Evaluating Predictive Algorithms using Receiver-Operative Characteristics for Coronary Illness among Diabetic Patients

Tahira Mahboob, Saman Sahaheen, Nuzhat Tahir, Mukhtiar Bano
Department of Software Engineering
Fatima Jinnah Women University, Pakistan

Abstract—The grouping of information is a typical method in Machine learning. Information mining assumes a crucial part to extract learning from vast databases from operational databases. In medicinal services Data mining is a creating field of high significance, giving expectations and a more profound comprehension of restorative information sets. Most extreme information mining technique relies on an arrangement of elements that characterizes the conduct of the learning calculation furthermore straightforwardly or by implication impact of the multifaceted nature of models. Coronary illness is the main sources of death over the past years. Numerous scientists utilize a few information digging methods for the diagnosing of coronary illness. Diabetes is one of the incessant maladies that emerge when the pancreas does not deliver enough insulin. The vast majority of the frameworks have effectively utilized Machine learning strategies, for example, Naïve Bayes Algorithm, Decision Trees, logistic Regression and Support Vector Machines to name a few. These techniques solely rely on grouping of the information with respect to finding the heart variations from the norm. Bolster vector machine is an advanced strategy has been effectively in the field of machine learning. Utilizing coronary illness determination, the framework presented predicts using characteristics such as, age, sex, cholesterol, circulatory strain, glucose and the odds of a diabetic patient getting a coronary illness using machine learning algorithms.

Keywords—*Artificial neural networks (ANN); Decision tree; Naïve Bayes; Logistic Regression and Clustering*

I. INTRODUCTION

The improvements of system integration as well as software development techniques have made an advanced generation for the complex computer systems. The researchers have presented numerous challenges by using systems. Machine learning can be defined as a scientific field. The developed algorithms associates the real time problem based on previous statistics, and performs to resolve a real time problem under definite set of instructions and rules. Both machine learning and data mining algorithms use design formatted by means of same set of fields such as features, attributes, inputs, or variables. When an example or an instance contains the correct output (class label) then the learning process is known as supervised learning. In other words, the process of machine learning without knowing the class label of instances is called unsupervised learning. Clustering is an unsupervised learning method used for classifying data. The main objective of clustering is to describe data in an unsupervised learning ways. On the other hand, classification and regression are predictive methods. In the present research, we focus on supervised machine learning. The proposed study applies diverse algorithms such as Naïve Bayes, Decision Trees (C4.5) and Logistic Regression for Classification. Receiving Operating Characteristics (ROC) curve on the classification of algorithms has been analyzed for evaluation of predicted results on basis of data set attributes and values.

Coronary heart disease: Heart is a basic organ of our body as life is all subject to capable working of heart. If spread of blood is inefficient in the body that leads to the organs like kidney, brain tending to deteriorate and if heart stops, the end will be happen within minutes. The word Heart sickness suggests disease of heart and vein structure inside the heart. There are number of segments that extend the threat of Heart disorder, for instance, smoking, cholesterol, not exactly stellar eating schedule, hypertension, blood cholesterol, physical absence of movement, hyper weight and family history of coronary disease.

The rest of the paper is structured as Literature Survey included in Section-II followed by methodology in Section-III. Section-IV covers the implementation of the proposed study which is analyzed in Section-V. The last Section Covers the Results and Analysis along with Conclusion.

II. LITERATURE SURVEY

In this segment, we review the existing literature and confer about different aspects of data mining applications in prediction of heart diseases.

In Year 2007, Choi S.et.al [1] presents discovery strategy for heart deformities utilizing the cardiovascular sound trademark waveform (CSCW) with Information Grouping Method. The research inferred that presented framework is reasonable for the recognizable proof of patients with high/low cardiovascular danger.

In year 2012 Nambiar V. P. et.al [2] presents another system to perceive driver sluggishness by requesting the power scope of a man's Distinctions Heart Rate variability (HRV) data, which is made using Genetic Algorithm. The maker presumes that the precise level of drowsiness is hard to

choose and is not secured in the database used as a piece of this paper.

In the year 2013 M Makkiet.al [3] proposed the prediction of coronary illness using two receiving wires situated on the mid-section to screen the heart action. Synchronous with the radar estimations, the Electrocardiogram (ECG) and pulse is measured. The author presumes that the capacity to distinguish changes in the heart affirms medicinal radar as a reasonable analytic apparatus.

In year 2008 Han C. H. et.al [4] presents that strategy MCG (Magnetocardiogram) is additionally utilized for coronary illness recognition like the ECG (Electrocardiogram). The author presumes that it can't give the propelled e-Health administrations in light of immense measure of information, which may be handled and overseen.

