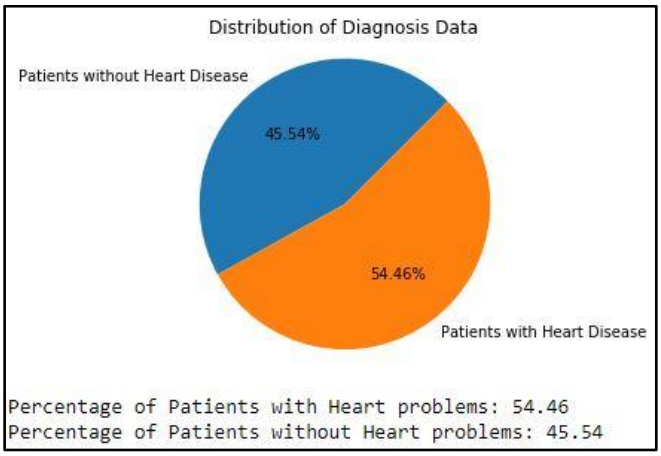

Figure 11: Proportionate of Response Class

The three main methods of Feature Selection are:

- Univariate Feature Selection
- Tree based Feature Selection
- Recursive Feature Elimination

In this research, we have implemented all the above methods described above for observing the impact caused by each of them.

As discussed earlier, Univariate Feature Selection works by selecting the best features based on the univariate statistical tests. Each feature is compared to the target variable to check if any statistically significant relationship exists between them. The feature is analyzed with the target independently without considering any other feature. It is also known as Analysis of Variance (ANOVA).

As the end of this analysis, the features have been reduced from 30 to only 5 features with not much decrease in the accuracy of the model.

A one-way ANOVA is used to select the top 5 features with the highest importance in model building in this method.

Similarly, Tree based Feature selection is implemented on the same dataset. This method returned a set of 5 features with high importance. The top 5 features returned by the method are **"thalach", "oldpeak", "cp", "exang"** and **"thal".**

Recursive Feature Elimination (RFE) is totally different from the above discussed feature selection methods. Univariate and Tree based Feature Selection works based on the dataset whereas RFE fits a model first and removes the weakest feature (or features) until the specified number of features is reached. Features are ranked by the model's *coef_* or *feature_importances_* attributes, and by recursively eliminating a small number of features per loop, RFE attempts to eliminate dependencies and collinearity that may exist in the model.

RFE further reduces the dataset to 4 features which has high importance correlating with the target variable.

h. Building Models

In this research, four different classification models are built.

Logistic Regression

The first model built on the dataset is Logistic Regression with "liblinear" as solver parameter. The model is trained without using any hyper parameter optimization technique and the accuracy was recorded. The accuracy score for the model without optimization was 78.69%.

The model is re-tuned with Regularization strength value which is an important parameter in Logistic Regression model. The best value of Regularization value which return the highest model accuracy is noted.

The model is again rebuilt with Recursive Feature Elimination (RFE) method to further optimize the number of features used for model training.

The model is built on top of 4 features selected by RFE method and this has now increased the accuracy of the model to 83.61%.

Gaussian Naïve Bayes

This algorithm is based on "Bayes Theorem" for classification. The parameter tuning is very limited respective to this algorithm.

The classifier is trained on two sets of data. One set of data with complete features without any data reduction. In this model, the accuracy was observed as 75.41 % and the model was also overfitted.

The other set of data was with features selected by Univariate Feature Selection method. This dataset has accuracy improve by 5%.

K-Nearest Neighbors

This algorithm stores all available cases and classify new cases based on a similarity measure i.e., distance functions. KNN has been used in statistical estimation and pattern recognition.

A case in KNN is classified by a majority vote of its neighbors , with the case being assigned to the class most common among its K nearest neighbors measured by a distance function.

For continuous variables, three distance functions such as Euclidean, Manhattan and Minkowski are used.While for categorical variables, Hamming Distance is used.

The algorithm is tuned on hyper parameters such as 'leaf_size', 'n_neighbors' and 'p' values. The best fit model is chosen based on the value of best hyperparameters.

This algorithm showed an accuracy about 83.61% for the complete dataset and a decreased accuracy of 81.97% with a dataset containing only 5 of the total 30 features.

Random Forest Classifier

It is one of the widely used algorithm for classification problems. Random Forests or Random Decision Forests is an ensemble learning method for classification, which works by constructing a multitude of Decision Trees at training time and outputting the class that is the mode of the classes (classification) of the individual trees.

It also corrects the decision trees habit of overfitting the training model.

In this research, Random Forest Classifier is implemented and the parameters such as 'n_estimators' and 'max_features' are hypertuned to get the best results of accuracy.

IV. EVALUATION

A comparative analysis of various classification algorithms on the Cleveland Heart Disease dataset has been performed. Some algorithms showed good accuracy whereas some algorithms performed average. But after implementing feature selection methods, all the algorithms modelled on training data showed an increase in the accuracy score. This work has used feature selection methods such as Univariate Feature Selection, Tree based Feature Selection and Recursive Feature Elimination. The best fit of all the algorithms is then ensembled using technique such as voting. The end results show that the weak classifiers can perform better when they are ensembled.

At the outset, the dataset is cleaned and pre-processed for outliers, missing values. Then, the dataset is split as training and testing datasets. The features in the training and test dataset is identified using Feature Selection methods. Classifiers such as Logistic Regression, Gaussian Naïve Bayes, K-Nearest Neighbors and Random Forest are used to train the model.

The model is trained using the above classifiers for two types of dataset. The dataset with the complete list of features and the dataset with only features selected by techniques such as Univariate Feature Selection and Recursive Feature Elimination.

