International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 12 (2020) pp. 222-230
© MIR Labs, www.mirlabs.net/ijcisim/index.html

Received: 29 Jan, 2020; Accepted: 5 April, 2020; Published: 20 June, 2020

Linear and non-linear HRV features for the prediction of heart disease among smokers: a predictive evaluation of machine learning model

Rathod S R[1] and Patil C Y[2]

[1] Department of Instrumentation and Control,
College of Engineering Pune, India
rshashiikant@gmail.com

[2] Department of Instrumentation and Control,
College of Engineering Pune, India
cypatil@gmail.com

***Abstract*: In smokers it is found that they have increased sympathetic and reduced parasympathetic activity in the heart rate variability (HRV) analysis by researchers. Smoking lowers HRV, it harms heart health function as it influences the Autonomous Nervous System (ANS). Therefore, HRV features are useful for the prediction of heart disease. Due to poor health awareness and inadequate lifestyle, heart patients are proliferating. Hence, having a model that can easily recognize the predominance of heart disease potentially is essential. In this study, an HRV parameter is used to assess the potential of four machine learning techniques to predict heart disease among smokers. These techniques assessed on classification indices of accuracy, precision, sensitivity, specificity, misclassification rate, F1 score, area under the curve (AUC), kappa value, and mean square error (MSE). These techniques also evaluated on the receiver operating characteristic curve (ROC). The final model has the highest classification accuracy of 0.94 which was reported using a k-nearest neighbors (k-NN) method with 0.95, 0.92, 0.96, 0.94, 0.061, 0.878, 0.95 and 0.246 precision, sensitivity, specificity, F1 score, kappa value, AUC, and MSE respectively.**

***Keywords*:** Heart Rate Variability, Autonomous nervous system, Machine learning, Heart disease, k-nearest neighbors**.**

I. Introduction

Smoking is a well - known cardiovascular risk factor such as sudden cardiac arrest (SCA), myocardial infarction, and ventricular fibrillation. Sympathetic, parasympathetic, and hemodynamic responses to smoking concerning cardiovascular events [1]. The effect of smoking on the autonomous nervous system ANS resulted into changes in HRV. HRV is a non - invasive tool for evaluating cardiovascular function. The HRV method analyzes linear parameter in time domain, frequency domain, and non-linear parameters [2]. Different clinical results and cardiac mortality were also associated with HRV analysis. Therefore, in this study, the dataset of smokers suffering from heart disease was taken and a predictive model for heart disease prediction based on their HRV parameter was developed using a machine learning technique. Indeed, the diagnosis of heart disease is a complicated and critical task that is effectively, accurately, and skillfully achieved. This task causes excessive time, costs, and it often depends on physician understanding. HRV parameter give indication about the complication associated with heart earlier, so that effective treatment must start prior it becomes serious [1, 2].

Machine learning from the responded medical care data sources plays a vital role in anticipating diseases. It analyzes various features and records of the patient's details. It predicts if the patient has a particular type of disease based on an appropriate learning approach. It can also anticipate the extent of the disease by assessing the results of the different characteristics. Machine learning is about developing methodologies and strategies that enable computer systems to understand and collect information based on previous experiences.

The techniques of machine learning are widely classified into two types; supervised and unsupervised. It has various uses and effectiveness in anticipating diseases, as well as analyzing dataset for the disease. Machine learning computation is being used to examine multidimensional biomedical data instantly. This study uses a supervised machine learning model to predict presence and absence of heart disease using the HRV parameters.

In this work, four different models result based on k-Nearest Neighbors (k-NN), Support Vector Machine (SVM), Naive Bayes (NB), and Artificial Neural Network are used [3]. Various packages are used to implement the machine learning algorithms namely sklearn, pandas, numpy, matplotlib, and metrics. The data processing and evaluation were carried out on Python tool. Machine learning technique was evaluated using accuracy, precision, sensitivity, specificity, misclassification rate, F1 score, Kappa coefficient, mean square error and ROC curve. In that work, the highest accuracy of 93.61% by random forest algorithm is obtained [4]. The following Table shows a comparison of the proposed model with the existing model.

Authors	Methods	Accuracy obtained
Ref [5]	ANN	0.8800
Ref [6]	DT	0.9100
Ref [7]	LOR	0.9161
Ref [8]	Simple CART	0.9220
Ref [9]	C5.0	0.9302

Accuracy considered in percentage (%).

