Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume12, Issue 5, July 2021: 3374-3400

Research Article

An Improved Ensemble Approach to Predict Cardiovascular Problems

[1]Dr. Mujtaba Ashraf Qureshi*, [2]Dr. Mohd Iqbal Sheikh
[1]Assistant Professor (C), Department-Information Technology, Cluster University, Srinagar, J&K.
Email:mujtaba170@gmail.com
[2]Assistant Professor (C), Department-Computer Science, Cluster University, Srinagar, J&K.
Email:iqbalsheikh915@gmail.com
(*Corresponding author's email: mujtaba170@gmail.com)

Abstract: Cardiovascular problems have fully dominated the world of humankind. Millions of people in the world die every year because of diverse existing cardiovascular problems. Data mining techniques plays imperative role to curb the death rate because diverse cardiac prediction models are devised using different mining techniques/algorithms. In this research work five models are devised to attain the performance measures related to the heart diseases prediction problems which include decision tree, J48, neural network, naïve Bayes and support vector machine. Among these employed classifiers J48 and neural network (MLP) depicts enhanced performance measures in comparison to other employed techniques. This experimented approach employed ensemble methodology to combine the predictive power of the two acceptable techniques i.e. J48 and neural networks (MLP) using WEKA simulation tool. Ensemble algorithms are tallied among the topmost machine learning techniques to combine the prediction power from desired multiple techniques. Ensemble includes several techniques but we engaged bagging method of ensemble to devise the prediction models i.e. model 1 and model 2 using J48 and neural networks. Both the developed models are satisfactory to employ for the prediction of cardiovascular problems however model 1 conquers to model 2 for the prediction of cardiovascular problems.

Keywords: *Cardiovascular problems, Data Mining, Ensemble Approach, J48, Neural Network*

1. Introduction

Data mining is defined as the extraction of useful and suitable patterns of data from the huge and diverse databases by the applications of diverse existing mining methods. The well-known definition of data mining is presented as Data mining is the non-trivial deduction of hidden previously unknown and possibly useful information about data (Frawley, 1996). To attain the concealed outlines of data, user oriented outline is made available by data mining process (Ralf Mikut, 2011). Medical databases contain huge, unstructured and distributed datasets. The manageable and existing unprocessed data needs to assemble and stock in an organized format (Wilson, 1998). To convert the raw data into useful and beneficial information is called as Knowledge discovery from database (KDD), which is the process to transform raw data into productive information. In the present world application of data mining techniques is visible and perceptible everywhere. Data mining process has revolutionized almost every field like education, banking, business, scientific methods, astronomy, census or healthcare. It helps these fields to detect the problems or difficulties and predict whether future outcomes would be beneficial or not for the business activities undertaken. So in this way counteractive and active measures are taken to resolve before they would create problems in future.

Medical industry is approaching at very fast pace towards the winning and preferred zone because of data mining methodology. It has completely changed the face of medical industry from traditional methods to modern technology. Data mining had also had made impossible things to move to the sphere of possibility. There is yet vast scope of data mining techniques as there survives huge and bulky medical databases which can be used to derive conclusions and decisions for the beneficial of humankind. The gathering of data about different diseases in today's medical science is very significant [Peyman Mohammadi et al. 2013].

According to Salim Diwani et al. (2013) data mining applications can be established to gauge the efficacy of medical treatments. This could also profit healthcare providers such as hospitals, physicians, and patients by categorizing active treatments and best performs.

Various applications of data mining to the field of medical science are presented below;

- To predict diverse problems or diseases,
- To perceive and early control of different diseases,
- To predict future consequences on the basis of available and collected data,

Heart problems are measured single chief source of death in developed countries and the main donors to disease distress in developing countries. In 2008 approximately 17 million people died all over the world due to the heart diseases. According to the estimation of WHO by 2030, approximately 24 million people will die because of heart diseases (Miss. Chaitrali S. Dangare, 2012). The figure 1 shows the influence of deaths happened due the cardiovascular diseases;

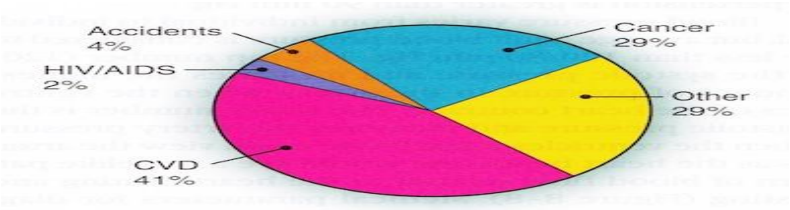

Figure 1: Mortality Rate due to CVD's

In this research paper role of data mining in medical science and some principal concerns confronted in diagnosis and preclusion of cardiac problems are emphasized. Also a proposed explanation is presented to overcome the problems connected to analysis and preclusion of cardiac problems using ensemble machine learning algorithms. To conclude, two classifiers namely J48 and neural networks (MLP) are used to predict the cardiac problems using ensemble methodology as both these techniques shows an enhanced performance measures over other employed techniques.