Figure 12 compares the classification accuracy of individual classifiers on the complete dataset without any data reduction.

When the dataset is classified using individual classifiers with complete dataset, the accuracy rates of Logistic Regression, Gaussian Naïve Bayes, K-Nearest Neighbors and Random Forests are found in the range of 75.41% - 83.61%. The Random Forest Classifier exhibits the highest accuracy of 83.61% whereas Gaussian Naïve Bayes and KNN algorithm show comparatively poor accuracy below 80%.

Figure 12: Accuracy Plot – Complete Feature Set

Contrarily, when the dataset is classified using individual classifiers with only a set of few features decided by feature selection techniques, the accuracy rates of Logistic Regression, Gaussian Naïve Bayes, K-Nearest Neighbors and Random Forests have increased significantly. The accuracy of these classifiers lies in the range of 80.33% - 85.25% which is show in Figure 13.

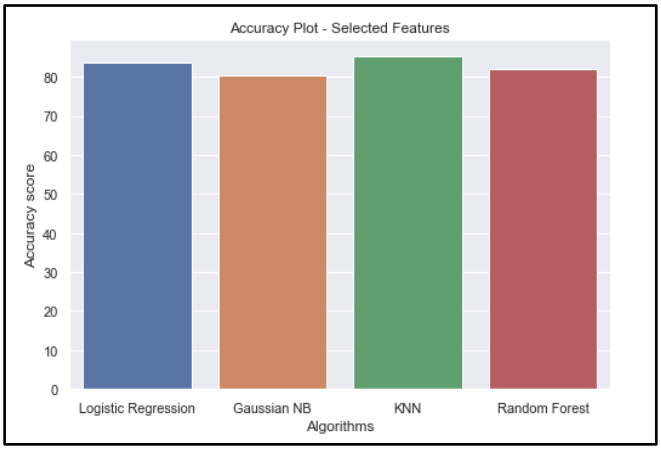

Figure 13: Accuracy Plot – Selected Features

Furthermore, the models are ensembled to improve the accuracy by learning from weak classifiers. Majority voting is one such ensemble strategy that combines multiple classifiers in order to improve their accuracy. In the proposed approach, for the Cleveland dataset, Gaussian NB and Random Forests classifiers turned out to be weak classifiers and they showed less accuracy. Logistic Regression and K-Nearest Neighbors performed well and had better classification accuracy.

By ensembling all the implemented classifiers using Voting classifier, we train a new model with K-fold Cross Validation on the feature selected dataset.

The accuracy of the ensemble classifier has showed an increase of the mean accuracy of about 83.9%, which is better accuracy than all the classifiers combined.

V. CONCLUSION

This paper analyses the accuracy of prediction of heart disease using appropriate feature selection method and an ensemble of classifiers. The Cleveland heart dataset from the UCI machine learning repository was utilized for training and testing purposes. The feature selection methods such as Univariate Feature Selection, Tree based Feature Selection, Recursive Feature Elimination and ensemble technique like Majority Voting were employed for experiments. When the weak classifiers are ensembled with majority voting with the model trained on selected features, the accuracy was improved by a maximum of 8.45%.

This research can further be extended to reduce the 'Misclassification Costs'. In a domain like Medical field, the prediction of False Negatives is worse when compared to the prediction of False Positives. Especially, in a scenario like prediction of diseases, wrong classification of a patient disease can be fatal. So, this classification error can be reduced as further continuation of this research paper.

REFERENCES

[1] "WHO | The Atlas of Heart Disease and Stroke," *WHO*. https://www.who.int/cardiovascular_diseases/resources/atlas/en/ (accessed May 17, 2020).

[2] M. Vasighi, A. Zahraei, S. Bagheri, and J. Vafaeimanesh, "Diagnosis of coronary heart disease based on 1 H NMR spectra of human blood plasma using genetic algorithm-based feature selection," *Journal of Chemometrics*, vol. 27, p. 318, Oct. 2013, doi: 10.1002/cem.2517.

[3] K. Uyar and A. Ilhan, "Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks," *Procedia Computer Science*, vol. 120, pp. 588–593, Jan. 2017, doi: 10.1016/j.procs.2017.11.283.

[4] "UCI Machine Learning Repository: Heart Disease Data Set." https://archive.ics.uci.edu/ml/datasets/Heart+Disease (accessed May 17, 2020).

[5] B. Fida, M. Nazir, N. Naveed, and S. Akram, "Heart disease classification ensemble optimization using Genetic algorithm," in *2011 IEEE 14th International Multitopic Conference*, Dec. 2011, pp. 19–24, doi: 10.1109/INMIC.2011.6151471.

[6] P. Chotwani, A. Tiwari, V. Deep, and P. Sharma, *Heart Disease Prediction System using CART-C*. 2018, p. 5.

[7] Y. A. Aslandogan and G. A. Mahajani, "Evidence combination in medical data mining," in *International Conference on Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004.*, Apr. 2004, vol. 2, pp. 465-469 Vol.2, doi: 10.1109/ITCC.2004.1286697.

[8] F. Le Duff, C. Muntean, M. Cuggia, and P. Mabo, "Predicting survival causes after out of hospital cardiac arrest using data mining method," *Studies in health technology and informatics*, vol. 107, pp. 1256–9, Feb. 2004.

[9] M. Gudadhe, K. Wankhade, and S. Dongre, "Decision support system for heart disease based on support vector machine and Artificial Neural Network," in *2010 International Conference on Computer and Communication Technology (ICCCT)*, Sep. 2010, pp. 741–745, doi: 10.1109/ICCCT.2010.5640377.