Table 1. Summary of heart disease prediction using UCI dataset

Authors	Methods	Accuracy obtained
Ref [10]	ANN	0.8396
Ref [11]	SVM-Poly	0.9233
Ref [12]	LMT	0.9250
Ref [04]	RF	0.9361
Proposed Model	**k-NN**	**0.9400**

Accuracy considered in percentage (%).

Table 2. Summary of heart disease prediction with HRV feature as input data

Most of the researchers proposed machine learning algorithm using Cleveland UCI repository heart disease dataset, and till date the highest accuracy achieved is 0.9302. Very few researchers have used HRV parameter to predict heart diseases listed in Table 2. HRV acts as a future predictor, it represents the early changes in their parameters when the pathological condition starts developing inside the body. Therefore, HRV can give early indication of a heart disease. The real motivation is to create a model which has the highest accuracy of heart disease prediction using HRV parameter.

The remaining part of the paper has the following organization: Section 2 deals with methodology. Section 3 deals with the classification algorithm used in the paper. Section 4 discussed ideas on performance measure, and Section 5 discusses the results obtained. The paper concludes with Section 6.

II. Methods

The dataset of smokers suffering from heart disease was obtained from the data science research group MITU skillogies Pune India. The dataset comprises of 1562 samples with 811 samples of smokers vulnerable to heart disease, and 751 are non-smokers sample.

A. Feature Extraction Methods

HRV features are derived using different linear and non-linear methods. The linear approach involves the technique of the time domain and frequency domain. The different methods and features are represented in fig. (1). The characteristics of the time domain, such as mean HR (HR-Heart Rate), mean RR (R to R Interval), SDNN (Standard Deviation of Normal to Normal Interval), and RMSSD (Root Mean Square of Standard Deviation).

Frequency-domain features such as Total Power (TP), low frequency (LF), high frequency (HF) in power (ms2), and normalized unit (nu), LF/HF ratio were measured. Using the Fast Fourier Transform (FFT) and Autoregressive (AR) technique, the frequency domain features were extracted.

The features obtained in the non-linear approach are the Poincare plot (SD1, SD2, and SD1/SD2), Detrended Fractal Analysis (DFA), Approximate Entropy (AppEN), and Sample Entropy (SampEN).

The machine-learning algorithm uses 4- time domain, 6- frequency domain, and 7- nonlinear parameter extracted as input (total=17 feature) to predict heart disease during HRV assessment. The dataset has two class existence and non-existence of heart disease.

B. Feature Selection Method

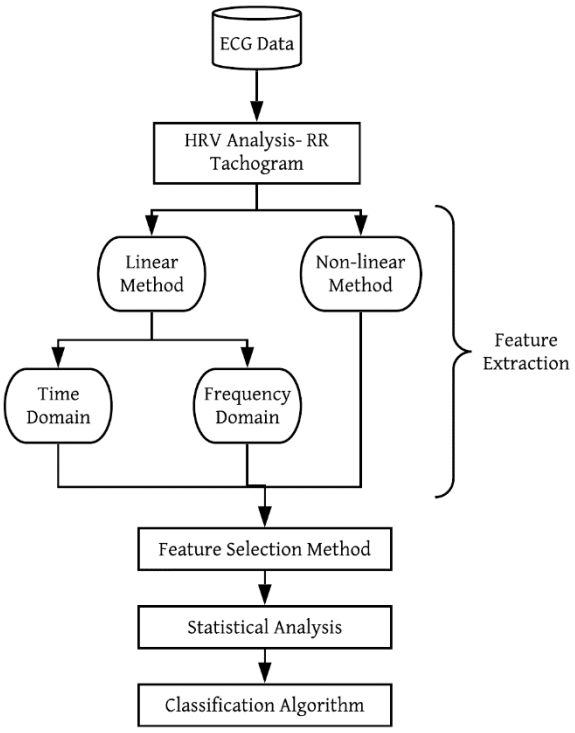

Figure1. Generalized diagram of automated computing of heart disease

Feature selection is applied for dimensionality reduction. While applying this it removes irrelevant, weakly relevant and redundant attributes. In this study, attribute selection is done using principle component analysis (PCA) method. PCA are used for reducing the number of features of the data. The main purpose of the feature space dimensionality reduction is to reduce the burden of computation to achieve higher accuracy. It helps in reducing features and improving predictions. Table 3 shows the number of features used before and after PCA application. Though the features are reduced but algorithm shows the same performance over different performance metrics.