2. Literature Survey

A brief survey of literature is accessible in this section correlated to heart diseases prediction using dissimilar data mining techniques;

The model, Intelligent Heart Disease Prediction System (IHDPS) formed by means of naïve bayes, decision tree and neural networks techniques by Sellappan Palaniappan et al (Sellappan Palaniappan, 2008).Complex queries were easily solved by developed IHDPS in comparison to other conventional system. Main features of this system are use-friendly, scalable, expandable, web-based etc.

The method called Neuro-fuzzy is estimated to predict cardiovascular diseases. This method presents very low error rate with an acceptable efficiency for analysis and occurrences of CVDs (A. K. Sen, 2013).

Four data mining techniques RIPPER classifier, Decision Tree, Support Vector Machine and Artificial Neural Network were employed to analyze cardiovascular problems (Godara, 2011). Ten-fold cross validation method is employed to evaluate and estimate the accuracies of the employed techniques. SVM overpowers to other employed techniques for the prediction of CVDs.

Cardiovascular Prediction model formation passes via various complex methods of data mining. Hidden patterns of data are obtained by user oriented approaches from data (Ralf Mikut, 2011).

Heart diseases prediction model is devised using neural networks, decision trees, naïve bayes and support vector machine to present acceptable results. The techniques are intelligent and efficient in nature and make them more applicable to medical science (S. Kiruthika Devi, 2016).

Two selected additional parameters are i.e. smoking and obesity are supplied in corporation with other thirteen selected parameters like sex, cholesterol to predict hear diseases. Here NB, DT, and NN were employed in WEKA 3.6.6 simulation tool. This work presents neural networks an acceptable technique over other techniques to predict heart diseases (Chaitrali S. Dangare, 2012).

In (Nidhi Bhatla, 2012, .) fuzzy logic using novel approach of data mining techniques to perform work. To decrease the number of tests to be done by patients, only four parameters are employed and the model developed presents an enhanced efficiency and accuracy. Here also 100% accuracy is achieved while using NBs and DTs along with four parameters only. A conclusion is drawn that NBs and DTs presents better results than others.

In (Moloud Abdar, 2015), thirteen selected parameters are used for to compare various techniques to predict diseases. Five techniques are employed namely NN, C5.0, SVM, Logistic Regression and KNN to devise the developed models. The results shown by K-Nearest Neighborhood, C5.0, NNs, SVM, is 88.37%, 93.02%, 80.23%, 86.05%, likewise. Here Decision tree presents acceptable results over other employed techniques for the prediction of heart diseases.

In (Hlaudi Daniel Masethe, 22-24 October,,2014) , a simulated to find out whether a person is suffering from cardiac problems or not by applying the selected eleven parameters is devised. The subsequent techniques given are J48, NBs, REPTREE, CART, and Bayes Net engaged for model enlargement and benevolences exactness 99.0741%, 97.222%, 99.0741%, 99.0741% and 98.148% respectively. The prominent exactness revealed by REPTREE, J48, and SIMPLE CART algorithms suggests that features used are reliable indicators to calculation the chances of diseases of heart in well-defined manner.

In (Resul Das, (2009)), a technique is accessible using 13 features in SAS simulation tool for analysis of CVDs. The SAS base simulation platform is an intellectual and competent platform to evaluate the performance results of the established systems in the various viewpoints. A collaborative model, using three self-regulating NN model, is established. Model is offered by using the features of three autonomous NN models. To enhance the performance results of the ensemble model, the application of nodes is amplified, but no development was attained. The experimental examination shows 95.91% specificity, 89.01% accuracy and 80.95% sensitivity values for calculation.

In (D'Souza, 2015) a model for CVDs using NN, K-means clustering and FIS generation techniques. Investigators used diverse number of features to accomplish the efforts. Here researcher employed 14 parameters of patients such as blood pressure, age etc. for this assignment. These nominated features are supplied to the established model to recognize the calculation of numerous cardiac problems.

