Volume 10, Issue 1
February 2021

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

Comparative Analysis of Feature Selection Based Machine Learning Methods for Heart Disease Prediction

[1]Suneeta Raykar and [2]Vinayak Shet

[1]Research Scholar, Department of Electrical & Electronics Engineering, Goa college of Engineering, Goa, INDIA

[2]Professor, Department of Electrical & Engineering, Goa college of Engineering, Goa, INDIA

E-mail: [1]suneeta@gec.ac.in, [2]vns@gec.ac.in

ABSTRACT

Smart healthcare application is an advanced field in the era of Internet of Things (IoT) technology. Digitization in healthcare focuses on predicting the health status of a patient with the aid of IoT based application. Machine learning algorithms can be applied to develop a decision framework to investigate whether an individual is likely to have heart disease or not. In the proposed study, machine learning algorithms are used in the evaluation of heart disease prediction framework. Feature selection (FS) method plays a vital role in improving the quality of prediction. Proposed prediction framework is analyzed using three important feature selection methods (six techniques) and seven supervised machine learning algorithms. Comparative performance analysis of machine learning algorithm is evaluated using most significant parameter selected using feature selection. The experiments are carried out using data set retrieved from Cleveland data set from UCI repository. Results show that Random Forest, a hybrid machine learning algorithm with LASSO regularization method for feature selection approach gives the best performance. Each algorithm's performance is measured using accuracy, precision, recall and specificity. Feature selection has an impact on classifier's accuracy and data storage. This approach can be further used in building IoT based heart disease decision support system with minimum number of sensor nodes to record the health parameters. Medical experts can diagnose the heart patient competently with the help of proposed method.

Keywords: *Internet of Things, Feature selection, Prediction, Machine Learning algorithms, Healthcare.*

1. INTRODUCTION

Owing to the current lifestyle and insufficient activities in day to day life is affecting human health. Heart being an essential organ in a human body which pumps blood throughout the body for the blood circulation is essential thus it's strength is to be preserved for a healthy living. According to the World Health Organization (WHO), every year more than 12 million deaths are occurring worldwide due to the various types of heart diseases which is also known by the term cardiovascular disease. WHO, has estimated that morality rate caused by heart diseases will becomes 23 million in a span of 20 years. The diagnosis of the heart diseases and treatment at early stage becomes essential and it is considered as difficult task in medical field. The symptoms of heart disease include shortness of breath, weakness of physical body, swollen feet, and fatigue with related signs [1]. Data set has many attributes related to a given disease. But not all the attributes have strong association with the output variable which decides prediction. Hence, finding the relevant attributes for a given prediction problem is important. In machine learning, accuracy improves if the data set has more features. However, including high dimensional data brings a high computational cost. High dimensionality lead to long training times, more features often lead to an algorithm overfitting as it tries to create a model that explains all the features in the data set. Feature selection is simply, the process of selecting most relevant attributes from the data set, which can be achieved by using filter method, wrapper method and embedded method. Feature Selection (FS) algorithms eliminate unrelated data from the healthcare data repositories and hence increase the performance of classifier accuracy used in different healthcare prediction model. It is essential to choose a feature selection algorithm carefully. The results obtained from different FS algorithms and classifier algorithm on heart disease data set with different number of features will also help researchers to find the best combinations of feature to design a framework for predicting heart disease. Selecting only most significant parameter helps to choose important sensors in IoT based health monitoring system.

2. RELATED WORK

Reviewing the related works in the field of data analytics and available machine learning methods in predicting heart disease are summarized. Data analytics is the development of originating knowledge from data, creating value like actionable insights. The Internet of Things (IoT) has been gaining drive in both the industry and research. Author Ehteshami A. Rezaei [2] proposed that an integration of clinical support with computer-based patient records could reduce medical errors, enhancement of patient safety, reduce undesirable laboratory tests there by improves patient outcome. Through immense literature survey, it is observed that early disease prediction is the most required area of research in healthcare sector. The data analysis involves more computational resources and devours much time when data set is of vast size. Therefore, feature selection procedure is used to eradicate the irrelevant or noisy features from the data [3] in order to reduce the