Classifier	Number of features without PCA	Number of features with PCA
k-NN	17	9
SVM	17	8
NB	17	8
ANN	17	9

Table 3. Number of features with and without PCA

C. Statistical Analysis

Method	Features	P-Value
Time Domain Feature of HRV	Mean HR	<0.001
	Mean RR	<0.001
	SDNN	<0.001
	RMSSD	<0.01
Frequency Domain Feature of HRV	TP (ms^2)	<0.001
	LF (ms^2)	<0.001
	HF (ms^2)	<0.001
	LF (nu)	<0.001
	HF (nu)	<0.001
	LF/HF	<0.001
Non-Linear Features	SD1	<0.001
	SD2	0.126
	SD1/SD2	<0.01
	DFA-α1	0.152
	DFA-α2	0.101
	AppEN	<0.01
	SampEN	<0.01

Table 4. Feature statistics between normal and smoker group

For statistical analysis, we have used unpaired students 't' test. Students 't' test is a statistical tool to determine whether the mean of two data groups are different or not [28]. The test gives p-values for the feature extracted for the two groups of data. In the case of classification, a low p-value is preferred; hence a small p-value feature is selected. The $p< 0.05$ is chosen as the level of significance.

D. Predictive model and Pre-processing Techniques

The dataset split into two parts for training and testing in our proposed predictive model. In Machine Learning, the training dataset is the actual and most massive data used to train the model for different actions. These are the actual datasets that the continuous development system models learn to train the machine to work automatically with different algorithms. The testing dataset is a data type used to provide a developed and fitting unbiased decisive assessment on the training dataset.

The next step is data pre-processing. Discretization, resampling, feature reduction are pre-processing techniques. Discretization is a method to transform the dataset from numeric to nominal value. Resampling forms a new dataset by creating a subset of the previous dataset, and also used to find missing values in the dataset. Pre-processing suggests removing the outlier, missing value, and normalizing dataset. Feature scaling is a pre-processing step for data to normalize within the given range. It mainly helps the data to normalize within the given range [3, 4]. This study uses min-max normalization feature scaling algorithm is used.

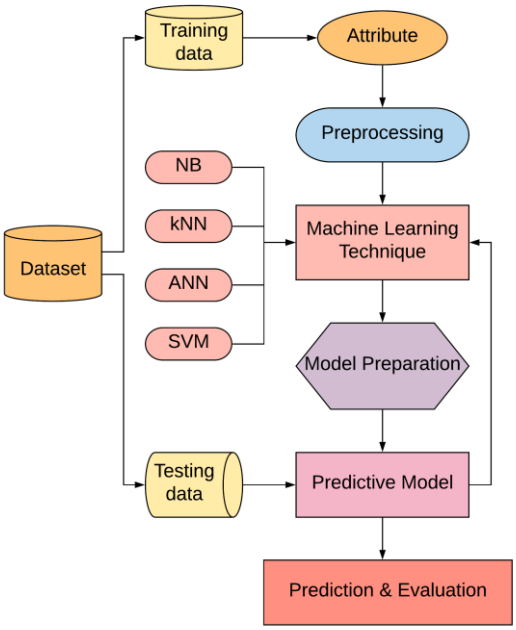

Figure 2. Framework for predictive model assessment

The dataset split into two parts for training and testing in our proposed predictive model. In Machine Learning, the training dataset is the actual and most massive data used to train the model for different actions. These are the actual datasets that the continuous development system models learn to train the machine to work automatically with different algorithms. The testing dataset is a data type used to provide a developed and fitting unbiased decisive assessment on the training dataset. The next step is data pre-processing. Discretization, resampling, feature reduction are pre-processing techniques.

Discretization is a method to transform the dataset from numeric to nominal value. Resampling forms a new dataset by creating a subset of the previous dataset, and also used to find missing values in the dataset. Pre-processing suggests removing the outlier, missing value, and normalizing dataset. Feature scaling is a pre-processing step for data to normalize within the given range. It mainly helps the data to normalize within the given range [3, 4]. This study uses min-max normalization feature scaling algorithm is used.

III. Classification Algorithm

A. K-Nearest Neighbor

K Nearest neighbor is a basic algorithm but the outcomes are quite good. It is a lazy, non - parametric learning algorithm based on instance. This method is used in concerns of regression and classification. k-NN is used in the classification to see which class belongs to the unique and unidentifiable object [13]. For this, k is the number of neighbors to consider, and the distance between the data points closest to the object is calculated by the distance formula of Euclidean [14, 15]. The Euclidean distance is determined by the equation between two points x and y –

$$\text{Euclidean} = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \quad \ldots (1)$$

By inspecting the dataset, the value of k is determined; generally, the odd 'k' value is considered to yield excellent outcomes. The higher k value is deemed to be more accurate [15]. K=9 is considered in this case.

