Comparison of Coronary Heart Disease Prediction models using various Machine Learning Algorithms

Sunil Kr. Tiwari* and Suresh Kumar Garg**

Department of Mechanical Engineering, Delhi Technological University, Delhi and Delhi Skill & Entrepreneurship University, Delhi
**Department of Mechanical Engineering, Delhi Technological University, Delhi*
Corresponding Author : sktiwari262001@yahoo.com

ABSTRACT

In the health sector, Data Analytics and Machine Learning (ML) methods are taking over role of skill and experience of a doctor especially in diagnosing diseases and preventive health measures. The health care industry is collecting very large amount of data related to patients, his medical history for preventive medication and diagnosing disease well in time and more accurately. In this paper, a comparison of five classification machine learning methods viz. Decision Tree, Random Forests, Support Vector Machine, Artificial Neural Network and Fuzzy Logic based soft computing method is done for heart disease diagnosis on the basis of data available on public domain. Out of 76 parameters collected for a patient, only 15 medical parameters such as blood pressure, sex, age, obesity and cholesterol level are used for predicting heart disease of patients.

Keywords: Artificial neural network (ANN); Coronary heart disease (CHD); Machine learning (ML); Random forest (RF); Decision tree (DT).

INTRODUCTION

There are many advances in medical services to treat patients with various diseases. However, correctly diagnosing disease and administering effective treatments of a patient is still a challenging task. Clinical decisions due to incorrect or delay in diagnosis of disease may end up resulting in death of patient, it also leads to loss of reputation of a hospital. Medical expenses for treating a heart disease patient is quite high, and sometimes not affordable by poor patients especially in country like India, having a large demographic diversity. Various Machine Learning methods can be used for exploring Computer-based information for achieving a correct and cost-effective treatment. Most of the hospitals maintain healthcare data/ history of a patient for their record as well as for easy and better diagnosis of disease of their patients. These information systems of hospitals generate large amount of data containing numbers, text, charts, images etc.

Predicting risk of Coronary Heart Disease (CHD) is a complicated task in medical field as it is closely related with lifestyle of a person. CHD is a type of disorder that influence capability of a heart to function normally. It develops when the coronary arteries supplying oxygen & blood to heart, gets contracted due to deposition of cholesterin on these artery walls. This leads to reduction in flow of blood to heart. A clot may block the flow of blood which may cause serious health issues. In case of narrowing down of these arteries, oxygen rich blood may not be received by the heart especially during physical activity, and it may lead to heart attack.

Several research has been carried out to help healthcare experts in diagnosing heart disease using statistical

and data mining tools. As per healthcare experts, major reasons for heart disease are poor life style, irregular & unhealthy diet, harmful use of tobacco, alcohol etc. These factors directly affects some body parameters like age, blood sugar, blood pressure, smoking habit, hypertension, total cholesterol, obesity, serum cholesterol levels, lack of physical activity and family history of heart disease etc and a variation can be recorded in these parameters. This paper is focusing on applying some Machine Learning techniques on these available parameters and try to find relation between these parameters and risk of heart disease. It will try to find a Machine Learning technique which has better accuracy in diagnosing the heart disease at an early stage, so that patient can be given treatment well in time and risk factor of life due to heart disease can be reduced. This paper is elaborated in seven sections. In section 2, work of previous researchers is discussed. In section 3, Problem is stated i.e. parameters that may cause heart disease and its affect on diagnosis of heart disease is defined. In section 4, different machine learning methods are elaborated. In section 5, different parameters of database are defined and elaborated. In section 6, results from different machine learning techniques are presented and discussed on the basis of their merits and demerits. The paper is concluded in section 7.

LITERATURE REVIEW

There are many factors which affects functioning of heart. These factors include week hard muscles, congenital heart defects, cardiovascular disease, blood vessel disorders such as irregular heart beat issue, and coronary artery diseases. CHD is a disease of heart which causes reduction in blood flow to the heart which leads to heart attack. Many authors applied different machine learning techniques on the open UCI datasets especially on Cleveland datasets and the results were compared on the basis of prediction accuracy for heart disease.

Rani, K., Chaitanya et al., 2018 proposed a rule based framework using combinations of SVM, decision trees, and Logistic Regression to predict heart disease and compared accuracy of these rules to the results of applying only Logistic regression.

Kumar, S. et al., 2018 applied enhanced decision tree algorithm for generating the decision rules and optimizing it using genetic algorithm for predicting heart disease. This proposed algorithm was applied on data sets from University of California, Irvine and it showed high accuracy as compared to SVM and Decision Tree (C4.5) algorithm. Simplicity of this proposed algorithm makes it ideal for pattern recognition applications.

Celik, S. et al., 2018 applied Multivariate Adaptive Regression Splines and Classification & Regression Tree approaches for estimating important variables in prediction of body weight of Turkish Tazi dogs and concluded that body weight of these dogs can be predicted using MARS algorithm which is more efficient compared to CART algorithm.

Narayan, S. et al., 2018 used decision tree and fuzzy algorithms along with cuckoo search (CS) algorithm as a hybrid classifier for heart disease problem.

Maheswari, S. et al., 2018 developed a windows based application through the intelligent prediction system using ID3 (Decision tree) & Navie Bayes techniques for instance guidance on heart disease to user.

Alsina E. et al., 2017 predicted reliability in 19 industrial components using supervised machine learning (ML) approach and concluded that machine learning models (especially random forest (RF) algorithm) predicts better results as compared to traditional approaches (e.g., Weibull distribution).

Kim, J. et al., 2015 formulated a rule-based predicting model applying decision tree (CART) and fuzzy logic for datasets derived from KNHANES-VI i.e. Korean National Health and Nutrition Examination Survey VI.

Akila, S., 2015 proposed two stage hybrid method integrating Multi Layer Perceptron (MLP) and Decision Tree for diagnosis of CHD risk. In first stage, CHD risk was identified on the basis of physical and bio-chemical factors using Decision Tree C4.5 algorithm. In the second stage, level of Coronary Heart Disease risk identified instances from decision tree are analyzed using MLP with medical history and habitation attributes.