Volume 10, Issue 1
February 2021

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

time and the resource usage. Author Shanta Kumar B. Patil [4] obtained important patterns from heart disease database for heart attack prediction. Authors have proposed MAFIA algorithm (Maximal Frequent Itemset Algorithm) implemented using Java. The data is pre-processed first, and then clustered using K-means algorithm into two clusters and the cluster significant to heart attack is obtained. Then frequent patterns are mined from data set to know significant feature to be chosen. Features of the data set are given some ranking to understand its significance in determining output variable. G Purusothaman [5] has surveyed and compared different classification techniques for heart disease prediction. Hybrid model of machine learning algorithm is used instead of using single algorithm such as Decision tree, Naive Bayes, Artificial Neural Network. Authors have demonstrated the effectiveness of hybrid models. The performances of single models such as Decision tree, Artificial Neural Network and Naive Bayes are 76%, 85% and 69% respectively. However, hybrid approaches show an accuracy of 96%. Therefore, hybrid models lead to reliable and promising classifiers for predicting heart diseases with good accuracy. Feature selection in medical diagnosis helps to improve clinical decision quickly. Heuristic methods help to resolve the problem of selecting best features. Author Song [6] proposed a feature selection approach using filtering the signal to noise ratio (SNR) score and Particle Swarm Optimization (PSO). Clustering data set using K means and SNR score is used to rank genes. SVM, KNN and Probabilistic Neural Network (PNN) classification methods are used for testing the performance. Author Ambarasi enhanced the prediction of heart disease with feature subset selection based on a Genetic algorithm [7]. Genetic algorithm is used to determine the attributes which contribute more towards the diagnosis of heart ailments which indirectly reduces the number of tests which are needed to be taken by a patient. Thirteen attributes are minimized to six attributes using Genetic search. Subsequently, three classifiers like Naive Bayes, Clustering and Decision Tree are used to predict the diagnosis of patients with the same accuracy with fewer attributes. Author Nahar, Jasmine proposed various methods like Bagging, Boosting, Stacking in prediction and achieved 84.15% accuracy by applying Stacking technique SVM, MLP as best accuracy in predicting heart disease [8]. Author Rajagopal [9] presented a classification of cardiac arrhythmia using five different linear and non-linear unsupervised dimensionality reduction techniques combined with a probabilistic neural network (PNN) classifier. The PNN classifier and the fast-independent component analysis (fastICA) obtained the best result Author Asl [10] presented a classification that used fifteen features extracted from heart rate variability (HRV) signal. The authors reduced the features to five using a GDA technique and increased the accuracy to 100% when combined with the SVM classifier.

3. PROPOSED METHODOLOGY

This paper proposes building a predictive framework to investigate person getting chances of heart disease using machine learning approach. It describes the combination of classification algorithms with feature selection techniques to accomplish two main objectives, first objective is to select the most important features of the data set using feature selection method, and second objective is to use machine learning algorithm as a classifier to measure performance of prediction framework. Data set is obtained from the Cleveland heart disease data available from University of California Irvine (UCI) data mining repository [11].

Seven supervised machine learning algorithms used are Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Artificial Neural Network (ANN), Support Vector Machine (SVM), Naive Bayes (NB) and K's Nearest Neighbour (KNN). The feature selection technique used are filter method, wrapper method and embedded method. Seven selected algorithms are evaluated by considering minimum number of most important parameter retrieved using three different feature selection methods. Figure 1 illustrates the layout of proposed predictive framework. It depicts three blocks, block I selects appropriate data set of heart disease, block II performs feature selection algorithm, to obtain most important parameters from the data set and block III contains the selection of machine learning algorithms. Two methods are