B. Support Vector Machine

Support Vector Machine is considered a classification method, but it can be used in both classification and regression issues. Multiple continuous and categorical variables can be handled [16]. SVM develops a multi-dimensional space hyperplane to segregate various classes. SVM computationally produces an appropriate hyperplane that is used to reduce an error. A hyperplane is a plane of decision segregating a set of objects with distinct memberships in the class. Support vectors are the nearest data points to the hyperplane. By determining margins, these features will define the segregating line [16, 17]. These features are more appropriate to the classifier construction. A margin is a gap in the nearest class points between two lines. It is computed as the distance perpendicular to support vectors or the nearest points from the line. If the class margin is more inclusive, it is considered a stable margin, and a smaller margin is a weak margin.

SVM algorithm is implemented using a kernel. The kernel transforms the input data space into the form required. Kernel transforms the non - separable problem to separable problems by incorporating more dimensions to it [18]. Kernel technique helps to create a more accurate classifier. Linear, polynomial, and radial - based kernel functions are kernel types. Linear kernel is used this case.

C. Naïve Bayes

Bayes Theorem: P (X|Y) = probability of X given that of Y is true

$$P(X \mid Y) = \frac{P(Y \mid X) * P(X)}{P(Y)} \quad \ldots (2)$$

X= Event and Y=Data sample

The Naive Bayes algorithm is a primary predictive classifier which calculates a probability collection by examining the frequency and value confluence in the given set of the dataset [19, 20]. Because of the Naive Bayes algorithm simple structure, it is quite easy to build and it has several benefits too. Naive Bayes construction is incremental, and it does not require a large set of data, a smaller set of data can also be used. The algorithm based on the theorem of Bayes with the assumption that the impact on a particular class of parameter value is independent of other parameter values [21].

D. Artificial Neural Network

Artificial neural network imitates human brain feature. It is considered as a set of artificial neurons called nodes. All these nodes can convey information to each other. A multilayer perceptron is referred to as artificial neural networks. A single layer perceptron can solve fundamental problems where data separated into ' n ' dimensions, where ' n ' is the number of features in the dataset. However, the accuracy of a single layer perceptron drops significantly in particular instance of non - linearly separable data. Multilayer perceptron can work effectively with data that is not linearly separable [22].

Artificial neural networks, in the form of a network, are a combination of many neurons. It has a layer of input, one or more hidden layers, and a layer of output. Each layer essentially turns the data into information, and the optimal output is eventually provided. The activation function in the functioning of neurons plays a crucial role [23]. The activation function summarizes all the input weighted as-

$$N = \sum_{m=1}^{n} w_i m_i + w_b b \quad \ldots (3)$$

where b represents bias

IV. Performance Measure

For the classification of heart disease samples, four supervised machine learning techniques were applied. The assessment of classification performance was the tenfold cross-validation. Six different classification quality assessment measures such as accuracy, precision, sensitivity, specificity, false positive rate, misclassification rate, and F1 score were used [24, 25]. Samples were treated as a positive class with the non - existence of heart disease, and as negative with the existence of heart disease. Confusion matrix was used to calculate the measures.

		Actual Value	
		Positive	**Negative**
Predicted Value	**Positive**	TP	FP
	Negative	FN	TN

TP – True Positive, FP – False Positive, FN – False Negative, TN – True Negative

Table 4. Confusion Matrix

- True positive (TP)- Number of samples with non – existence of heart disease is predicted as heart disease non - existence.

- False positive (FP) - Number of samples in the dataset with heart disease non - existence predicted as heart disease existence.
- True Negative (TN) - Number of samples in the dataset with the existence of heart disease predicted as the existence of heart disease.
- False Negative (FN) - Number of samples in the dataset with non- existence of heart disease predicted as the existence of heart disease.