In (Ansari) accomplished a work, heart prediction system is developed using neuro-fuzzy unified method. To scrutinize heart diseases, author established neuro-fuzzy combined arrangement. Work is accomplished using seven medical features by means of MATLAB simulation platform. The chief characteristics of this unified system are to envisage the heart problems with low or high strength.

In (Carlos Ordonez Teradata, 2006), an algorithm is offered to diminish the rule number, to inspect for association on the training sets and validate them by means of the test data. To appraise the implication in the medical industry revealed rule is used with support, lifts and assurance. An examination and investigational process is accomplished by means of the real data sets, by means of association rules and verified a satisfactory technique to forecast heart diseases in a efficient way.

In (Boshra Bahrami, 2015.) WEKA replication platform is employed to execute real-world project of this work. Investigator used four techniques specified consequently naïve bayes, SMO, KNN and j48 decision tree to inspect and match up. J48 offers advanced outcome as related to other employed techniques.

3. Framework

The framework employed for this experimental approach to be carried out is presented, shown in figure 2. An exploratory study of the output results revealed by employed techniques for the prediction of cardiovascular problems is accomplished.

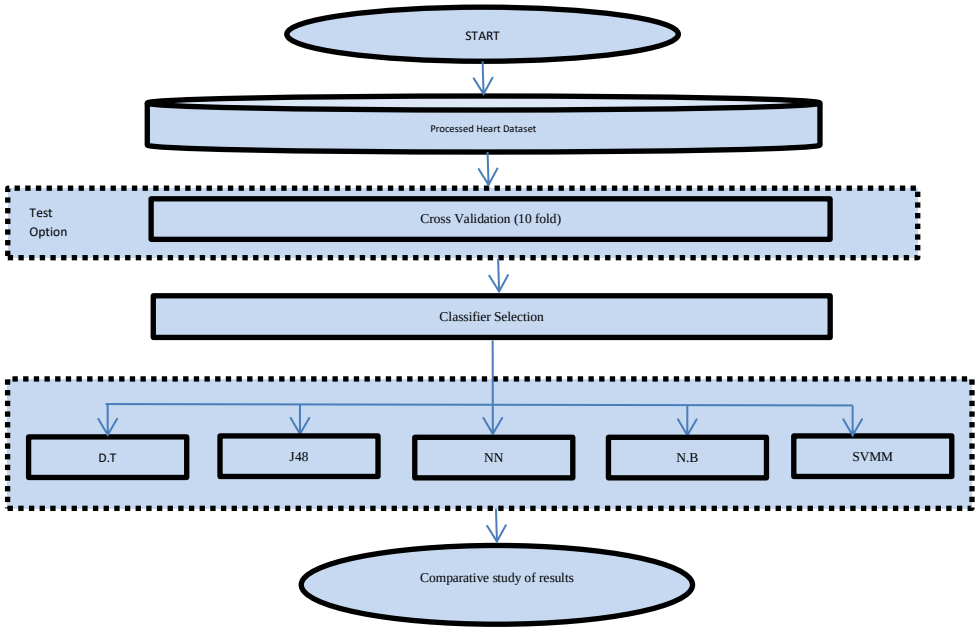

Figure 2: Framework

A comparative study of the results shown by employed techniques is executed, in which J48 and neural network overpowers to other techniques. So J48 and neural networks are employed to devise heart disease prediction model/models using ensemble approach.

4. Role of Data Mining in Cardiac Science

Data mining technology has played substantial and most imperative role to perceive and preclude the heart diseases well on time. Mining technology has taken upheaval in the cardiology field of the healthcare. There is urgent need to restrict heart problems at the present stage as this disease is considered one of the dreadful diseases in the world.
Dissimilar models are developed and drawn out for the diagnosis of CVDs by the submission of mining technology. An Intelligent Heart Disease Prediction System (IHDPS) model molded by means of NBs, Decision trees and NN mining techniques by (Sellappan Palaniappan, 2008). Multifaceted requests were given to this IHDPS and results are obtained in an intellectual and effective way which other conventional systems cannot do.
Neuro-fuzzy technique is predictable to estimate the CVDs. This method displays low error rate and high competence for investigation and incidences of heart diseases (A. K. Sen, 2013).
Heart disease forecast model expansion approaches pass through various multifaceted stages of data mining technology. Data mining offers a user oriented method to find the unseen designs in the data (Ralf Mikut, 2011). In

brief, it is a procedure of investigating data from different viewpoint and meeting the information from it. In healthcare industry the accessible unprepared medicinal data is diverse, gigantic and dispersed over various sites. The process of transforming the data into valuable evidence is known as Knowledge discovery from database (KDD), which is the process to change raw data into beneficial evidence. The various steps of KDD are shown as selection, preprocessing and transformation of data. Figure 3 shows how conclusions are made by means of the datasets of patients.