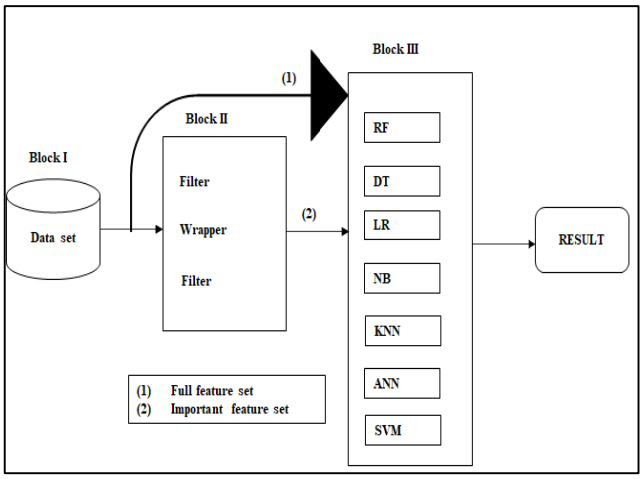

Fig. 1. Layout of proposed predictive framework

used to conduct the experimental analysis. Method 1 is represented as (1), which feeds the data set with all features to each machine learning algorithm to determine accuracy rate of prediction similarly method 2 shown as (2) feeds only most important parameters selected by the any one feature selection method used from block II. Performance of each machine learning algorithm is measured using accuracy, recall, precision and specificity.

Volume 10, Issue 1
February 2021

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

suggested for building IoT based heart prediction framework. IoT based healthcare prediction system, assists in recording. human body vital parameters using wearable sensors nodes and stores the collected data on smart gadgets like mobile device, tablet or personal computer system. The data set used is having 14 attributes. Pre-processing of data is carried out using MinMax Scalar to the data set for effective use in the classifiers. In this paper we have glanced upon seven different algorithms to predict the chance of getting heart disease. The algorithms used are listed as below.

A. Machine Learning Algorithms.

i. Logistic Regression is a statistical model that in its basic form uses a log function to model binary dependent variable. Regression can be defined by two categories; they are linear regression and logistic regression. Logistic regression basically is used to classify the low dimensional data having nonlinear boundaries. It also provides the difference in the percentage of dependent variable and provides the rank of individual variable according to its importance. Logistic regression is mainly used for prediction and also calculating the probability of success.

ii. Naive Bayes classifiers are also based on a statistical model. A Naive Bayes classifier assumes that all attributes are conditionally independent. This algorithm is built on conditional probability. The basic hypothesis is conditional independence so it is called "naive". The assumption that all input features are independent from one another assumes to be unrealistic but it performs surprisingly well on large data set where this condition is assumed to be holds good. Naive Bayes is mainly useful in recommending system and forecasting application.

iii. Support Vector Machine SVM is primarily a classifier method that performs classification tasks by constructing hyperplanes in a multidimensional space that separates cases of different class labels. SVM supports both regression, classification tasks and can handle multiple continuous and categorical variables. It uses a technique called the kernel trick to transform the input data into high dimensional feature spaces. Kernel used in SVM are linear, polylinear and radial basis function. The solution will not be stuck with local minima.

iv. K's Nearest Neighbour KNN classification, the output is a class membership. An object is classified by a plurality vote of its neighbors, with the object being assigned to the class most common among its K's nearest neighbors (K is a positive integer, typically small). If K equal to one then the object is simply assigned to the class of that single nearest neighbor.

v. Artificial Neural Network ANN is trained to predict the outcomes. The problem can be basically categorized into two types linearly separable and non-linearly separable problem. Most of the real-world problem belongs to non-linearly separable class problem. The solution to non- linearly separable problem needs multilayer perceptron model (MLP).

B. Data set Description

Data set has 303 records with 14 attributes. It has 165 instances with heart disease and 138 instances with normal cases. Data set has 14 different attributes. They are age of a person in years, sex, chest pain, resting blood pressure, cholesterol, fast blood sugar expressed as 1 or 0 (above 120mg/dl is 1), electrocardiographic (ECG) at rest position, heart rate or thalach, exercise induced angina, old peak, ca or major vessels coloured by fluoroscopy, thallium scan is a method of examining the heart to obtain information about the blood supply to the heart muscle, target variable to determine to have heart disease as 1 or 0. Target or output value 1 represents chances of heart disease and 0 represents healthy condition. Table I represents the range of all attributes used in data set. The attributes are expressed in nominal range.