Ramakrishnan, S. et al., 2015 applied decision tree model for accurate prediction of corporate bankruptcy in different industries and reduction in costs of risk management. This paper attempts to detail the characteristics of each industry using sector indicators. It was concluded that there is significant relationship between probability of default and sector indicators.

Al-batah, M., 2014 used Classification and Regression Tree (CRT) to predict heart disease and determine attributes contributing in diagnosis of heart ailments.

Mythili T. et al., 2013 proposed a rule based model using hybrid data mining algorithms to compare accuracy of applying rules to the individual results of SVM, logistic regression and decision trees to present an accurate heart disease prediction model.

Dhanashree S. et al., 2013 used Naive Bayes for classification purpose and presented a classifier approach for detecting heart disease.

Ishtake SH et al., 2013 developed a computer-based information and DSS (decision support system) for solving multi-layered problem of diagnosing heart disease based on the huge amount of data collected by hospitals in the form of numbers, text, charts and images.

Anooj, P., 2011 for heart disease diagnosis, presented a weighted fuzzy rule-based clinical DSS on the basis of the patient's clinical data and compared it with the neural network-based system on the basis of specificity, sensitivity and accuracy. The proposed clinical DSS consists of two phases: (i) developing a fuzzy rule-based DSS using Mamdani fuzzy inference system and (ii) automated approach for generation of weighted fuzzy rules using mining techniques, attribute selection and attribute weightage method.

Ozsen S et al., 2009 determined weights of attributes in artificial immune systems (AIS) using genetic algorithms and evaluated it with two datasets i.e. Statlog Heart Disease and BUPA Liver Disorders datasets.

Polat K et al., 2007 used k-Nearest Neighbour (k-NN) as pre-processor and Artificial immune recognition system (AIRS) with fuzzy resource allocation mechanism as main classifier for diagnosis of heart disease and obtained 87% classification accuracy.

Table 1. Comparative study of research articles related to prediction models for diagnosing coronary heart disease with various machine learning methods

Name of Author	Database used	No of datasets used	No of parameters used	Methodology used	Prediction Accuracy/ Remarks
Rani, K., Chaitanya et al., 2018	Cleveland Database	303	13	a rule based framework using combinations of SVM, logistic regression, and decision trees	compared accuracy of the rules to individual results of logistic regression
Kumar, S. et al., 2018	UCI breast cancer dataset	286	34	enhanced decision tree algorithm	77.4
Narayan, S. et al., 2018	Cleveland Heart Disease Database	303	13	fuzzy and decision tree algorithm combined with cuckoo search (CS) algorithm as a hybrid classifier	-
Maheswari, S. et al., 2018	Cleveland Heart Disease Database	303	13	Decision tree (ID3) and Navie Bayes techniques	-
Kim, J. et al., 2015	Korean National Health and Nutrition Examination Survey VI (KNHANES-VI)	748 (70% for training, and 30% for a testing)	6	rule-based fuzzy logic and decision tree	69.51
Akila, S., 2015	Datasets from Institute of Road and Transport Perundurai Medical College	375	6	two stage hybrid method integrating Decision Tree (C4.5 algorithm) and Multi Layer Perceptron	96.66
Al-batah, M., 2014	Statlog Heart Disease datasets	270 (90% for training, and 10% for a testing)	13	Classification and Regression Tree (CRT)	prediction accuracy is 90.3%

Mythili T. et al., 2013	Cleveland Heart Disease Database	297	13	rule based model using hybrid data mining algorithms	-
Dhanashree S. et al., 2013	Cleveland Database	303 (303 for training and 240 for testing)	13	Naive Bayes for classification	89.58
Ishtake SH et al., 2013	Cleveland Heart Disease Database	909 (data for training and testing is 1:1)	13	computer-based data and Systems supporting decision for finding relation ship of heart disease with various parameters	-
Anooj, P., 2011	Cleveland Database	720 (data for training and testing is 2:1)	6 to 8	weighted fuzzy rule-based clinical decision support system	62.35 (for clevland dataset of 303 rows)
Ozsen S et al., 2009	Statlog Heart Disease datasets	270	13	genetic algorithm	87.43
Polat K et al.,2007	Statlog Heart Disease datasets	270	13	k-nearest neighbour (k-NN) & Artificial immune recognition system (AIRS) with fuzzy resource allocation mechanism	87% classification accuracy

PROBLEM STATEMENT

With the advancement in healthcare industry, diagnosing a disease at an early stage on the basis of body symptoms and body parameters is a challenging task for doctors. To meet this challenge of diagnosing a disease, Machine Learning (ML) is playing an important role in predicting Locomotor disorders, Heart diseases and may more diseases. Prediction of such diseases well in advance, may provide important insights to doctors and they can then accordingly adapt their diagnosis process and its treatment.

In the present day, heart disease and cancer are two major diseases which if not diagnosed in time may cause severe impact on life of a person. In this research paper, different machine learning methods are applied on data related to heart patients and results of all these machine learning methods are compared. Though 76 parameters of a patient were collected, but for for diagnosing heart disease only 14 parameters are found relevant. These parameters are defined in detail in next section.

Python language and its package is a open source code language and is easy to use, due to which program is developed in Python language for applying different machine learning method. Following libraries/packages of python are used in developing program of five different machine learning methods for prediction, finding accuracy and data visualization:

1. **Pandas**: It is used for reading data file, for analyzing data, create multi-dimensional data-tables

2. **Numpy**: It helps in managing multi-dimensional arrays very efficiently.

3. **Scikit-Learn (sklearn)**: It has several methods including regression methods, classification methods, and clustering. It can also be used for train-test split of data, model validation and model selection, dimensionality reduction and feature extraction etc.