The classification measures defined are as given below:
Accuracy the ratio of the correctly classified instance to the total number of instances is defined.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad \ldots (4)$$

Precision Number of positive samples is divided by the total number of positive samples predicted by the classification algorithm.

$$\text{Precision} = \frac{(TP)}{(TP + FP)} \quad \ldots (5)$$

Sensitivity or recall or true positive rate It is the proportion of positively identified instances.

$$\text{Sensitivity} = \frac{(TP)}{(TP + FN)} \quad \ldots (6)$$

Specificity It is the possible negative sample proportion with all negative instances.

$$\text{Specificity} = \frac{(TN)}{(TN + FP)} \quad \ldots (7)$$

Misclassification rate It is known as the ratio of wrongly classified samples to the total number of samples.

$$\text{MCR} = \frac{(FP + FN)}{(TP + TN + FP + FN)} \quad \ldots (8)$$

F1 Score It is a combined model of precision and recall. Its criterion is one for the best performances, and poor performance is zero.

$$\text{F1 score} = \frac{(2TP)}{(2TP + FP + FN)} \quad \ldots (9)$$

V. Results and Discussion

The performance of four machine learning algorithms to predict heart disease was evaluated using 17 HRV features. The dataset for training and testing is divided into two parts. The model was trained with 80% training data and tested the remaining 20% data. Tenfold validation method was used in this research to evaluate the efficiency of the developed method of classification.

k-Nearest Neighbors			
	Actual class		
	Non-Existence	Existence	Total
Predicted Class			
Non-Existence	141	7	148
Existence	12	153	165
Total	153	160	313

Normalization: Yes (Min-max Normalization), Number of neighbors: 9

Table 6. Confusion matrix of the k-NN algorithm

Support Vector Machine			
	Actual class		
	Non-Existence	Existence	Total
Predicted Class			
Non-Existence	127	21	148
Existence	12	153	165
Total	139	174	313

Normalization: Yes (Min-max Normalization), Kernel- Linear

Table 7. Confusion matrix of the SVM algorithm

Support Vector Machine			
	Actual class		
	Non-Existence	Existence	Total
Predicted Class			
Non-Existence	127	21	148
Existence	11	154	165
Total	138	175	313

Table 8. Confusion matrix of the NB algorithm

Support Vector Machine			
	Actual class		
	Non-Existence	Existence	Total
Predicted Class			
Non-Existence	143	5	148
Existence	15	150	165
Total	158	155	313

Table 9. Confusion matrix of the ANN algorithm

Confusion matrix of k-Nearest Neighbors (k - NN), Support Vector Machine (SVM), Naive Bayes (NB) and Artificial Neural Network (ANN) are computed in Table 6 through Table 9. The current study found that the k-nearest neighbors model achieved a classification accuracy of 0.94 with a precision of 0.95, sensitivity of 0.92, specificity of 0.96, F1 score of 0.94, MCR of 0.061 and AUC of 0.95; the artificial neural network (ANN) reach to an accuracy of 0.93, precision of 0.97, sensitivity of 0.97, specificity of 0.97, F1 score of 0.93, MCR of 0.064 and AUC of 0.95 (Table 11).

The performance measurement shown by the linear SVM kernel and Naive Bayes algorithm is similar, but when compared both algorithms accuracy, sensitivity is higher and lower MCR is found in Naive Bayes (Table 11). The predictive model of k - nearest neighbors and artificial neural

network showed a better performance than support vector machine and Naive Bayes. Moreover, in the k - nearest neighbors, the accuracy, precision, sensitivity, specificity, F1 score, AUC are higher and lower misclassification rate than the artificial neural network. Table 10 indicates the performance measure of the training dataset, and Table 11 indicates the performance measure of the testing dataset. The model must perform well on testing dataset; hence testing accuracy is very important.

Sr. No.	Predictive Model	Evaluation Parameter								
		Accuracy	Precision	Sensitivity	Specificity	F1 Score	MCR	Kappa	AUC	MSE
1	K-NN	0.93	0.97	0.89	0.97	0.93	0.074	0.852	0.97	0.271
2	SVM	0.90	0.91	0.89	0.91	0.90	0.099	0.801	0.94	0.315
3	NB	0.90	0.91	0.89	0.91	0.90	0.099	0.801	0.92	0.315
4	ANN	0.94	0.99	0.89	0.99	0.94	0.065	0.870	0.97	0.255

Accuracy, Precision, Sensitivity, and Specificity are considered in percentage (%).