In the year 2009 Wu W. et.al [5] presents a non-intrusive multi-channel ECG checking that is planned and executed to dissect the HRV of people amid various visual boosts. It is reasoned that a large portion of the HRV parameters changed fundamentally, and long haul negative visual jolt may render the capacity of ANS powerless which influence our body adversely. So positive visual incitement is required that helps us for more noteworthy fixation and spotlight on our undertakings, objectives and desires under the positive visual boost.

In year 2007, Manriquez A. et.al [6] presents another calculation proposed for QRS onset and counterbalance discovery in single lead electrocardiogram (ECG) records. It is inferred that the new calculations can be assessed with the PhysioNet QT database. As far as STD quality, it beats alternate calculations assessed on the same database. The location blunders in these outcomes are likewise around the resiliences acknowledged by specialists.

In the year 2014, Masethe H. D. et.al [7] exhibited that information mining calculations, for example, J48, Naïve Bayes, REPTREE, CART, and Bayes Net are connected in this examination for anticipating heart assaults. The author reasoned that the prescient precision controlled by J48, REPTREE and SIMPLE CART calculations proposes that the parameters utilized are solid pointers to anticipate the proximity of heart illnesses.

In year 2010 F. Sufi et.al [8] presents the risk of Cardiovascular Disease (CVD) related to the use of cellular telephone based computational stages, body sensors and remote correspondences is multiplying. Since cell phones have limited computational resources, existing PC based complex CVD disclosure figuring are every now and again prohibited for remote telecardiology applications Moreover, if the current Electrocardiography (ECG) based CVD recognition calculations are embraced for portable telecardiology applications, then there will be a need to handle delays because of the computational complexities of the current calculations.

In Year 2010, J. Bushra et.al [9] proposed another technique to identify the QRS Complex by wavelet based methodologies. The author investigates a non-stationary sign utilizing Gaussian wavelet and the recognized interest is indicated by relating them through the nearby extrema in wavelet change.

In the year 2009 Aardal, Ø., et.al [10] exhibited body sensor systems (BSNs) as high preparation is required in decompression that would squander profitable vitality in the asset and force obliged sensor hubs. In this paper, to analyze cardiovascular irregularity, for example, Ventricular tachycardia, a novel framework to break down and characterize compacted ECG signal by utilizing a PCA for highlight extraction and k-mean for bunching of ordinary and unusual ECG signals is proposed.

In year 2014 J. Lee et.al [11] presents that strategy ECG which analyzes cardiovascular ailments utilizing blood testing, ought to give early location to the ailment and more dependable checks. In year 2014 D. Khemphila, A., et.al [12] presents that technique to gain the clinical and ECG information, in order to prepare the Artificial Neural Network to precisely analyze the heart and foresee variations from the norm.

In year 2014 Mishra, S. K., et. al[13] presents telemetry framework to procure the Electrocardiogram (ECG) waveform and examine it utilizing a calculation created to identify cardiovascular irregularities in patients. The ECG information is sent continuously to the patient's cell phone from a Bluetooth sensor. Two methodologies are being produced to prepare the information. The Web server methodology is to send the information from the telephone to a Web server where the information will be examined and prepared and the outcomes will be sent back; this should be possible through Wi-fi or a 3G association.

In year 2014 P. Kaur et.al [14] presents in this work helps in location of heart rate, ECG variations from the norm and ensuing estimation of related infection utilizing different modules. The calculation outlined can get the information from measured document or recreate the ECG signal, form the information, and shows ECG waveform heart rate and its variations from the norm.

III. METHODOLOGY

The proposed strategy of the subsequent study is diagnosing powerlessness of patients of heart maladies. The methodology of the proposed study is presented in Fig.1.We took 303 records of patients to perform the experimentation. Thus the data set used for setting up the classifier contains 303 diabetic patient records out of which 175 records are of those having coronary sickness (positive cases) and the staying 128 records are of those not having coronary disease (negative cases). A sample of the characteristics making up each record/dataset is presented in Table1 and the attributes chosen for the prediction is presented in Table.2. The dataset is available at [*]"UCL Machine Learning Repository*"
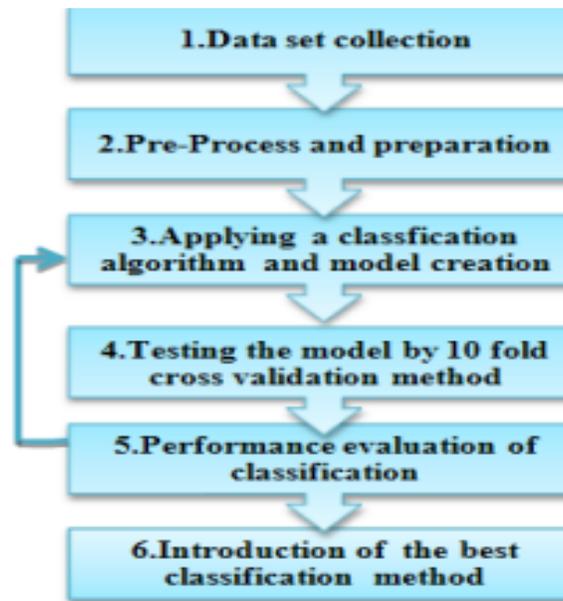
Fig. 1. Proposed Model of Data Flow