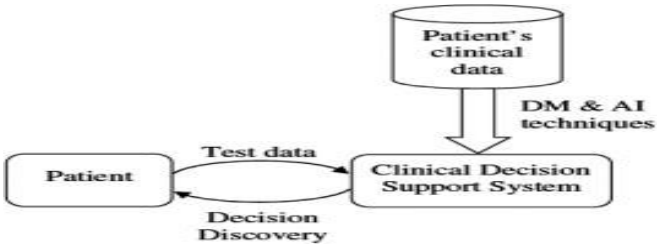

Figure 3: Decision Making Steps.

Role of data mining technology for the heart diseases diagnosis is summed as below;

i. Diverse heart disease prediction models are developed with varied levels of efficiency and accuracy.
ii. Preventive measures are taken immediately when prediction is made well on time.
iii. Undiscovered and distributed data is made available to analyze and draw conclusions related to solve heart related problems.
iv. From the sea of hidden data, useful patterns are obtained.
v. Prediction of diseases is done well on time with the help of developed models.
vi. Patient not required to go for numerous medical tests and thus saves both the cost and time of the patient.
vii. Detection and diagnosis of heart diseases prevalence is determined on the basis of various factors like geographical, dietary/food habits, work, age, family background and so on.
viii. Work plan and other surgical plans are made very successful for the heart patients using data miming methodology.

5. Methodology

5.1. Dataset

In this experimental work, WEKA 3.9 simulation tool is employed for performance analysis of the employed techniques for cardiac problems. WEKA is basically an open source platform employed for preprocessing methods, to implement several machine learning algorithms and visualization methods to develop different models to solve real world problems related to data mining. Moreover the potentials like easy to employ, graphical features, well defined interface and high model building capability enhances the success of WEKA tool further Raw data is always accompanied by irrelevant, noisy and redundant data but preprocessing methods in WEKA handles that data efficiently. WEKA delivers dissimilar algorithms like classification, clustering and association and allows their applications based on the requirement of the assignment.

The effective output of the algorithms is very much inflated by the nature and rationality of employed datasets. The dataset engaged to accomplish this investigational method is acquired from the premises of JLNM Hospital *(Srinagar, Kashmir, India) to perform experimental work*. Heart linked datasets are available in adequate expanse. Preprocessing methods and feature selection techniques accessible in WEKA replication tool are integrated to achieve the high profile attributes only to train and testify the established models. The selection of only 12 high

profile attributes is arranged which consists of 621 instances only. A valuation of the nominated attributes is presented below in table 1.

Table 1: Attributes used to predict Heart Diseases.

S.NO	*Attributes*	*Description*	*Type*
1	*age*	*In years*	*Numeric*
2	*trestbp*	*Resting blood pressure*	*Numeric*
3	*Sex*	*Male or female*	*Nominal*
4	*Cp*	*Chest pain type*	*Nominal*
5	*Chol*	*Cholesterol level in mg/dl*	*Numeric*
6	*Fbs*	*Fasting blood sugar*	*Nominal*
7	*Restecg*	*Electrographic results at rest.*	*Nominal*
8	*Thalach*	*Maximum Heart Rate Achieved.*	*Numeric*
9	*Exang*	*Exercise Induced Angina*	*Nominal*
10	*Oldpeak*	*ST Depression Induced By Exercise.*	*Numeric*
11	*Slope*	*Slope of the Peak Exercise ST Segment.*	*Nominal*
12	*ThalBlood*	*Type of Defect of blood/Heart.*	*Nominal*
13	*Diagnosis*	*Absence or presence of disease.*	*?*

5.2. Data Preprocessing and Feature Selection

The collected data sets were full of redundant and noisy data and consist of 706 instances. However after applying various preprocessing methods available in the WEKA we obtained 601 processed instances with only 12 significant attributes mentioned in table 1. Binary classification system is used to classify the data of patients whether a person falls in yes class (heart patient) or no class (healthy person). After preprocessing we obtained 278 instances as positive (1) and 343 as negative (0) class. Positive class refers to those people who are involved in heart diseases and the value is set to 1 (yes). Similarly negative class refers to that category of people who are not involved in heart problems and the value is set to 0 (no). Here we employed four machine learning algorithms to develop heart diseases prediction models based on the application of the processed dataset.