Table 10. Classification performance measure- Training

Sr. No.	Predictive Model	Evaluation Parameter								
		Accuracy	Precision	Sensitivity	Specificity	F1 Score	MCR	Kappa	AUC	MSE
1	K-NN	0.94	0.95	0.92	0.96	0.94	0.061	0.878	0.95	0.246
2	SVM	0.88	0.86	0.91	0.88	0.89	0.11	0.787	0.94	0.324
3	NB	0.89	0.86	0.92	0.88	0.89	0.10	0.787	0.93	0.324
4	ANN	0.93	0.97	0.91	0.97	0.93	0.064	0.872	0.95	0.252

Accuracy, Precision, Sensitivity, and Specificity are considered in percentage (%)

Table 11. Classification performance measure- Testing

A. Performance Evaluation Using Kappa Statistics Value and Mean Square Error

The kappa statistics are often used to assess the reliability of the inter-rater [26]. The significance of rater reliability resides in the reality that it represents the magnitude of accurate depictions of the measured factors and the information gathered in the research [26, 27]. The Kappa outcome viewed as follows: values ≤ 0 as showing no agreement and 0.01–0.20 as none to slight, 0.21–0.40 as fair, 0.41–0.60 as mild, 0.61–0.80 as substantial, and 0.81–1.00 as almost perfect [26, 27]. Interpretation of the kappa statistic value is shown in Table 12.

Value of Kappa	Level of Agreement
0-0.20	None
0.21-0.39	Minimal
0.40-0.59	Weak
0.60-0.79	Moderate
0.80-0.90	Strong
Above 0.90	Almost Perfect

Table 12. Interpretation of kappa statistic value

Another performance metrics is the mean square error (MSE) to check the performance of the trained model over the test set. It is the mean of the squared difference between the actual and predicted values. MSE's lower value confers the best predictive model. The formula to find MSE is as follow-

$$MSE = \frac{1}{n}\sum_{i}^{n}(y_i - \hat{y}_i)^2 \quad \ldots (10)$$

where n is the number of data sample, y_i represents actual values and $\hat{y}_i$ represents predicted values. The model of SVM and Naive Bayes indicate a comparable value of kappa and MSE. Kappa value marginally improved in the k-NN model, and MSE much lower compared to the ANN model mentioned in Table (11).

B. Performance Evaluation Using Training Time and Prediction Time

The performance of a machine learning algorithm tested based on Training time and Prediction time. The model that requires less time to get the best model is predicted. The training time and prediction time depend on the processing speed of the processor. Table 13 shows the training and predicting time of the machine learning model. In this study, it is found that k-NN algorithm has less training time and prediction time.

	k-NN	**SVM**	**NB**	**ANN**
Training Time	0.00147	0.16215	0.01596	1.09927
Prediction Time	0.00044	0.00314	0.00143	0.0049

Time measured in seconds

Table 13. Performance measure based on time

C. Performance Evaluation Using ROC

Receiver operating characteristic curve is another classification measure. It is a plot between false positive rate (x-axis) and true positive rate (y-axis). ROC shows fluctuation

of both classes during training [25]. The area under the curve must be equal to one for best classifications. k-NN and ANN provide maximum value close to one. ROC curve of four predictive models are -