*The dataset of heart diseases is obtained from "UCL Machine Learning Repository" available at the link: "http://archive.ics.uci.edu/ml/"

TABLE I. DATA SET USED FOR DETECTION OF HEART ABNORMALITIES

Relation: cleveland-14-heart-disease

No.	age Numeric	sex Nominal	cp Nominal	trestbps Numeric	chol Numeric	fbs Nominal	restecg Nominal	thalach Numeric	exang Nominal	oldpeak Numeric	slope Nominal	ca Numeric	thal Nominal	num Nominal
1	63.0	male	typ_a...	145.0	233.0	t	left_v...	150.0	no	2.3	down	0.0	fixed_...	(50
2	67.0	male	asympt	160.0	286.0	f	left_v...	108.0	yes	1.5	flat	3.0	normal	)50_1
3	67.0	male	asympt	120.0	229.0	f	left_v...	129.0	yes	2.6	flat	2.0	revers...	)50_1
4	37.0	male	non_a...	130.0	250.0	f	normal	187.0	no	3.5	down	0.0	normal	(50
5	41.0	female	atyp_...	130.0	204.0	f	left_v...	172.0	no	1.4	up	0.0	normal	(50
6	56.0	male	atyp_...	120.0	236.0	f	normal	178.0	no	0.8	up	0.0	normal	(50
7	62.0	female	asympt	140.0	268.0	f	left_v...	160.0	no	3.6	down	2.0	normal	)50_1
8	57.0	female	asympt	120.0	354.0	f	normal	163.0	yes	0.6	up	0.0	normal	(50
9	63.0	male	asympt	130.0	254.0	f	left_v...	147.0	no	1.4	flat	1.0	revers...	)50_1
10	53.0	male	asympt	140.0	203.0	t	left_v...	155.0	yes	3.1	down	0.0	revers...	)50_1
11	57.0	male	asympt	140.0	192.0	f	normal	148.0	no	0.4	flat	0.0	fixed_...	(50
12	56.0	female	atyp_...	140.0	294.0	f	left_v...	153.0	no	1.3	flat	0.0	normal	(50
13	56.0	male	non_a...	130.0	256.0	t	left_v...	142.0	yes	0.6	flat	1.0	fixed_...	)50_1
14	44.0	male	atyp_...	120.0	263.0	f	normal	173.0	no	0.0	up	0.0	revers...	(50
15	52.0	male	non_a...	172.0	199.0	t	normal	162.0	no	0.5	up	0.0	revers...	(50
16	57.0	male	non_a...	150.0	168.0	f	normal	174.0	no	1.6	up	0.0	normal	(50
17	48.0	male	atyp_...	110.0	229.0	f	normal	168.0	no	1.0	down	0.0	revers...	)50_1
18	54.0	male	asympt	140.0	239.0	f	normal	160.0	no	1.2	up	0.0	normal	(50
19	48.0	female	non_a...	130.0	275.0	f	normal	139.0	no	0.2	up	0.0	normal	(50
20	49.0	male	atyp_...	130.0	266.0	f	normal	171.0	no	0.6	up	0.0	normal	(50
21	64.0	male	typ_a...	110.0	211.0	f	left_v...	144.0	yes	1.8	flat	0.0	normal	(50
22	58.0	female	typ_a...	150.0	283.0	t	left_v...	162.0	no	1.0	up	0.0	normal	(50
23	58.0	male	atyp_...	120.0	284.0	f	left_v...	160.0	no	1.8	flat	0.0	normal	)50_1
24	58.0	male	non_a...	132.0	224.0	f	left_v...	173.0	no	3.2	up	2.0	revers...	)50_1
25	60.0	male	asympt	130.0	206.0	f	left_v...	132.0	yes	2.4	flat	2.0	revers...	)50_1
26	50.0	female	non_a...	120.0	219.0	f	normal	158.0	no	1.6	flat	0.0	normal	(50
27	58.0	female	non_a...	120.0	340.0	f	normal	172.0	no	0.0	up	0.0	normal	(50
28	66.0	female	typ_a...	150.0	226.0	f	normal	114.0	no	2.6	down	0.0	normal	(50
29	43.0	male	asympt	150.0	247.0	f	normal	171.0	no	1.5	up	0.0	normal	(50
30	40.0	male	asympt	110.0	167.0	f	left_v...	114.0	yes	2.0	flat	0.0	revers...	)50_1
31	69.0	female	typ_a...	140.0	239.0	f	normal	151.0	no	1.8	up	2.0	normal	(50
32	60.0	male	asympt	117.0	230.0	t	normal	160.0	yes	1.4	up	2.0	revers...	)50_1
33	64.0	male	non_a...	140.0	335.0	f	normal	158.0	no	0.0	up	0.0	normal	)50_1
34	59.0	male	asympt	135.0	234.0	f	normal	161.0	no	0.5	flat	0.0	revers...	(50
35	44.0	male	non_a...	130.0	233.0	f	normal	179.0	yes	0.4	up	0.0	normal	(50
36	42.0	male	asympt	140.0	226.0	f	normal	178.0	no	0.0	up	0.0	normal	(50
37	43.0	male	asympt	120.0	177.0	f	left_v...	120.0	yes	2.5	flat	0.0	revers...	)50_1
38	57.0	male	asympt	150.0	276.0	f	left_v...	112.0	yes	0.6	flat	1.0	fixed_...	)50_1
39	55.0	male	asympt	132.0	353.0	f	normal	132.0	yes	1.2	flat	1.0	revers...	)50_1