TABLE I
DESCRIPTION OF DATA SET

Attributes	Description	Range (min-max)
age	Age of the person	29-77
sex	Gender (male/female)	1: male, 0:female
chest pain	pain type	1- 4
trestbps	resting blood pressure	94- 200 mm Hg
chol	cholesterol	120 – 564 mg/dl
fbs	fasting blood sugar	0 - 1
restecg	ECG at rest position	0-2
thalachh	max heartrate	71 - 202 BPM
exang	excercise induced angina	0-1
oldpeak	ECG ST segment	0 - 6.2
slope	ST waveform	1 - 3
ca	major vessels colour by floroscopy	0 - 3
thal	thallium scan	3- 7
target	Heart disease	1: Yes 0: No

C. Feature selection methods

As data generation and collection keeps increasing, visualizing it and drawing inferences becomes more and

more challenging. The medical data records used in predicting the chances of disease collected from Electronic Health Records (EHR) is having high dimension which needs to be reduced in order to reduce training time, improvement in accuracy of the model and overfitting issues. The basic steps involved in feature selection are generation of subset, evaluation of subset, stopping criterion and validating the subset [12]. Usually, three types of methods are used in feature selection filter method, wrapper method and hybrid (embedded) method.

- Filter method: The attributes are selected on the basis of their output results/scores from diverse statistical methods. These set of features are checked for their inter- dependency with output feature so that the correct subset is chosen. Filter method are Linear discriminate analysis (LDA), Pearson's correlation, Chi-Square method.
  - Chi Square test: The scikit-learn library provides the Select K Best class that can be used with a suite of different statistical tests to select a specific number of features. Chi-squared (chi) statistical test for deriving non-negative values for features. It selects eight of the best features from the data set, if Select K Best is set to eight.
  - Pearson's Coefficient: The filtering is done using correlation matrix and it is most commonly done using Pearson correlation. The result is further an alyzed by plotting the Pearson's correlation heatmap and the correlation of independent variables with the output variable helps to select the feature subset. The correlation coefficient has values between -1 to 1.
  - Feature importance: Feature importance of each feature of the data set is calculated by using the feature importance property of the model. Bagged decision trees like Random Forest and Extra Trees can be used to estimate the importance of features.

- Wrapper Method: They will create a subset of features from the data set in training the prediction model. Accuracy of the model, is validated by considering the features either by adding or removing. This iteration goes until, the best subset is found. of these types of methods are forward feature selection, backward feature elimination, recursive feature elimination methods.
  - Backward selection: This is implemented by feeding all the possible features to the model at first. The performance of the model is then checked by iteratively removing the worst performing features one by one till the overall performance of the model comes in acceptable range value.
  - Recursive Elimination method (RFE): It recursively removes the attributes and building a model on those attributes that remain. It uses accuracy metric to rank the feature according to their importance.

- Embedded Method: This is a combination of both filter and wrapper methods. Algorithms have their own built in feature selection criteria. This helps in generating the best subset and provide the same to the training model. This type of method normally gives much more accurate prediction. Embedded (hybrid) method include
  - LASSO regularization (L1), Ridge regularization (L2) and combination of both Lasso and Ridge regularization. LASSO Regularization: Least absolute shrinkage and selection operator is the selection performed on the selected data source. If the feature is irrelevant, lasso penalizes its coefficient and make it 0. The features with coefficient equal to 0 are removed and the remaining features are considered.

4. PERFORMANCE ANALYSIS

Performance of each algorithm is determined using processed data set. The simulation is carried out using Spyder software on Windows operating system. Software used is Spyder, Scientific Python Development Environment, is a free integrated development environment (IDE) that is included with Anaconda Navigator. Anaconda is a free and open Python programming language toolkit. Spyder features a unique combination of the advanced editing, analysis, debugging and profiling functionality of a comprehensive development tool with the data exploration, interactive execution, deep inspection and beautiful visualization. Spyder offers built-in integration with many popular scientific packages, including NumPy, SciPy, Pandas, IPython, QtConsole, Matplotlib, SymPy. Data mining algorithm for feature selection is implemented using Spyder tool. The experimental analysis carried out using scenario I and II.