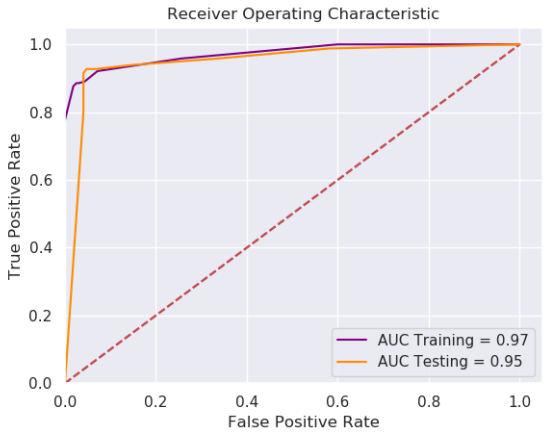

Figure 2. ROC curve for k-NN algorithm

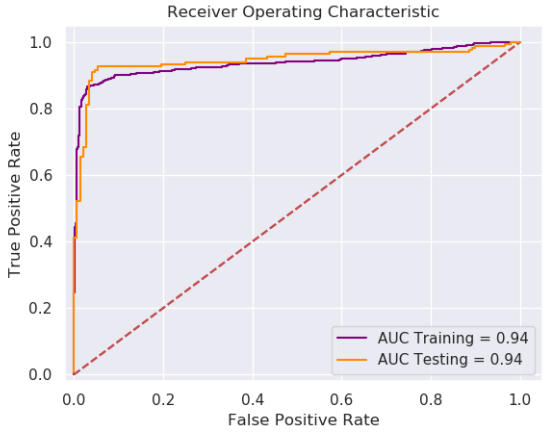

Figure 3. ROC curve for SVM algorithm

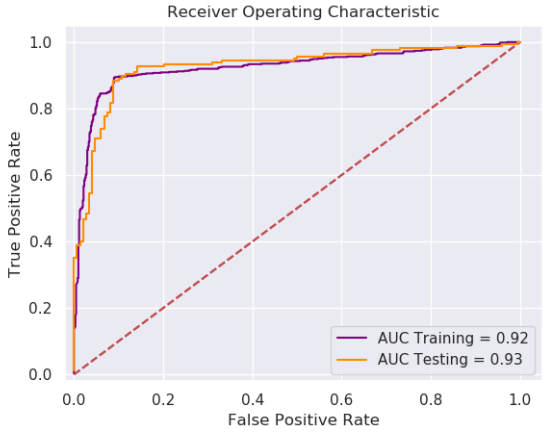

Figure 4. ROC curve for NB algorithm

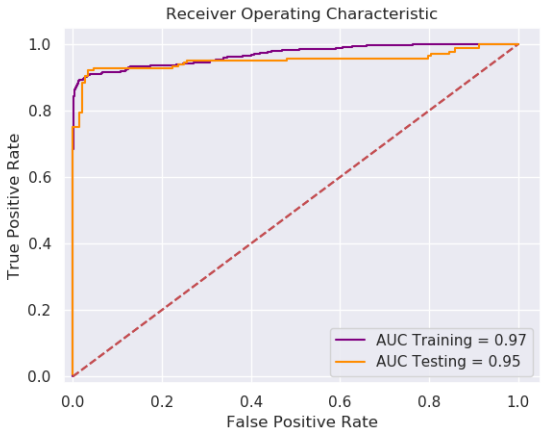

Figure 5. ROC curve for ANN algorithm

D. Hyperparameter optimization

Optimization or tuning of hyperparameters is the concern in machine learning to determine a set of perfect hyperparameters for learning algorithm. A parameter that measures the learning process using its value is a hyperparameter. They are metaparameters associated with learning algorithm. It is the best value for hyperparameters that generalizes the model for better accuracy. Before training the model explicitly the value of the hyperparameter can be changed manually by the machine learning expert. Machine learning model performance depends on the variety of hyperparameters such as hidden layers, multiple units per layer, activation function, regularizer, learning rate, number of neighbors, number of estimators and primary learners [4].

Algorithm	**Hyper parameters tuned to**
kNN	• N_neighbors = '9' • Weight = 'uniform'
SVM	• Kernel= 'linear' • Probability='True' • tol=0.001
NB	• Priors='array'
ANN	• Hidden_layer_sizes = (10,10,10) • max_iter=1000

Table 14. Hyper parameter of the model

In this research, kNN, SVM, NB, and ANN are machine learning algorithms. These models have default hyper-parameters, to achieve improved performance of model some of their hyper-parameters are tuned as shown in Table 14.

VI. Conclusion

Predicting heart diseases can save human lives and have a significant effect on their treatments. This study has implemented four machine learning technique for the prediction of heart diseases in smokers through Heart rate variability (HRV) parameter. These predictive models are evaluated on different performance measures. It is found that k-nearest neighbors show the highest classification accuracy.

This study can encourage researchers to select the ensemble and deep learning method to achieve more precise outcomes.

Conflict of Interest

No conflict of interest

References

[1] Barutcu I, Esen AM, Kaya D, Turkmen M, Karakaya O, Melek M, Esen OB, Basaran Y "Cigarette smoking and heart rate variability: dynamic influence of parasympathetic and sympathetic maneuvers." *Annals of non-invasive electroc*ardiology 10(3):324-9, 2005.

[2] Shashikant, R. and Chetankumar, P., March. "A Review on Impact Application of Heart Rate Variability (HRV)". In *International Conference on E-Business and Telecommunications* (pp. 1-8). Springer, Cham 2019.

[3] Dwivedi AK "Performance evaluation of different machine learning techniques for prediction of heart disease." *Neural Computing and Applications* 1:1-9,2018.

[4] R. Shashikant and P. Chetankumar, "Predictive model of cardiac arrest in smokers using machine learning technique based on Heart Rate Variability parameter", *Applied Computing, and Informatics*, 2019.