TABLE II.   ATTRIBUTES TAKEN FOR DETECTION OF HEART DISEASE

Sr.#	Attribute name	Depiction/values
1	Age	Age of the patients in year
2	Sex	Sex of patients takes two value(Female ,Male)
3	chest pain(Cp)	chest pain type in the heart patients Value 1: typical angina Value 2: atypical angina Value 3: non-angina pain Value 4: asymptomatic
4	resting blood pressure(Trestbps)	resting blood pressure (in mm Hg on admission to the hospital)
5	cholesterol (chol)	Heart patient serum cholesterol in mg/dl
6	fasting blood sugar (fbs)	(fasting blood sugar > 120 mg/dl)
7	resting electrocardiographic (restecg)	latent electrocardiographic effects Value 0: normal Value 1: having ST-T wave abnormality
8	thalach	maximum heart rate achieved
9	(exang)	implementation prompted angina (1 = yes; 0 = no)
10	old peak	ST dejection made by exercise relative to rest
11	slope	the slope of the peak exercise ST segment Value 1: upsloping Value 2: flat Value 3: down sloping
12	ca	quantity of main vessels (0-3) colored by fluoroscopy
13	thal	3 = normal; 6 = fixed defect; 7 = reversible defect
14	num	analysis of core disease (angiographic disease status) Value 0: < 50% diameter narrowing Value 1: > 50% diameter narrowing

IV.   IMPLEMENTATION

Data Mining Techniques Used For Predictions: The diverse information mining arrangement methods utilized as a part of our examination, i.e. Neural Networks, Decision Trees, bunching, logistic relapse, and Naive Bayes are utilized to break down the dataset of coronary illness patients targeting the prediction of coronary illnesses. As a means to evaluate the accuracy of predictive capacity of the applied algorithms the receiving operating characteristics [15] of each have been plotted followed by a comparison to find the best fit algorithms for predictions. [16] R.O.C outline is a valuable visual instrument for looking at order techniques. It demonstrates the exchange off between the genuine positive rate and the false positive rate for a given model. ROC outline depends on the restrictive probabilities affectability and specificity.

A. Neural Networks:

A simulated neural system Artificial Neural Networks (ANN) regularly called as "neural system" (NN) [17]. It is a numerical model or computational model taking into account natural neural system. As it were, it is a reproduction of organic neural framework in the field of restorative. The receiving operating characteristics have first been plotted in Fig. 2 depicting the correctness of the results using ANN algorithm and the ROC measurement observed is 0.891.
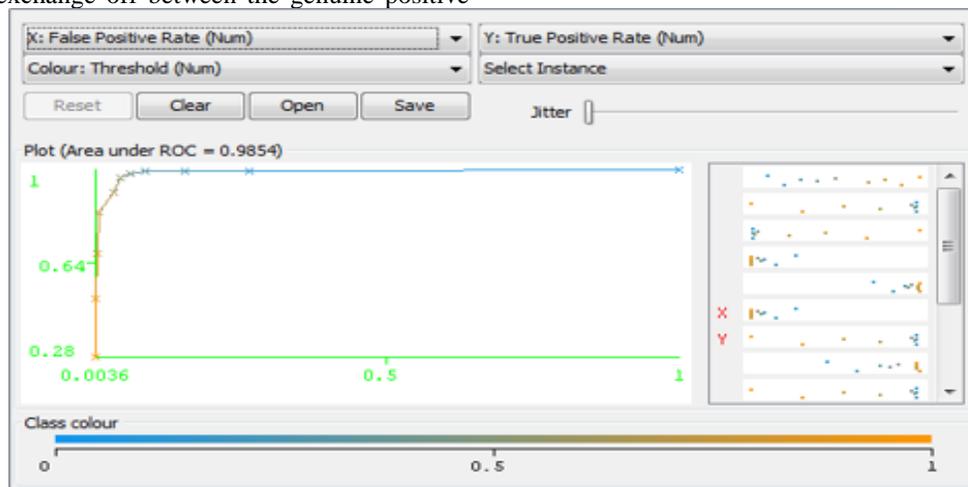