4. **Scipy**: It provides efficient solutions for various numerical routines (mathematical formulae for machine learning models). These includes optimization, integration, interpolation, etc.

5. **PyTorch**: It provides two high-level features: Deep neural networks (built on a tape-based autograd system) and Tensor computation (like NumPy).

6. **TensorFlow**: It is used for numerical computation via utilising the graphs of data flow. Here we represents mathematical operations by graph nodes, and the tensors (arrays of multidimensional data) that flow inbetween them is represented by graph edges. Due to this computation, and flexible architecture can be distributed systematically and strategically to one or more GPUs and CPUs without editing the existing code.

7. **Matplotlib**: It is mainly used for data visualization, plotting the graphs etc.

DESCRIPTION OF DATABASE

The database used in this paper is available in public domain and consists of records of patients from V.A. Medical Center, Long Beach and Cleveland Clinic Foundation. This database was created in July 1988 and was used by many researchers using machine learning methods. Though there are 76 medical parameters, but only 14 medical parameters were applied for diagnosing the diseases of heart. The combined database from all these four sources contains 920 rows (i.e. record of 920 patients). Presence of disease of heart in patient is referred by "num (goal)" which is mentioned in the last column. It is an integer with values from 0 to discern existence. This value is considered as output and other 13 parameters as input variable. Details of these 14 medical parameters used in this paper from the available database is described in Figure 1.

MACHINE LEARNING METHODS

In our present systematic investigation, we have used six ML methods for diagnosing heart problem i.e. Linear Regression, Decision Tree, ANN, SVM, Random Forest, and Soft Computing methods. Brief description of all these methods are as under:

1. **Logistic Regression**: It is a prediction based analysis & is applicable in case of dichotomous (binary) dependent variable i.e. binary output. It is commonly used due to its low calculation time to describe data and relating one dependant binary variable with one or more autonomous variables that may be ratio-level type, interval, ordinal or nominal/.

2. **Decision Tree:** Decision tree, also referred as CART (Classification and Regression Trees), is a decision support tool used for classification or regression of decisions with conditional control statements. There are four main components of decision tree -

i. "test" is represented by a Node on an attribute,

ii. Branch represents the outcome of test,

iii. Leaf represents a decision of class label

iv. Link/ paths from root to leaf representing classification rules.

For predictive models, it is believed to be the better monitored learning method with ease of interpretation, along with high stability and accuracy.

S No	Parameter	Sequence in database	Parameter Description
1.	age	3	age in years
2.	sex	4	sex (1 = male; 0 = female)
3.	cp	9	chest pain type -- Value 1: typical angina -- Value 2: atypical angina -- Value 3: non-anginal pain -- Value 4: asymptomatic
4.	trestbps	10	resting blood pressure (in mm Hg on admission to the hospital)
5.	chol	12	serum cholestoral in mg/dl
6.	fbs	16	fasting blood sugar > 120 mg/dl (1 = true; 0 = false)
7.	restecg	19	resting electrocardiographic results -- Value 0: normal -- Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) -- Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
8.	thalach	32	maximum heart rate achieved
9.	exang	38	exercise induced angina (1 = yes; 0 = no)
10.	oldpeak	40	ST depression induced by exercise relative to rest
11.	slope	41	the slope of the peak exercise ST segment -- Value 1: upsloping -- Value 2: flat -- Value 3: downsloping
12.	ca	44	number of major vessels (0-3) colored by flourosopy
13.	thal	51	3 = normal; 6 = fixed defect; 7 = reversable defect
14.	num (goal)	58	diagnosis of heart disease (angiographic disease status) -- Value 0: < 50% diameter narrowing -- Value 1: > 50% diameter narrowing (in any major vessel: attributes 59 through 68 are vessels)

Figure 1: Description of important parameters of datasets (from UCI datasets)

3. **Support Vector Machine:** SVM analysis is applied for classificiying and regression. It is a non-parametric technique as it relies on kernel function. Nonlinear boundaries are produced by constructing linear boundaries in a large. As the output in SVM analysis can be any rational or irrational number, predicting the information becomes unmanageable. To overcome this issue, a margin of tolerance (epsilon) is set in approximation to the SVM to minimize error, individualizing the hyperplane to maximize margin. Support Vector Regression (SVR) works on similar principle of SVM but in it, predictor variables and observed response values are included in set of training data.

   Support Vector Machine uses mainly four kernel functions:

   a) Linear function: A scalar response/ output (or dependent variable) have linear relationship with one or more explanatory/ input variables (or independent variables)

   b) Polynomial function: It fits a nonlinear relationship between y (dependent variable) and nth degree polynomial of x (independent variable). Least squares method are used for polynomial regression models by minimizing the variance.

   c) Radial Basis Function (RBF) /Gaussian function: It is a real-valued function which depends on radial distance from origin/ center eg. Euclidean distance etc. Sums of RBFs are used to approximate given functions and this approximation process of given functions may be interpreted as a simple neural network. In support vector classification, these radial basis functions are also used as a kernel.

   d) Sigmoid function: It is commonly used due to its low calculation time.

4. **Random Forest:** It is used for high prediction accuracy. Clustering with this method saves efforts in feature pre-processing. There may be multiple random decision trees in Random forest which can be of two types:

   a) Each decision tree built on a random sample data from original data.

   b) Decision tree at each tree node.

   A sub-set features is selected in a random manner for generating most beneficial split. For reduction in variance of an prediction function estimation or bootstrap aggregation is used as a selection parameter in these decision tree. Random forest method averages multiple deep decision trees trained on different subsets of the same training dataset with the goal of minimizing variance. It works well for low-bias procedures and high-variance, such as trees.

5. **k-nearest neighbors (kNN) algorithms:** It is also a non-parametric method like SVM & can be applied for both analysis of regression and classification. The k closest training datasets are in the input of the in the feature space and the output reckons on whether this algorithm is being used for classification (output is class membership) or for regression (output is average property value of k nearest neighbors for the object). In it, weights are assigned for the contributions of neighbors, eg. giving each neighbor a weight of 1/d, where d is the distance from its neighbor. This algorithm is commonly used for its good predictive power, low calculation time and ease of interpretation.