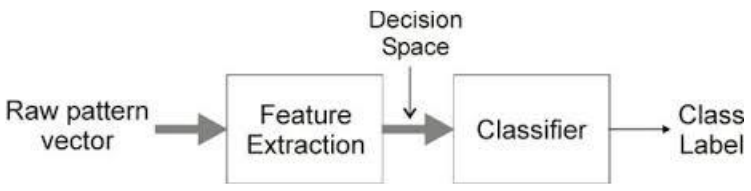

Figure 4: Feature extraction.

6. Classification Modeling

In WEKA we employed four machine learning algorithms - Decision Tree, J48, Neural Network (MLP), Naïve Bayes and Support vector Machine (SVM) to develop heart diseases prediction models based on the submission of the processed dataset. Cross-validation method is used as resampling practice to devise prediction models based on the processed but limited data. In CV methodology actually dataset is divided into several folds to train and test the model and finally presents output as an average of the result. More recommended number of folds for a dataset is 10, i.e. k=10, denotes dataset is divided into ten folds and among which 9 folds provides training to the model and one fold acts as test fold to the said model. This procedure continues until all the folds will not take part in testing at least once. Finally output is obtained as an average of the results of the 10 folds.

To illustrate efficiency and capability of the developed models/techniques or the process of collecting, analyzing and recording the act of the system or model is called performance measurement. There are diverse metrics chosen to evaluate and estimate the capabilities of the devised models. In this research work following performance measures is taken into consideration to analyze and evaluate the prediction models.

Confusion Matrix: the confusion matrix is one of the most acceptable and easiest metrics used for problems related to classification. Confusion matrix performs well whether output is binary or multiclass system. Different performance measures are defined and explained based on the confusion matrix only. Thus it is good to say that confusion matrix is a complex structure and in real logic confusion matrix itself is not any measurement. Figure… depicts the overall structure and definition of confusion matrix as;

Several significant measurements are taken into consideration to indicate the importance and derive the performance measures of the devised models as;

1. Accuracy is demarcated as the how well our model/classifier is performing to divide tuples into corresponding classes.

$$Accuracy = (TP+TN)/ (TP+TN+FP+FN)$$

2. Precision refers to the dimension of correctness or percentage of tuples labeled as positive or negative are actually as such by the developed classifiers.

$$Precision = (TP)/ (TP+FP)$$

3. Sensitivity states to positive tuples properly labeled by the classifier or to recognize suitably those entities actually related to diseases. Sensitivity is also referred as true positive rate.

$$Sensitivity\ (Recall) = (TP)/ (TP+FN)$$

4. Specificity is defined as the number of negative tuples correctly recognized.

$$Specificity= (TN)/ (TN + FP)$$

5. Elapsed Time of Training is the time taken.

Here various data mining techniques are employed to compare and present the suitable techniques for the prediction of heart diseases, which are presented as below;

Decision Tree: Decision tree is one of the most significant data mining technique used for the analysis, comparison of datasets and finally to develop predictive models for the given variables in the cardiovascular predictive system. CART, ID3, C4.5, CHAID, and J48 are the various prominent decision tree algorithms. This technique forms an

inverted tree like structure to the given data. This tree forms various structures known as internal nodes, root node and leaf nodes.

J48: J48 (C4.5) is employed to generate decision tree. This is extension of ID3 algorithm. This is considered suitable for classification problems so often are called as statistical classifier.

Neural Network (MLP): An artificial neural network is a mathematical model or computational model also called as neural network. This network is based on biological neural networks. Artificial neural network is based on observation of a human brain [51].Brain of human beings consists of complex network of neurons. Figure 3.2 shows the basic structure of biological neuron.

Naïve Bayes: Naïve Bayes works on the conditional probability means that something will happen, provided that something has occurred in the past. Most appropriate and most possible output is possible to obtain using input data by using the Bayesian theorem. One of the realistic and weighty features of the Bayes theorem is to add the new unprocessed data at runtime. This classifier has also better probabilistic nature.

Support Vector Machine (SMO): Support Vector Machine is supervised machine learning technique. In data mining process SVM is used for both regressions as well as for classification techniques. "SVM recently introduced to progress in methods for prediction of diseases [21, 22]." Each data item in the SVM is represented as a point in n-dimensional space. These items are separated into two classes by plane called hyper plane.