Fig. 2.   ROC of Artificial Neural Networks

B. Decision Trees:

The decision tree methodology is all the more effective for the arrangement of issues. There are two phases used as a piece of this framework; building a tree and applying the tree to the dataset of coronary ailment patients. The Algorithm count uses pruning methodology to create a tree. Pruning is a system which diminishes the degree of tree by ousting over fitting data, which may prompts poor precision in predications examination.[18] [19]The DT figures recursively gatherings' data until it will sort data as sublimely as could sensibly be

normal. This methodology gives most amazing exactness on planning data sets. A sample DT result is given in Fig. 4 depicting/Classifying attribute 'Chest Pain' as 'unstable angina', 'stable angina' and "non-angina'. The receiving operating characteristics of the DT are shown in Fig. 4 giving a value of 0.952.
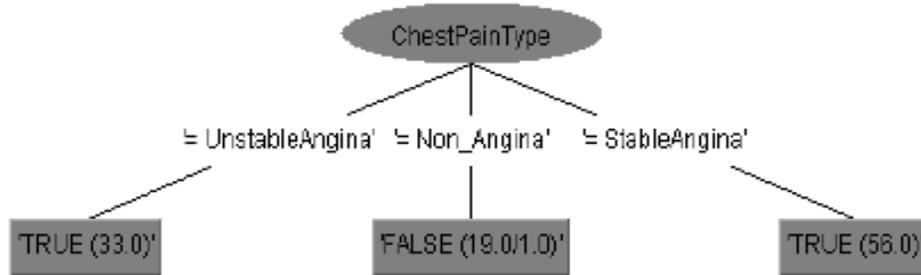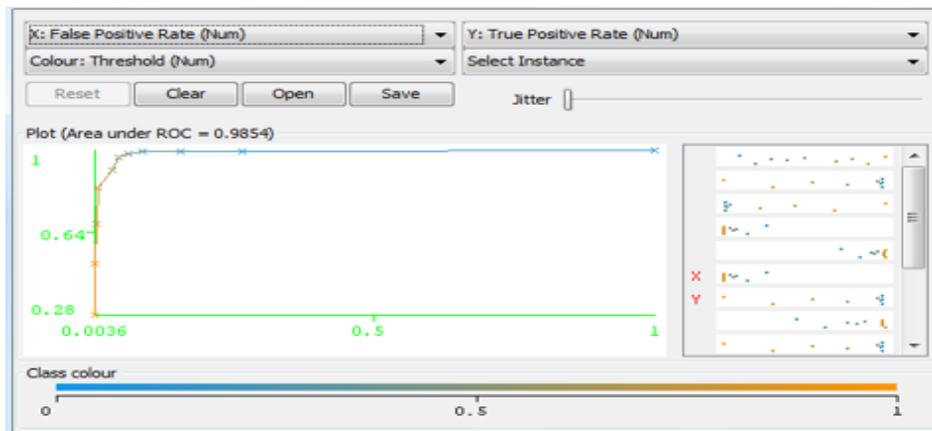
Fig. 3.   Decision Tree of chest pain type

Fig. 4.   ROC Decision Tree

C. *Naive Bayes:*

Naive Bayes classifier depends on Bayes hypothesis. This classifier calculation utilizes contingent freedom, implies as it accept that a quality worth on a given class is autonomous of the estimations of different traits. [20][21]The ROC curve for the NaiveBayes algorithm is given in Fig.5. Though very simple to implement but doesn't give a high value for ROC measurement as compared to other implemented algorithms. i.e. 0.91
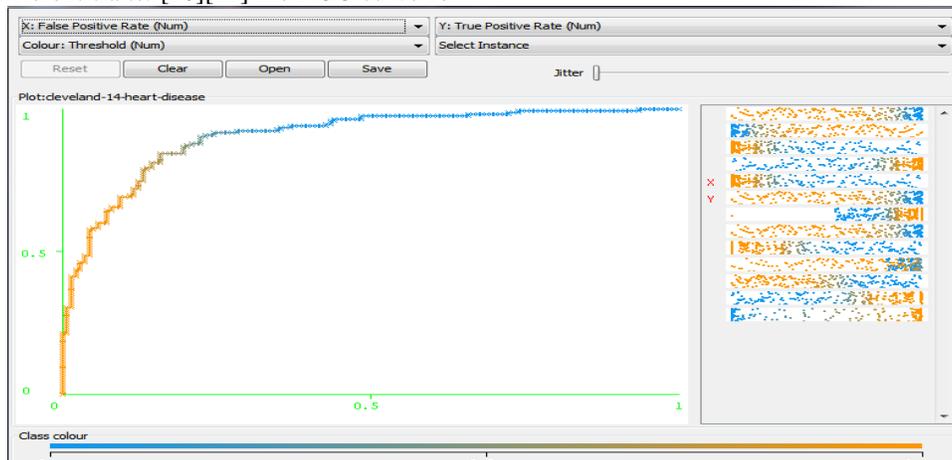
Fig. 5.   ROC Naive Baye