A. Experimental Scenario - I

Selected data set is pre processed and applied to six different feature selection algorithms to select eight most significant parameters from the data set. Results obtained using various feature selection method is tabulated in Table II. The feature selection methods used for selecting the most important parameter are Chi square method, Pearson's coefficient, Feature importance, Backward elimination, Lasso and Recursive Elimination (RFE) method.

Most important parameters selected from all the above method commonly selected the following attributes from the data set, they are sex, chest pain, thalach, restecg, ca, exang and oldpeak. This clearly demonstrates that data obtained from ECG sensor heart rate sensor plays crucial role in deciding the rate of heart

Volume 10, Issue 1
February 2021

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

disease prediction. The analysis of accuracy and other parameter of machine learning is carried out. All the selected seven machine learning algorithm have been tested on full features (13 features) from selected data set.

TABLE II
FEATURE SELECTION METHODS USED

Methods	Selected Attributes
Chi-square	thalach, oldpeak, ca, cp, exang, chol, age, trestbps
Pearson's coefficient	cp, thalch, exang, oldpeak, slope ca,thal
Feature importance	cp, ca, thalach, oldpeak, thal, age, chol, trestbps
Backward elimination	sex, cp, ca, thal, oldpeak, exang, thalach, trestbps
RFE	sex, cp, restecg, exang, oldpeak, slope, ca, thal
LASSO	sex, exang, thal, ca, oldpeak, thalach, cp, slope

B. Experimental Scenario II

The most significant parameter selected from feature selection method listed in Table II is further used to train the seven selected machine learning algorithms. Figure 2 demonstrates the flow diagram used to evaluate the accuracy of the prediction framework. Random sampling is used to select data set into two vectors. Training data and test data vectors. Random sampling method of selecting train and test vectors eliminates memorizing the patterns. The processed data set is applied to machine learning algorithm. The training of algorithm is carried out using various algorithm listed above.

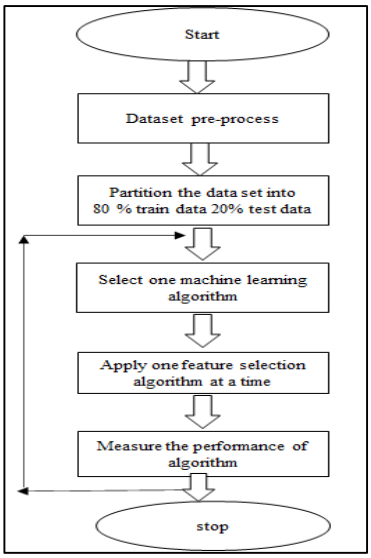

Fig. 2. Flow diagram to evaluate prediction framework

The results obtained using each of the machine learning algorithm using python code is expressed using confusion matrix as shown in Table III. Confusion matrix demonstrates The result of the algorithm is tested against new test vector whose output variable is predicted by learned algorithm at the time of training. In supervised algorithm target variable is given at the time of training. the number of instances which are unhealthy (sick) and diagnosed accurately. Confusion Matrix is used to specify information about actual and predicted classification done by a classification. Table III represents the information present within a confusion matrix. The entries in the confusion matrix have the following meaning

- TN is the number of correct predictions that an instance is negative.
- FP is the number of incorrect predictions that an instance is positive.
- FN is the number of incorrect of predictions that an instance negative.
- TP is the number of correct instances that an instance is positive.

TABLE III
CONFUSION MATRIX

	Negative	Positive
Negative	TN	FP
Positive	FN	TP

False positive (FP) demonstrates the number of instances which are healthy and diagnosed wrongly as they are un- healthy (having chances of heart disease). False negative (FN) demonstrates the number of instances which are sick but the instances are diagnosed wrongly as healthy. True negative (TN) contains a number of instances which are healthy and the instances are diagnosed accurately. True positive (TP) contains number of instances which are unhealthy and predicted also as unhealthy patterns. The accuracy of the network is calculated as ratio between total instances predicted correctly to total number of all instances recorded in the confusion matrix. N is number of samples in the given data set (303 samples) . Recall is the ratio of correctly classified to actual positive instances or having heart disease records. Precision is the proportion of the predicted positive cases that are correct. Specificity determines among healthy instances how many of such instances model predicted perfectly. Performance parameters of the machine learning algorithm is calculated with the help of confusion matrix are