RESULT AND DISCUSSION

In this research paper, a model is developed on the basis of applying five different machine learning methods on an open source database available at UCI repository. Before applying these machine learning methods for diagnosing heart disease, all the available 920 records collected from four different locations with 13 input parameters and one output parameters, were merged, cleaned and then it is reshuffled many times for having a better prediction. The dataset is split in 80:20 ratio, i.e. 80% for training the model and remaining 20% for testing this model.

For better prediction of results, it is to be ensured that datasets on which these machine learning methods are being applied are balanced i.e. number of samples for each target values are almost equal. In the dataset which is used for prediction of heart disease, two target values were allotted in output variable i.e. 0 for those not suffering from heart disease, and 1 for those suffering from heart disease.
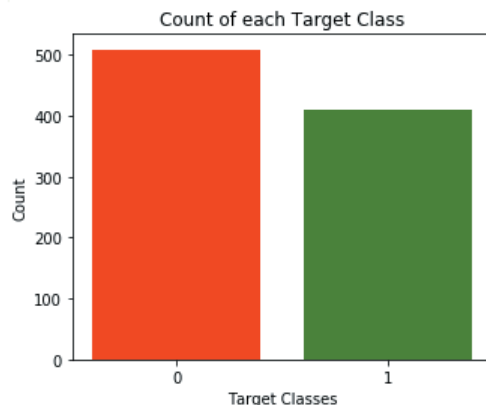

Figure 2: Count of each Target class

As shown in above graph (Figure 2), out of available records of 920 persons, 500 persons are not suffering from heart disease and 420 are suffering from heart disease. This data can be considered balanced and hence, machine learning methods can be applied.

As prediction of heart disease is a classification problem, its performance can be described by a table called confusion matrix (Figure 3).

Figure 3: Confusion matrix

For prediction of accuracy, following parameters can be calculated:

Accuracy = [(tp + tn) / (tp + fp + tn + fn)] x 100

Precision = tp / (tp + fp)

Recall = tp / (tp + fn)

where   tp - True Positive,   tn - True Negative,

fp - False Positive,   fn - False Negative

In this paper, confusion matrix is created on the basis of predictions from all the five machine learning methods

which are applied on the heart disease dataset after pre-processing and checking whether it is balanced. Confusion matrix (Figure 4, 6, 8, 10, 12) generated from these machine learning methods and the parameters calculated are shown below:
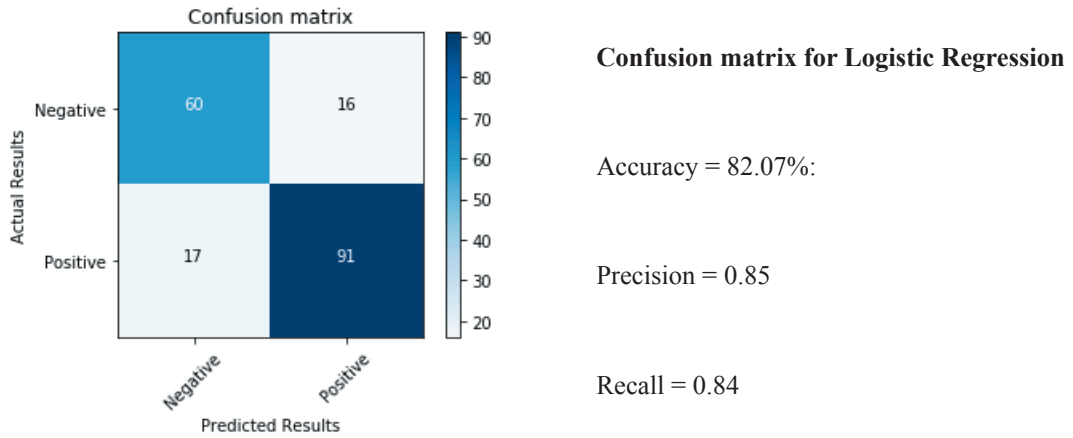

Confusion matrix for Logistic Regression

Accuracy = 82.07%:

Precision = 0.85

Recall = 0.84

Figure 4. Confusion matrix for Logistic Regression

In this database, there are total 30 features for 13 input parameters. Each input parameters have two or more different values, and each such different value is considered as a feature (eg. male, female are tow features in sex parameter). A graph is plotted between number of parameters and score (accuracy) and it is observed that the maximum score (accuracy) is obtained for 23 features while applying decision tree machine learning method.
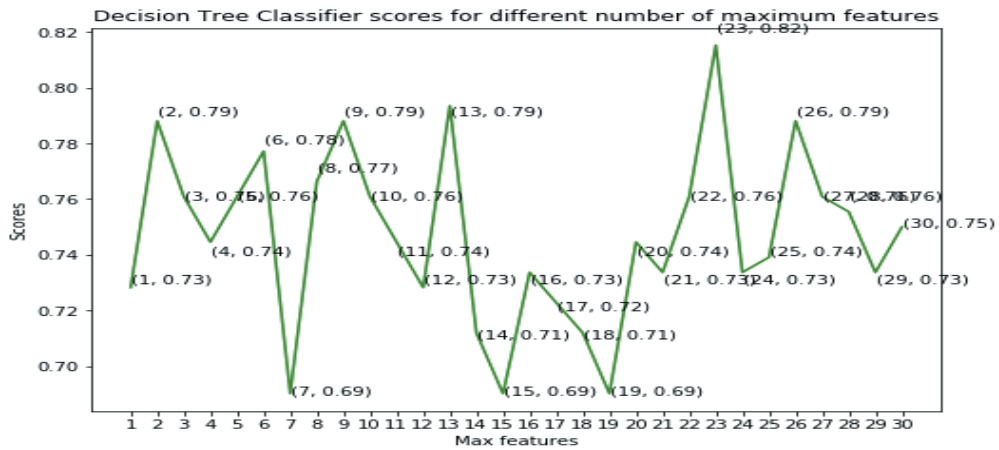
Figure 5: Decision tree classifier scores