D. *Clustering:*

Clustering is an information mining procedure that makes valuable group of items that have same trademark utilizing this system. Not quite the same as grouping, clustering strategy likewise characterizes the classes and place objects in them, while in order articles are doled out into predefined classes [22]. For instance in expectation of coronary illness by utilizing grouping we get bunch or we can say that rundown of

patients which have same danger variable such as different rundown of patients with high glucose and related danger

=== **Model and evaluation on training set** ===
Clustered Instances
0   175 (58%)
1   128 (42%)

E. Logistic regression:

component. Results predominant cluster having instances 175 in cluster 1 and 128 in cluster 2.

Logistic regression is a measurable technique for examining a dataset in which there are one or more autonomous variables that decide a result. [23] The result is measured with a dichotomous variable (in which there are just two conceivable results). The ROC of the Logistic regression is presented in Fig.6 giving an ROC measurement of 0.938.
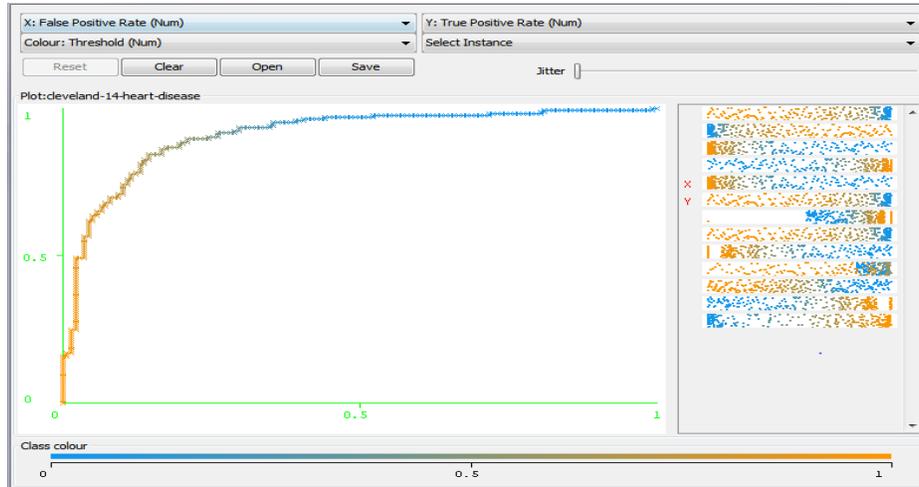

Fig. 6.   ROC Logistic regression

F. Prediction Analysis:

The prediction as it name suggested is one of an information mining strategies that finds relationship between autonomous variables and relationship amongst reliant and free variables. For example, expectation examination system can be utilized as a part of offer to anticipate benefit for the

future on the off chance that we consider deal is an autonomous variable, benefit could be a variable. At that point taking into account the chronicled deal and benefit information, we can draw a fitted relapse bend that is utilized revenue driven expectation.   The ROC measurement for Prediction Analysis is. 0.908.
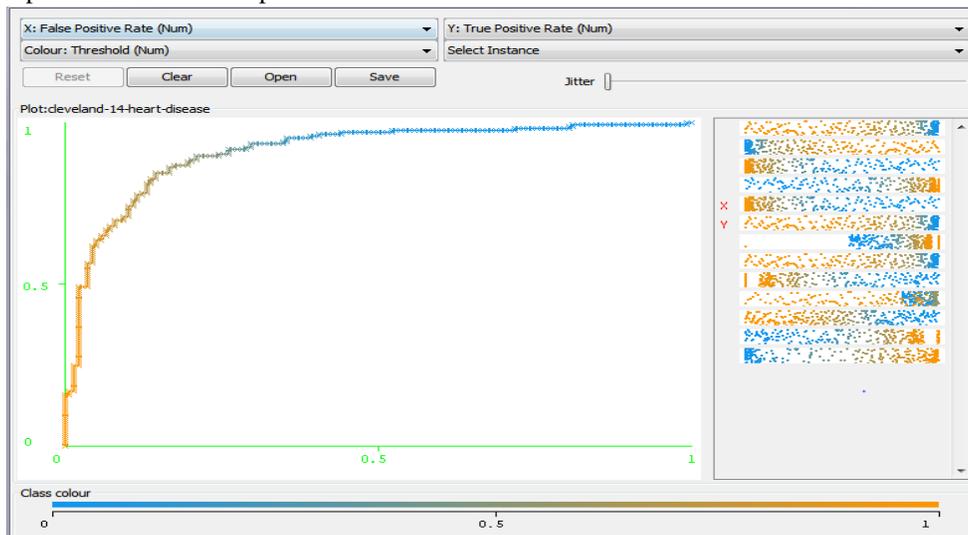