$$\text{Accuracy} = (TP + TN)/ (TP + TN + FP + FN)$$

$$\text{Recall} = TP/ (TP + FN).$$

$$\text{Precision} = TP/ (TP + FP).$$

$$\text{Specificity} = TN /(TN+FP)$$

The parameter setting used for each algorithm is given in Table IV. Number of iterations, types of Kernel, number of nodes

Volume 10, Issue 1
February 2021

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

are modified to obtain good accuracy results. Final values used in the machine learning algorithm is tabulated to achieve best performance.

TABLE IV
MACHINE LEARNING TUNING PARAMETERS

Algorithm	Tuning Parameters
Decision Tree	iterations 2000
Random Forest	iterations 2000
SVM	Kernel Linear
KNN	Neighbours 05
ANN (I)	hidden nodes 8 input nodes 13, iteration 200
ANN (II)	hidden nodes 6 input nodes 8, iterations 200

The Random Forest and Decision Tree give best accuracy rate by setting number of iterations used in training as 2000. Support Vector Machine yields best result when kernel function is set as Linear, other kernel function tested are radial basis and poly linear. Number of neighbors used in KNN are five. Artificial Neural Network with multilayer perceptron model is set as 13 input neurons, one hidden layer with 8 neurons and one output layer during full feature set and ANN-II, used 6 neurons at the hidden layer when input neurons are reduced to eight most significant parameters.

The figure 3 shows the performance of selected machine learning algorithm using dataset with all features.

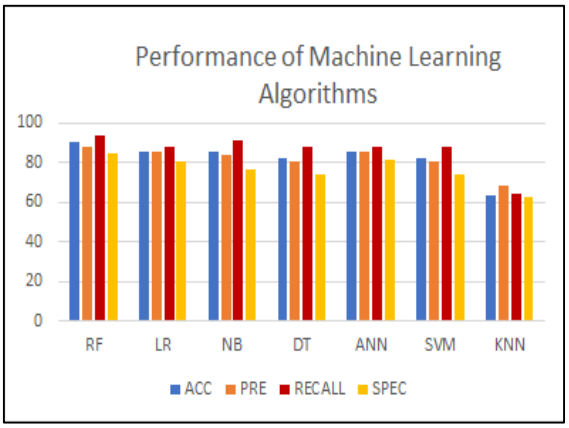

Fig. 3. Performance of machine learning algorithms with all features.

The bar graph shows accuracy (blue color), precision (orange), recall (red) and specificity (yellow) of each algorithm respectively. X axis shows various algorithm with performance parameter and Y axis shows the accuracy score obtained in percentage. Graph illustrates that accuracy and recall and specificity factors are high in Random Forest, precision of Decision Tree and Support Vector Machine algorithms shows better result.

This result obtained using method 1 are tabulated in Table V. It illustrates performance parameters of each algorithm calculated using complete features of the dataset. The performance of LASSO regularization shows accuracy (Acc) of 90.16% prediction rate in detecting heart disease. Other parameter precision (Pre), recall (Rec) and specificity (Spec) are 88.18%, 94.11%, 85.18% respectively. The value obtained using other algorithms are also tabulated. KNN shows least performance in prediction of heart disease.

TABLE V
PERFORMANCE MATRIX – I

Algorithm	Performance Matrix-I			
	Acc	Pre	Rec	Spec
Random Forest	90.16	88.18	94.11	85.18
Decision Tree	81.97	81.04	88.29	74.07
Logistic Regression	85.25	85.71	88.23	81.48
Naive Bayes	85.25	83.78	91.17	77.77
Neural Network	85.25	85.72	88.23	81.48
SVM	81.97	81.04	88.23	74.07
KNN	63.93	68.75	64.70	62.96

Table VI gives insight of performance observed by each algorithm when feature selection methods are applied. Per- formance matrix of optimal values obtained by each machine learning algorithm clearly illustrates that LASSO method is giving good prediction accuracy. Recursive feature elimination method (RFE) also gives satisfactory response.