7. Experimental Approach

The experimental approach is divided into two sub-sections as 7.1 and 7.2. In the section 7.1, a comparative analysis of the results revealed by employed techniques, is performed. An analytical study shows that shows that neural networks and J48 models subjugated to other predictive techniques in performance measures. In section 7.2, an ensemble method by means of J48 and neural network algorithms is employed to develop heart diseases prediction models to augment heart disease prediction results further.

7.1: Experimental Setup:

This section presents experimental results, as shown in table 2, of the employed techniques to verify the supremacy and legitimacy for the prediction of cardiovascular problems.

Table 2: Performance Results shown by employed Techniques.

Summary	Classifiers				
	Decision Tree	**J48**	**Neural Network (MLP)**	**Naïve Bayes**	**Support Vector Machine (SMO)**
Correctly Classified Instances	494.757 (79.671 %)	521.4325 83.9666 %	529.7328 85.3032 %	498.13 (80.2143 %)	502.75 (80.95%)
Incorrectly Classified Instances	126.243 (20.329 %)	99.5675 16.0334 %	91.2672 14.6968 %	122.8695 (19.7857 %)	118.24 (19.04%)
Kappa statistic	0.5934	0.6793	0.7061	0.6043	0.6192
Mean absolute error	0.3046	0.1906	0.1506	0.2089	0.1904
Root mean squared error	0.3832	0.3686	0.3544	0.3789	0.4364

Relative absolute error	60.9249 %	38.1201 %	30.116 %	41.7792 %	38.0826%
Root relative squared error	76.6466 %	73.7212 %	70.8823 %	75.7836 %	87.2718%
Total Number of Instances	621	621	621	621	621

Table 3 presents the performance results shown by decision tree related to cardiovascular problems.

=== Detailed Accuracy shown by Classes ===

Table 3: Detailed Accuracy of Decision Tree

	TP Rate	**FP Rate**	**Precision**	**Recall**	**F-Measure**	**MCC**	**ROC Area**	**PRC Area**	**Class**
	0.824	*0.230*	*0.781*	*0.824*	*0.802*	*0.594*	*0.869*	*0.855*	***Yes***
	0.770	*0.176*	*0.814*	*0.770*	*0.791*	*0.594*	*0.869*	*0.869*	***No***
Weighted Average	***0.797***	***0.203***	***0.798***	***0.797***	***0.797***	***0.594***	***0.869***	***0.862***	

Table 4 presents the performance results shown by J48 related to cardiovascular problems.

Table 4: Detailed Accuracy of J48

	TP Rate	**FP Rate**	**Precision**	**Recall**	**F-Measure**	**MCC**	**ROC Area**	**PRC Area**	**Class**
	0.831	*0.152*	*0.846*	*0.831*	*0.838*	*0.679*	*0.874*	*0.882*	*yes*
	0.848	*0.169*	*0.834*	*0.848*	*0.841*	*0.679*	*0.874*	*0.816*	*no*
Weighted Average	***0.840***	***0.160***	***0.840***	***0.840***	***0.840***	***0.679***	***0.874***	***0.849***	

Table 5 presents the performance results shown by neural networks related to cardiovascular problems.

Table 5: Detailed Accuracy of N.N (MLP)

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.849	*0.143*	*0.856*	*0.849*	*0.852*	*0.706*	*0.910*	*0.902*	*yes*
	0.857	*0.151*	*0.850*	*0.857*	*0.854*	*0.706*	*0.910*	*0.906*	*no*
Weighted Average	***0.853***	***0.147***	***0.853***	***0.853***	***0.853***	***0.706***	***0.910***	***0.904***	

Table 6 presents the performance results shown by naïve bayes related to cardiovascular problems.

Table 6: Detailed Accuracy of Naïve Bayes

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.773	*0.169*	*0.821*	*0.773*	*0.796*	*0.605*	*0.895*	*0.884*	*yes*
	0.831	*0.227*	*0.786*	*0.831*	*0.808*	*0.605*	*0.895*	*0.902*	*no*
Weighted Average	***0.802***	***0.198***	***0.803***	***0.802***	***0.802***	***0.605***	***0.895***	***0.893***	

Table 7 presents the performance results shown by SVM related to cardiovascular problems.