Fig. 7.   ROC Prediction Analysis

G. Support Vector Machine:

The data set used for setting up the classifier contains 303 diabetic patient records out of which 175 records are of those having coronary sickness (positive cases) and the staying 128 records are of those not having coronary disease (negative cases). These records after satisfactory pre-get ready are given

as commitment to set up the SVM classifier. Support Vector machines have been used as a classifier for feature selection in CAD disease in combination with genetic algorithms [24]. Also [SVM technique has been used to predict heart abnormalities as Babaoğlu, I., et al. 25]

The confusion matrix demonstrating the precision of the SVM classifier for the given information set is presented in Table 4. The confusion matrix is a Visualization apparatus utilized as a part of administered realizing which contains real and predicated grouping. Every section speaks to example in a predicated class and every line speaks to case in a genuine class.

Fig. 8 shows the comparative plots of the ROC of that servers as a basis for analysis.

TABLE III.   CONFUSION MATRIX

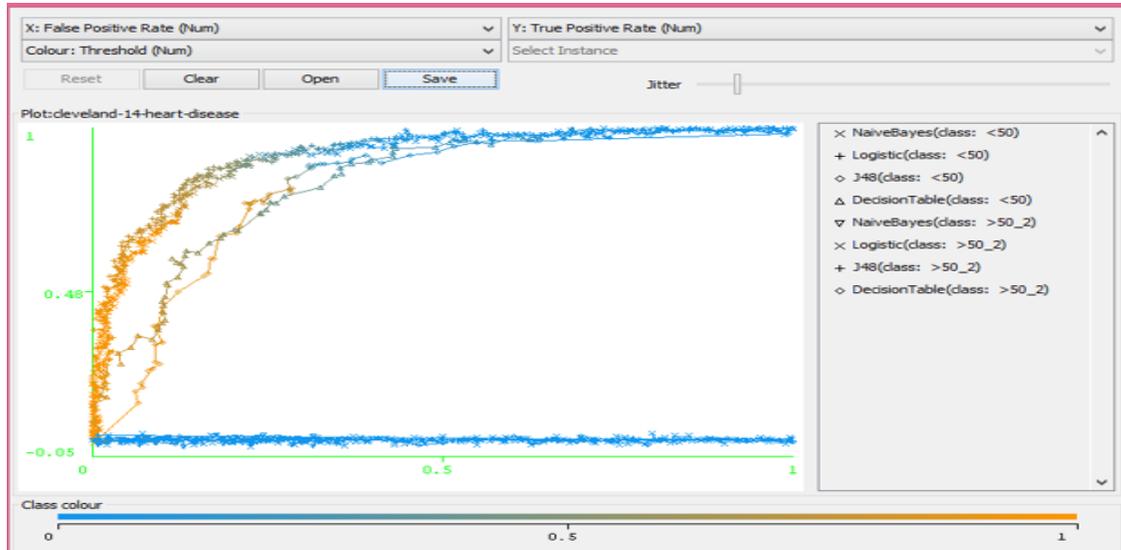| a	b	<-- classified as
181	23	a = no
48	51	b = yes

Fig. 8.   Comparative ROC Curve for Various Techniques

V.   ANALYSIS AND RESULTS

A comparative analysis of the ROC curves for implemented techniques has been presented in Fig. 8 where the Decision Tree algorithm produces the best curve having value of 0.952. The Accuracy measurements of the results obtained are presented in Table 4. Hence it can be concluded that the DT are best classifier for the prediction of heart disease given the dataset. The results shows that the DT algorithm applied on the data set is more accurate than the rest of the algorithm used such as Naive Bayes, Logistic Regression, ANN, The DT shows the 92% result to predict the abnormality found in the human heart. The whole model is proposed to collect data and analyzed using machine learning algorithm one by one. The value is predicted by correctly classified the instances. As the instances found that are correctly classified shows that the people who have chronic heart diseases. By predicting the results we reach the results to find the heart abnormalities in patients. As the rest of algorithms such as Naive Bayes show 83.49% correctly specified instances then decision tree specified 82.83% results which are correctly specified. ANN specifies correct parameters about 80.85%. Logistic Regression specifies correct instances 87.13% and Prediction 76.57%. A comparative graph clearly depicts the discussed results. The results show that we can find a way to specify correct instances and which algorithm is best to use.