TABLE VI
PERFORMANCE MATRIX -II

Algorithm	Performance Matrix-II			
	Acc	Pre	Rec	Spec
Random Forest (Lasso)	90.16	77	91	88
Decision Tree (Feature importance)	86	83	91	77
Logistic Regression (RFE)	85.26	83	91	77
Naive Bayes (RFE)	83.6	83	88	77
Neural Network (RFE)	80.33	82	82	77
SVM (Feature importance)	85	83	91	77
KNN(Lasso)	82	82	85	77

The prediction accuracy rate does not differ to large extent by reducing number of feature set from given dataset. Each algorithm's accuracy is different when method of feature selection is changed. The bar graph in figure 4 gives various algorithms vs accuracy obtained when six different feature selection methods are applied. Graph legend shown in figure 4

Volume 10, Issue 1
February 2021

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

show various feature selection method applied to measure accuracy rate of prediction.

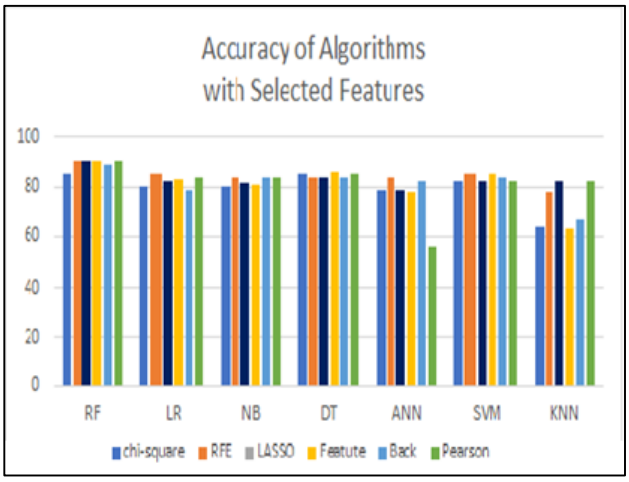

Fig. 4. Accuracy of Algorithms with selected features.

Most common attributes selected using six different feature selection process used in prediction illustrated that chest pain, sex, heart rate, ECG parameter like slope, old peak are very crucial parameters. Evaluation of prediction accuracy between various feature method and machine learning method helps to select suitable algorithm and features selection method to build prediction framework. Experimental results of the proposed work show that size of the deduced data set was 45% less as compared to original data set. This makes a remarkable significance when large data sets are used for developing prediction frame work. The size of original data set consumed 11 KB (14 attributes) space on hard disk and reduced data set using LASSO method consumed (9 attributes) nearly 6 KB.

5. CONCLUSION

The objective for precisely predicting existence of heart disease with the condensed number of attributes was achieved in this paper. Initially, data set has thirteen attributes involved in predicting the heart disease. In the proposed work LASSO regularization is selected among other feature selection method to determine the most significant features which contribute more towards output variable. Thirteen attributes given in the selected data source is reduced to eight important attributes. Seven supervised machine learning algorithm such as Random Forest, Naive Bayes, Decision Tree, Logistic Regression, Support Vector machine, KNN are used to predict the diagnosis of patients is evaluated using all thirteen attributes and only eight important attributes. Observations show that the Random Forest algorithm give good results. Additionally, novel feature-selection methods can be established to get a wider awareness of the important features to intensify the performance of heart disease prediction. The designed framework can be further used to receive test data from wearable sensor nodes to predict the outcome. IoT based health monitoring system collects and store the sensor data in the cloud. Work can be extended to develop Graphical user interface (GUI) which uses Random Forest algorithm with Lasso method as a backend for training the model which further can predict the results for new test parameters obtained from wearable sensor nodes as test vectors. Employing effectual, modernized techniques will assist quick and accurate diagnosis. Proper treatment, early diagnosis and reduced expenses helps in improving healthcare facility in future.