Table 7: Detailed Accuracy of SVM

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.806	*0.187*	*0.812*	*0.806*	*0.809*	*0.619*	*0.810*	*0.751*	*yes*
	0.813	*0.194*	*0.807*	*0.813*	*0.810*	*0.619*	*0.810*	*0.750*	*no*

Weighted Average	***0.810***	***0.190***	***0.810***	***0.810***	***0.619***	***0.619***	***0.810***	***0.751***	

Table 8 presents Confusion Matrix of the applied algorithms. Confusion matrix is one of the best visualization tool employed to determine the performance results of employed techniques.

Table 8: Confusion Matrix shown by Employed Algorithms.

Classified as	**Confusion Matrices**									
	Decision Tree		**J48**		**MLP**		**Naïve Bayes**		**SVM**	
a = yes	*256*	*55*	*258.01*	*52.49*	*263.59*	*46.91*	*240.13*	*70.37*	*250.19*	*60.31*
b = no	*72*	*239*	*47.07*	*263.43*	*44.36*	*266.14*	*52.5*	*258*	*52.5*	*252.6*

7.2 Ensemble Approach

Ensemble method uses multiple algorithms to acquire acceptable results in comparison to be acquired from any discrete or integral machine learning algorithms unaided. In real world applications with reference to machine learning, it presents determinate set of alternative models and depicts a flexible model from the alternatives to existence. Existing and popular, three ensemble methods are Bagging, Boosting and Blending.

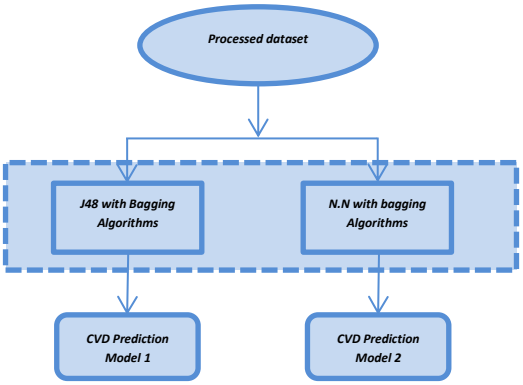

Figure 5: Ensemble Approach.

Let's proceed out by the application of J48 and NN (MLP) algorithms to experimental approach to compare the performance results using ensemble algorithmic approach in reference to heart disease prediction. Table 9 presents a interpretation of the experimental results exposed by the two proposed models/techniques in association with bagging technique.

Table 9: Performance Results using Ensemble Approach.

Classifiers with Ensemble (Bagging)	Performance Measures							
	Percent Correct %	T.P Rate %	F.P Rate %	Area under ROC %	Area under PRC %	F Measure %	Precision %	Matthews Corr. %
J48	*87.81*	*86.01*	*11.01*	*95.01*	*94.01*	*86.01*	*87.01*	*76.01*
N.N (MLP)	*87.02*	*86.01*	*12.01*	*94.01*	*93.01*	*86.01*	*86.01*	*74.01*

Performance results of ensemble approach are depicted in graphical form in figure 6.

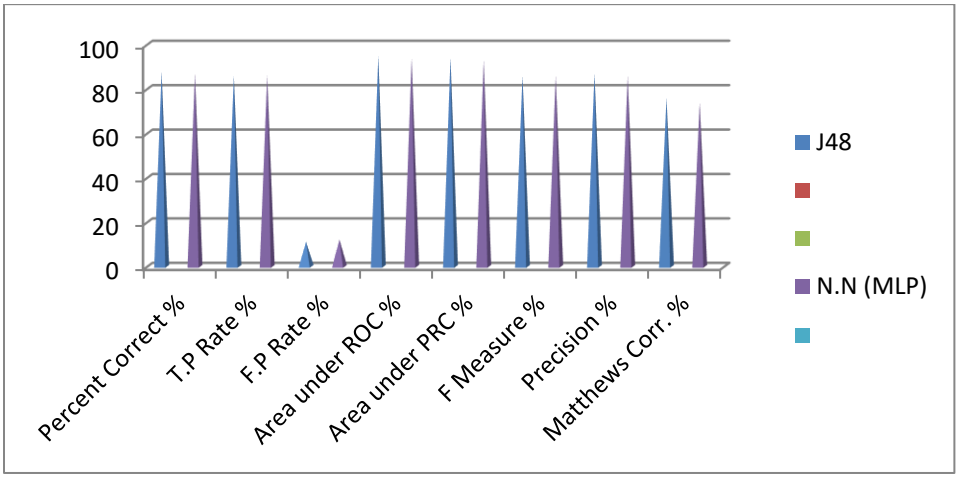

Figure 6: Performance Results shown by means of Ensemble Approach.