[5] Ajam N. "Heart Diseases Diagnoses using Artificial Neural Network". *Network and Complex Systems.*; 5(4):7-10, 2015.

[6] Nikhar S, Karandikar AM. "Prediction of heart disease using machine learning algorithms". *International Journal of Advanced Engineering, Management, and Science*; 2(6), 2016.

[7] Dinesh KG, Arumugaraj K, Santhosh KD, Mareeswari V. "Prediction of Cardiovascular Disease Using Machine Learning Algorithms". In *International Conference on Current Trends towards Converging Technologies (ICCTCT)*, Mar 1 (pp. 1-7), 2018.

[8] Sharmila S. "Analysis of heart disease prediction using data mining techniques". *International Journal of Advanced Networking & Applications (IJANA)*; 8(05):93-5,2018.

[9] Khanna D, Sahu R, Baths V, Deshpande B. "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*, Oct 1; 5(5):414, 2015.

[10] Ebrahimzadeh E, Pooyan M, Bijar A. "A novel approach to predict sudden cardiac death (SCD) using nonlinear and time-frequency analyses from HRV signals". *PloS one*.Feb 4; 9(2), 2014.

[11] Houshyarifar V, Amirani MC. "Early detection of sudden cardiac death using Poincaré plots and recurrence plot-based features from HRV signals". *Turkish Journal of Electrical Engineering & Computer Sciences,* Apr 10; 25(2):1541-53, 2017.

[12] Hussain L, Aziz W, Nadeem SA, Abbasi AQ. "Classification of normal and pathological heart signal variability using machine learning techniques". *International Journal of Darshan Institute on Engineering Research and Emerging Technologies*; 3(2):13-8(2014).

[13] Aha DW "Feature weighting for lazy learning algorithms." In Feature Extraction, Construction, and Selection 13-32.

[14] Kanmani S, Uthariaraj VR, Sankaranarayanan V, Thambidurai P "Object-oriented software fault prediction using neural networks." *Information and software technology* 49(5):483-92, 2007.

[15] Kandhasamy JP, Balamurali S "Performance analysis of classifier models to predict diabetes mellitus." *Procedia Computer Science* 47:45-51, 2015.

[16] Vapnik VN "An overview of statistical learning theory." *IEEE transactions on neural networks* 10(5):988-99.

[17] Vapnik V "The nature of statistical learning theory." *Springer science & business media*, 2013.

[18] Burges CJ "A tutorial on support vector machines for pattern recognition." *Data mining and knowledge discovery* 2(2):121-67.

[19] Heckerman D, Geiger D, Chickering DM ("Learning Bayesian networks: The combination of knowledge and statistical data." *Machine learning* 20(3):197-243.

[20] Jensen FV "An introduction to Bayesian networks." *London: UCL Press*.

[21] Patil TR, Sherekar SS "Performance analysis of Naive Bayes and J48 classification algorithm for data classification." *International journal of computer science and applications* 6(2):256-61, 2013.

[22] Bellazzi R, Zupan B "Predictive data mining in clinical medicine: current issues and guidelines." *International journal of medical informatics*. Feb 77(2):81-97, 2008.

[23] Lee TT, Liu CY, Kuo YH, Mills ME, Fong JG, Hung C "Application of data mining to the identification of critical factors in patient falls using a web-based reporting system." *International journal of medical informatics* 80(2):141-50, 2011..

[24] Lipton ZC, Elkan C, Naryanaswamy B "Optimal thresholding of classifiers to maximize F1 measure." In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* 225-239, 2014.

[25] Rice ME, Harris GT "Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r." *Law and human behavior* 29(5):615-20, 2005.

[26] McHugh, M.L.,"Interrater reliability: the kappa statistic*". Biochemia medica: Biochemia medica*, 22(3), pp.276-282.

[27] Marston, L., "Introductory statistics for health and nursing using SPSS". *Sage Publication*,2010.

[28] Boneau CA., "The effects of violations of assumptions underlying the t test". *Psychological bulletin*, 57(1):49-64.

Author Biographies

Rathod S R received M.Tech degree in Biomedical Instrumentation from College of Engineering Pune (COEP), India, in 2016. His research interest includes Artificial Intelligence, Machine learning, Ensemble Learning, Biomedical Electronics, Biomedical Instrumentation and Biomedical Signal Processing.

Patil C Y received his PhD from University of Pune, Pune, India. He is Professor in Instrumentation and Control Engineering department at College of Engineering Pune (COEP), India. His research interest includes Signal Processing and Soft Computing.