REFERENCES

[1] Subanya B and Rajalaxmi R, “A Novel Feature Selection Algorithm for Heart Disease Classification”, International Journal of Computational Intelligence and Informatics, Vol. 4(2), pp. 117–124, 2014.

[2] Ehteshami, Asghar, Peyman Rezaei, Nahid Tavakoli, and Mahtab Kasaei, “The role of health information technology in reducing preventable medical errors and improving patient safety”, International Journal of Health System and Disaster Management, vol.1, pp. 4-195, 2013.

[3] Sahu, Barnali, and Debahuti Mishra, “A novel feature selection algorithm using particle swarm optimization for cancer microarray data”, Procedia Engineering, vol. 38, pp. 27-31, 2012.

[4] Shantakumar B. Patil,and Dr. Y. S. Kumaraswamy, “Extraction of Significant Patterns from Heart Disease Warehouses for Heart Attack Prediction”, IJCSNS International Journal of Computer Science and Network Security, vol. 9 , no. 2, pp. 228-235, 2012.

[5] G. Purusothaman, and P. Krishnakumari, “A Survey of Data Mining Techniques on Risk Prediction Heart Disease”, Indian Journal of Science and Technology, vol. 8(12), pp. 1-5, 2015.

[6] Song, M. H, Lee.J, Cho. S.P, Lee. K.J and Yoo. S.K, “Support vector machine Based Arrhythmia classification using reduced Feature”, International Journal of Control, Automation and Systems, vol 3, pp. 571-579, 2005.

[7] M. Anbarasi, E. Anupriya, N. Ch. S. N. Iyengar, “Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm”, International Journal of Engineering Science and Technology, vol. 2(10), pp. 5370-5376, 2010.

[8] Nahar, Jesmin, “Computational intelligence for heart disease diagnosis A medical knowledge driven

approach", Expert Systems with Applications, vol. 40(1), pp. 96-104, 2013.

[9] Rajagopal, R. and Ranganathan, V, "Evaluation of effect of unsupervised dimensionality reduction techniques on automated arrhythmia classification", Biomedical Signal Processing and Control, vol.34, pp.1-8, 2017.

[10] Asl, BM, Setarehdan, S.K,Mohebbi, M, "Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal", Artificial Intelligence in Medicine, vol.44(1), pp.51-64 2008.

[11] UCI Machine Learning Repository: Heart Disease Data Set 2018. [Online]. Available: http://archive.ics.uci.edu/ml/data sets/heart+disease

[12] Dash, Manoranjan, and Huan Liu, "Feature selection for classification", Intelligent data analysis, vol.1, no.3, pp. 131-156, 1997.

AUTHOR PROFILES

Suneeta S Raykar graduated from Electrical & Electronics Engineering, from Karnataka University Dharwad, in the year 1993. She completed M.Tech in System Analysis and Computer Application from National Institute of Technology Surathkal Karnataka in the year 2004. Presently she is pursuing Ph.D. at Goa college of Engineering, under Goa University. She has 20 years of teaching experience in Department of EEE at Goa College of Engineering Goa India. Her research interest includes Soft computing, Computer Network, Network Security, Embedded System. She has guided more than 30 under graduate projects and 5 post graduate projects.

Vinayak N. Shet received his B. E, M. Tech and Ph.D. in Electrical Engineering, Power Electronics and Power System and Electrical Engineering from Karnataka University, IIT Bombay, IIT Bombay, India in 1984, 1995 and 2002 respectively. He is currently working as Professor in Electrical & Electronics department at Goa College of Engineering, Farmagudi, Goa. He worked as Lecturer, Assistant Professor, Professor, Head of Department and then as Principal of Goa College of Engineering during 1986-95, 1995-2003, 2003-14 and 2014-18 respectively. He is the Chief Investigator for Information Security Education and Awareness under the Department of Information technology, Ministry of Electronics and Communication, Government of India. He is actively involved as reviewer for various Journals and Conferences. He has published number of papers in various international Journals and Conferences. He has been awarded with patents based on his research work. His area of interest is Power Factor Correction, switch mode converter, power quality mitigation, grid connected systems.