8. Conclusion and Future Work

This experimental work highlights the issues related to heart diseases and heart disease related prediction problems faced by the medical science. A group (five) of machine learning techniques is employed to devise various prediction models to compare and analyze the performance results related to cardiovascular diseases. Two classifiers namely J48 and neural networks conquer to others devised classifiers in the performance measures with reference to heart disease prediction capabilities. Ensemble methodology is employed to combine the prediction power of J48 and neural networks for the prediction of CVD's. Ensemble algorithms are considered influential class of machine learning techniques to combine prediction power of various/desired models. This experimental approach utilized the power of ensemble algorithms and combined the power of J48 and neural networks using WEKA platform. Both the employed techniques i.e. J48 and neural networks shows an enhanced performance results as compared to results shown by J48 and neural networks without ensemble algorithms. The comparative results undoubtedly depicts from the table 2 and table 8 that J48 and neural networks are more acceptable than other techniques and both these techniques shows feeble enhancement using ensemble (bagging) approach.

An open confront is available to research community to exploit more techniques in ensemble nature for the prediction of cardiovascular diseases. Moreover to enhance the predictive power of the developed models using bulky but related datasets could be employed in future also.

9. References

A. K. Sen, S. B. (2013). A Data Mining Technique for Prediction of Coronary Heart Disease Using Neuro-Fuzzy Integrated Approach Two Level. *International Journal of Engineering and Computer Science, 2*, 1663-1671.

Ansari, A. (n.d.). Automated Diagnosis of Coronary Heart Disease Using Neuro-Fuzzy Integrated System . *2011 World Congress on Information and Communication Technologies 978-1-4673-0125- 1@ 2011 IEEE* , (pp. 1383-1388).

Boshra Bahrami, M. H. (2015., February). Prediction and Diagnosis of Heart Disease by Data Mining Techniques. *Journal of Multidisciplinary Engineering Science and Technology (JMEST) ISSN: 3159-0040 , 2*(2).

Carlos Ordonez Teradata. (2006). Association Rule Discovery with the Train and Test Approach for Heart Disease Prediction. *Published in Transactions on Information Technology in Biomedicine (TITB Journal)*, 10(2):334-343.

Chaitrali S. Dangare, S. S. (2012, June). Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques. *International Journal of Computer Applications (IJCA), 47*, 44-48.

D'Souza, A. (2015, March). Heart Disease Prediction Using Data Mining Techniques Andrea. *International Journal of Research in Engineering and Science (IJRES) ,ISSN (Online): 2320-9364, 3*(3), 74-77.

Frawley, P.-S. (1996). Knowledge Discovery in Databases: An Overview.

Godara, M. K. (2011). Comparative study of data mining classification methods in cardiovascular disease prediction. *IJCST , 2*, 304-305.

Hlaudi Daniel Masethe, M. A. (22-24 October,,2014). Prediction of Heart Disease using Classification Algorithms. *WCECS 2014, San Francisco, USA.*

Miss. Chaitrali S. Dangare, D. M. (2012). A data mining approach for prediction of heart disease using neural networks . *International journal of computer engineering and technology*.

Moloud Abdar, S. R. (2015, December). Comparing Performance of Data Mining Algorithms in Prediction Heart Diseases. *5*.

Nidhi Bhatla, K. J. (2012, ., September). A Novel Approach for Heart Disease Diagnosis using Data Mining and Fuzzy Logic. *International Journal of Computer Applications ISSN 0975 – 8887, 54*, 16-21.

Ralf Mikut, M. R. (2011, September/October). Interdisciplinary Reviews:Data Mining and Knowledge discovery. *1*(5), 431-443.

Resul Das, I. T. ((2009)). Effective diagnosis of heart disease through neural networks ensembles. . *Expert Systems with Applications*, 7675–7680.

S. Kiruthika Devi, S. K. (2016, October). Prediction of Heart Disease using Data Mining Techniques. *Indian Journal of Science and Technology , 9(39)*.

Sellappan Palaniappan, R. A. (2008, August). Intelligent Heart Disease Prediction System Using Data Mining Techniques. *IJCSNS International Journal of Computer Science and Network Security , 8*.

Wilson, P. e. (1998). Prediction of Coronary Heart Disease using Rusk Factor Categories. *American Heart Association Journal*.