Confusion matrix for Decision Tree

Accuracy = 81.52%:

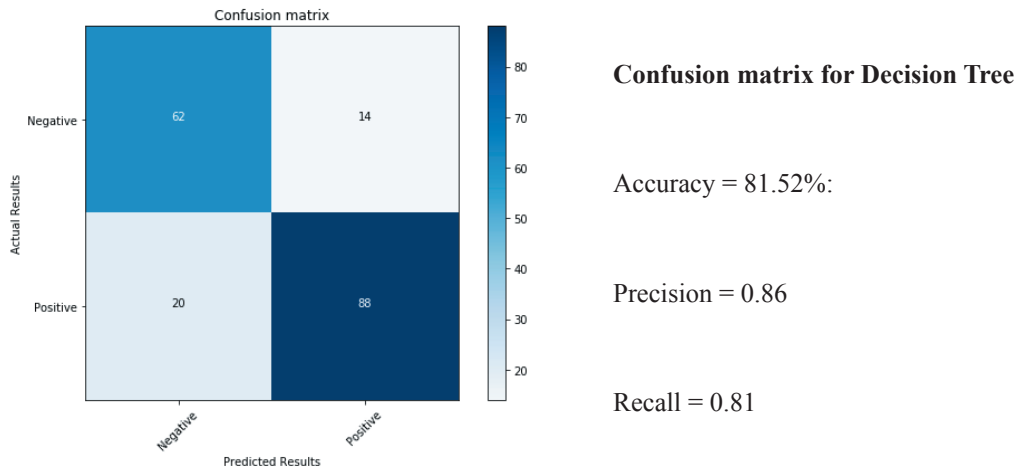Precision = 0.86

Recall = 0.81

Figure 6. Confusion matrix for Decision Tree

For Support vector machine, database is analyzed on the basis of four different function i.e. linear, polynomial, RBF, and gaussian function. The accuracy with all these function is calculated and compared. A histogram is prepared for all these function and a confusion matrix is created for best accuracy.
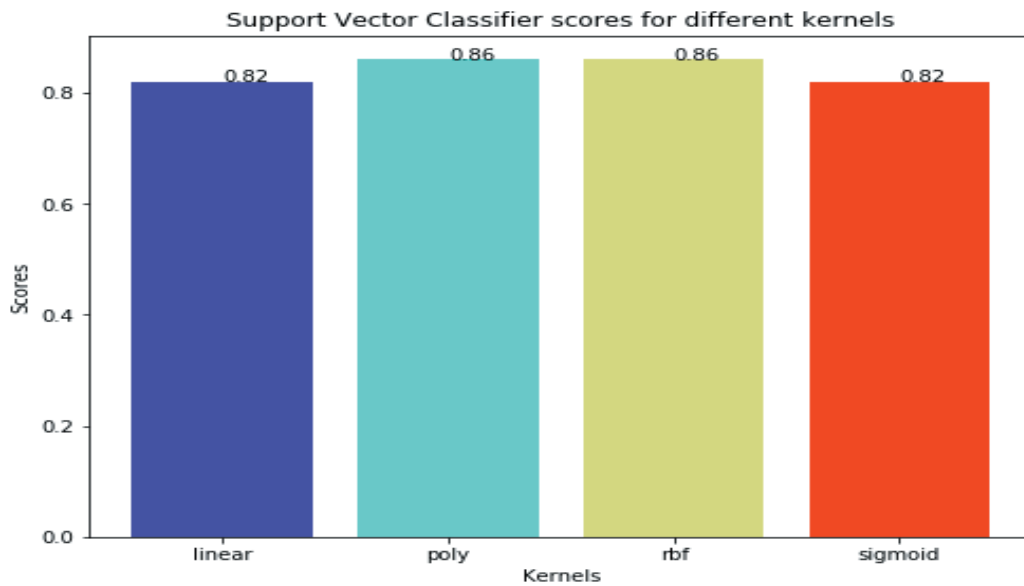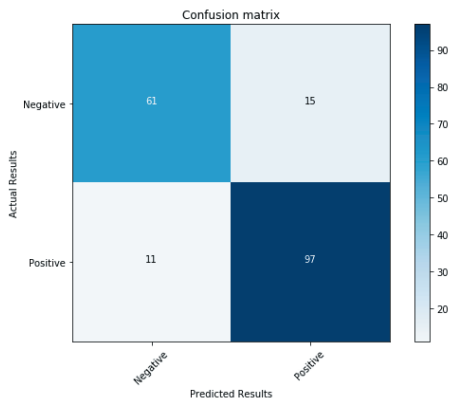

Figure 7. Support Vector classifier scores

Confusion matrix for SVM

Accuracy = 85.87%:

Precision = 0.87

Recall = 0.90

Figure 8: Confusion matrix for SVM

For random forest method, a histogram is created for different no of tree (in a forest) and corresponding score (accuracy). On the basis of best accuracy, a confusion matrix is created.
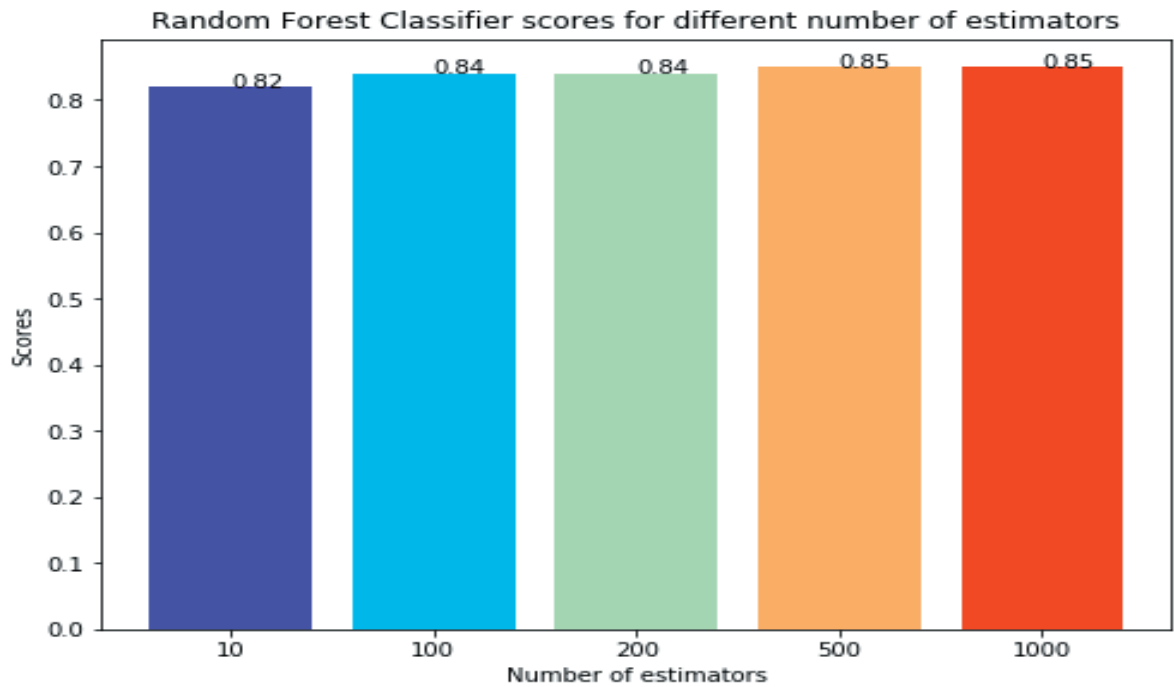

Figure 9. Random Forest classifier scores

Confusion matrix for Random Forest

Accuracy = 85.33%:
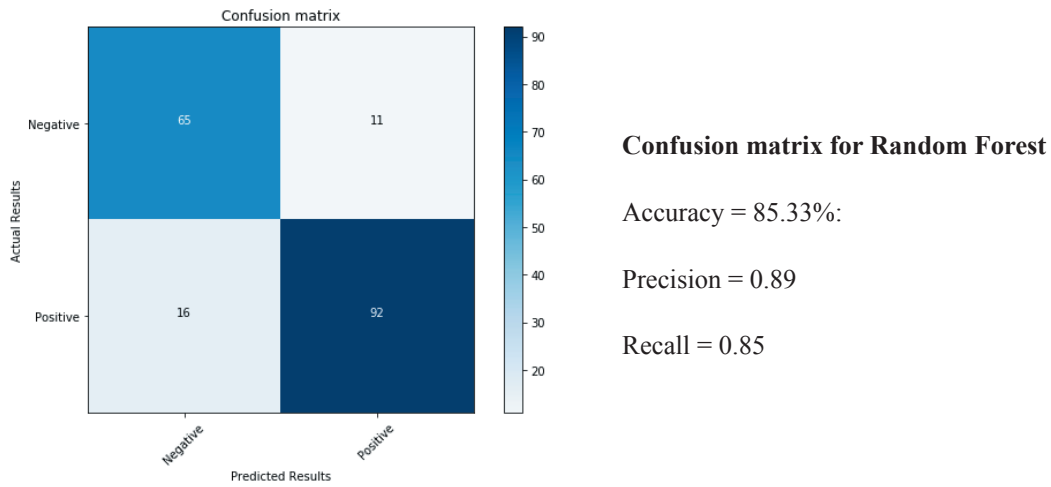
Precision = 0.89

Recall = 0.85

Figure 10: Confusion matrix for Random Forest

For k-NN (k nearest neighbors) machine learning method, a graph is plotted for different values of k i.e. different number of neighbors and it is observed that the maximum score (accuracy) is obtained for k=9 and on the basis of best accuracy, a confusion matrix is created.
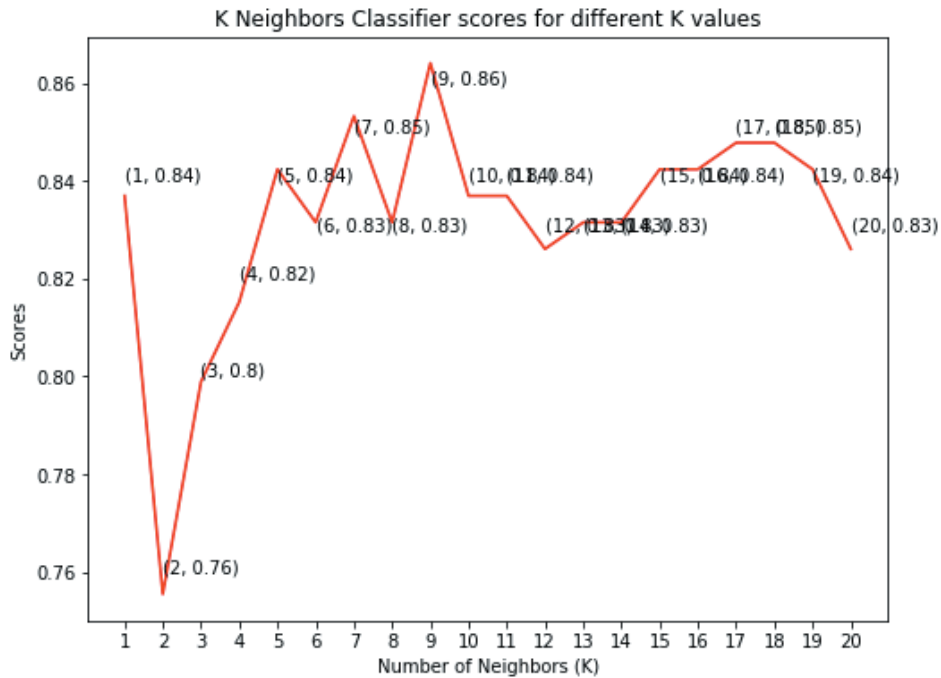

Figure 11. k- neighbors classifier scores for different k values

Confusion matrix for k-NN (k nearest neighbors)

Accuracy = 86.41 %:

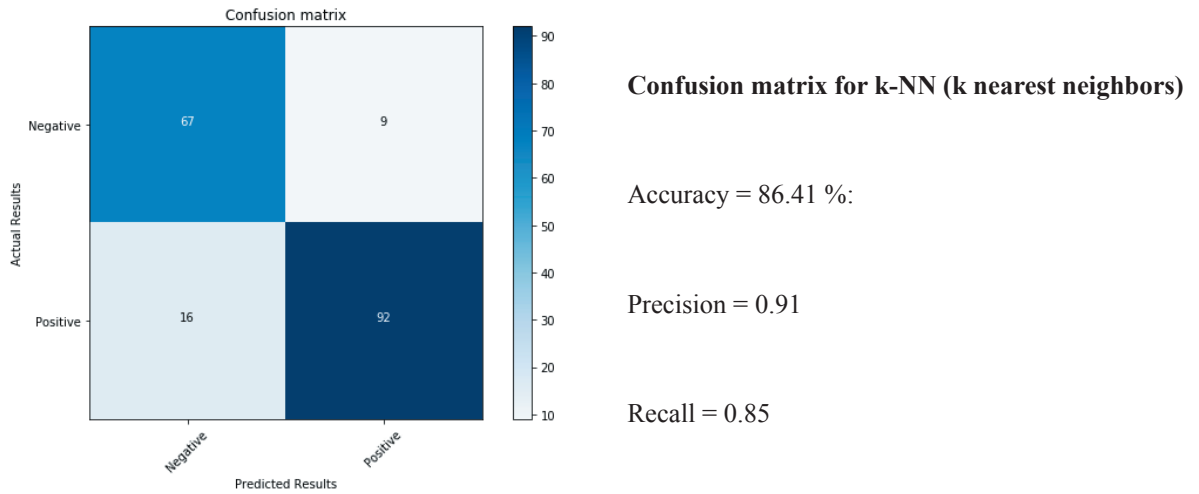Precision = 0.91

Recall = 0.85

Figure 12. Confusion matrix for k-NN

A histogram is drawn for maximum score (accuracy) for all these five machine learning methods and on the basis of the accuracy, it is found that k-NN machine learning method is best for predict heart disease, which predicts 86.41 % accurate results. This k_NN also proves itself better than others with 91 % precision value.
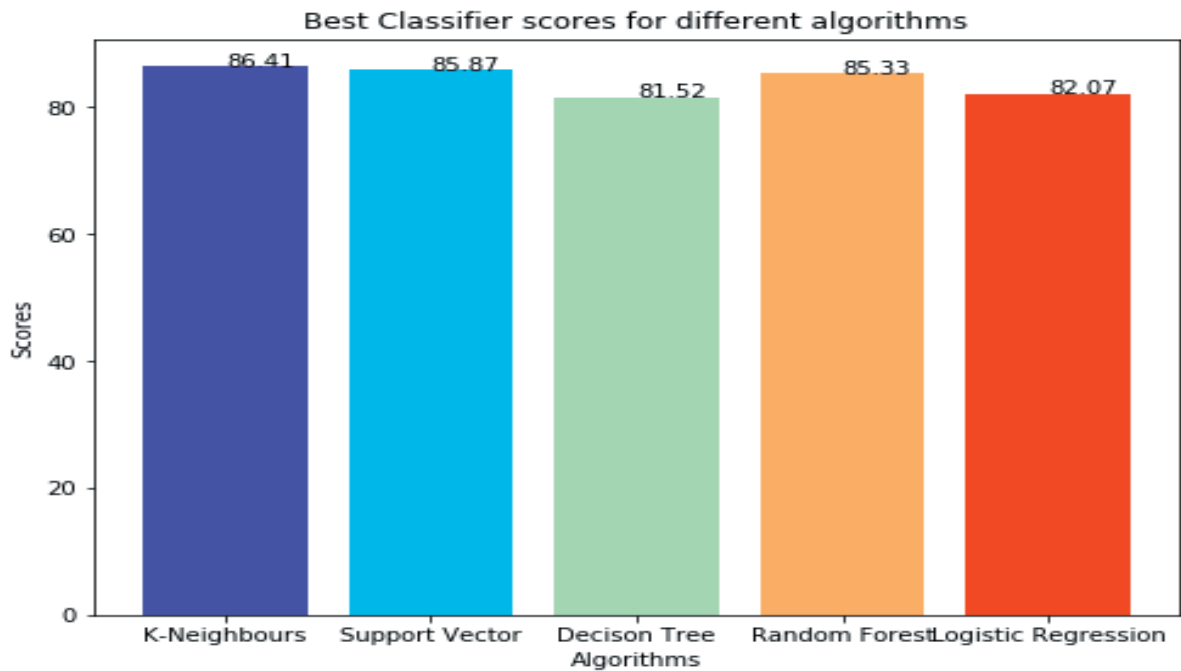

Figure 3: Best classifier scores

The result may vary if ratio of train & test dataset is varied, however, the ratio of 80:20 provides good results and can be relied on. The result may also vary if SVM machine learning algorithm proves itself better, if results are compared on the bases of recall value.

CONCLUSION

In this paper, an open source database available at UCI repository with 920 records collected from four different locations and having 13 input parameters and one output parameters was used for developing various prediction models for coronary heart disease using machine learning techniques. Five different machine learning algorithms with multiple parameters were applied to develop multiple prediction models for each algorithm. A comprehensive comparative study was carried out to compare the results of these prediction models within each of the 5 ML algorithms and then across the 5 ML algorithms to determine the best prediction model for the problem. A comparative study tool in python programming language was used for developing prediction models and comparing their prediction accuracy to determine the best model. It is observed that k-NN method is a better prediction algorithm with better accuracy and precision value.

FUTURE WORK

1. This comparative tool can be adapted for the purpose of multi-classification problem of diagnosing severity of heart disease for a patient on the basis of the data available.

2. Machine Learning methods using Artificial Neural Network and/ or Fuzzy Logic can be used for this classification problem of diagnosing heart disease and its level (severity of disease) for a patient on the basis of the data available.

3. This comparative tool based on five different machine learning algorithms can also be used on a coronary heart disease database or any other medical database related to diagnosis of disease collected from Indian origin.

4. A window based graphic user interface (GUI) can be developed for predicting the disease on the basis of available input parameters and with capability of self-improvement in the background i.e. updating database, training the model and improving its performance from the data entered for prediction of disease.

5. This comparative tool can also be adapted for classification purpose for a database collected from medical records, financial records, social media posts, emails, digital media files, mobile phones, and records collected from engineering systems such as factories through sensors, aircraft and road networks etc.

6. This comparative tool can also be adapted in the mechanical engineering field to identify patterns, for making decisions, drawing conclusions related to material, alloys and their properties, machine tools and their life, fluids and its properties, predictive maintenance, Supply chain, ergonomics etc.

REFERENCES

Rani, K., Chaitanya, M., & Kiran, G. (2018). A Heart Disease Prediction Model using Logistic Regression By Cleveland DataBase. International Journal Of Trend In Scientific Research And Development, Volume-2(Issue-3), 1467-1470. doi: 10.31142/ijtsrd11402

Kumar, S., & Sahoo, G. (2018). Enhanced decision tree algorithm using genetic algorithm for heart disease prediction. International Journal Of Bioinformatics Research And Applications, 14(1/2), 49. doi: 10.1504/ijbra.2018.089199

Celik, S., & Yilmaz, O. (2018). Prediction of Body Weight of Turkish Tazi Dogs using Data Mining Techniques: Classification and Regression Tree (CART) and Multivariate Adaptive Regression Splines (MARS). Pakistan Journal Of Zoology, 50(2). doi: 10.17582/journal.pjz/2018.50.2.575.583

Narayan, S., & Gobal, J. (2018). Optimal Decision Tree Fuzzy Rule Based Classifier (ODT-FRC) For Heart Disease Prediction Using Improved Cuckoo Search Algorithm. International Journal Of Business Intelligence And Data Mining, 1(1), 1. doi: 10.1504/ijbidm.2018.10008934

Maheswari, S., & Pitchai, R. (2018). Heart Disease Prediction System Using Decision Tree And Naive Bayes Algorithm. Current Medical Imaging Reviews, 14. doi: 10.2174/1573405614666180322141259

Prediction of Heart Disease using Decision Tree and Neural Network Data Mining Techniques-A Review. (2018). International Journal Of Advance Engineering And Research Development, 5(02). doi: 10.21090/ijaerd. etiece024

Alsina, E., Chica, M., Trawiński, K., & Regattieri, A. (2017). On the use of machine learning methods to predict component reliability from data-driven industrial case studies. The International Journal Of Advanced Manufacturing Technology, 94(5-8), 2419-2433. doi: 10.1007/s00170-017-1039-x

Kim, J., Lee, J., & Lee, Y. (2015). Data-Mining-Based Coronary Heart Disease Risk Prediction Model Using Fuzzy Logic and Decision Tree. Healthcare Informatics Research, 21(3), 167. doi: 10.4258/hir.2015.21.3.167

Akila, S., & Chandramathi, S. (2015). A Hybrid Method for Coronary Heart Disease Risk Prediction using Decision Tree and Multi Layer Perceptron. Indian Journal Of Science And Technology, 8(34). doi: 10.17485/ijst/2015/ v8i34/85947

Ramakrishnan, S., Mirzaei, M., & Bekri, M. (2015). A Multi-industry Default Prediction Model using Logistic Regression and Decision Tree. Research Journal Of Applied Sciences, Engineering And Technology, 9(10), 856-861. doi: 10.19026/rjaset.9.2635

Al-batah, M. (2014). Testing the Probability of Heart Disease Using Classification and Regression Tree Model. Annual Research & Review In Biology, 4(11), 1713-1725. doi: 10.9734/arrb/2014/7786

Mythili T., Mukherji, D., Padalia, N., & Naidu, A. (2013). A Heart Disease Prediction Model using SVM-Decision Trees-Logistic Regression (SDL). International Journal Of Computer Applications, 68(16), 11-15. doi: 10.5120/11662-7250

Dhanashree S. Medhekar, Mayur P. Bote, Shruti D. Deshmukh (2013). Heart Disease Prediction System using Naive Bayes International Journal Of Enhanced Research In Science Technology & Engineering Vol. 2 Issue 3, March.-2013 ISSN NO: 2319-7463

Ishtake SH, Sanap SA. (2013). Intelligent Heart Disease Prediction System Using Data Mining Techniques. International J. of Healthcare & Biomedical Research. 1(3):94-101.

Anooj, P. (2011). Clinical decision support system: risk level prediction of heart disease using weighted fuzzy rules and decision tree rules. Open Computer Science, 1(4). doi: 10.2478/s13537-011-0032-y

Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome (٢٠٠9). The Elements of Statistical Learning (٢nd ed.). Springer. ISBN 0-387-95284-5.

Ozsen S, Gunes S. (2009). Attribute weighting via genetic algorithms for attribute weighted artificial immune system (AWAIS) and its application to heart disease and liver disorders problems. Expert Systems with Applications. 36(1):386-392

Polat K, Sahan S, Gunes S. (2007). Automatic detection of heart disease using an artificial immune recognition system (AIRS) with fuzzy resource allocation mechanism and k- nn (nearest neighbour) based weighting preprocessing. Expert Systems with Applications. 32(2):625-631.