TABLE IV.   RESULT ANALYSIS ON THE BASIS OF ACCURACY

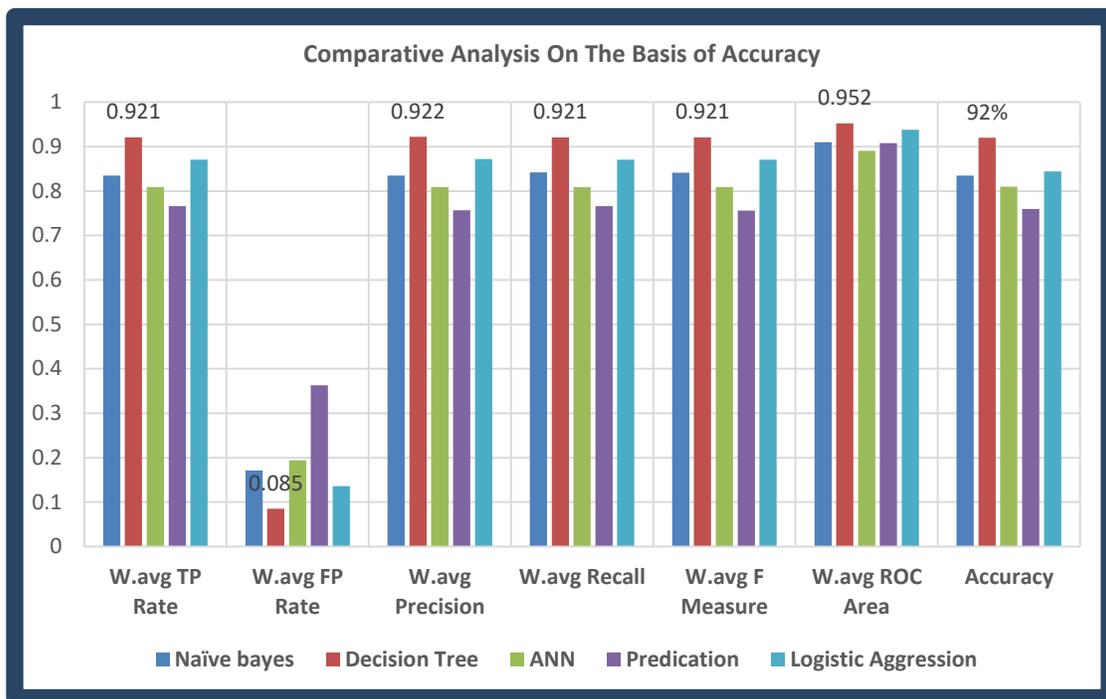| Sr.#	Algorithm	W.avg TP Rate	W.avg FP Rate	W.avg Precision	W.avg Recall	W.avg F-Measure	W.avg ROC Area	Accuracy
1.	Naïve bayes	0.835	0.171	0.835	0.842	0.841	0.91	83.5%
2.	Decision Tree	0.921	0.085	0.922	0.921	0.921	0.952	92%
3.	ANN	0.809	0.194	0.809	0.809	0.809	0.891	81%
4.	Predication	0.766	0.363	0.757	0.766	0.756	0.908	76%
5.	Logistic Aggression	0.871	0.136	0.872	0.871	0.871	0.938	84.4%

Fig. 9.   Accuracy Measures Comparison of Algorithms

TABLE V.   COMPARATIVE ANALYSIS OF RESULTS

Algorithms	Naïve Bayes		Decision Tree		Prediction		ANN		Logistic Regression	
Correctly Classified instances	253	83.49%	279	92.08%	232	76.57%	245	80.85%	264	87.13%
Incorrectly Classified instances	50	16.50%	24	7.92%	71	23.43%	58	19.14%	39	12.87%
Kappa Statistics	0.6661		0.8396		0.4304		0.6141		0.7391	
Mean Absolute Error	0.0738		0.0532		0.2976		0.0772		0.0775	
Root mean squared error	0.2299		0.1624		0.4449		0.2544		0.1954	
Relative absolute error	36.80%		26.56%		67.58%		38.48%		38.65%	
Root relative squared error	72.97%		51.54%		94.85%		80.74%		62.02%	
Total Number of instances	303		303		303		303		303	

VI.   CONCLUSION

Application of Machine learning in analyzing the therapeutic information is a decent technique for considering the current connections between variables. From our proposed approach we have demonstrated that it recovers valuable connection even from qualities which are not immediate pointers of the class we are attempting to foresee. In our work we have attempted to foresee the odds of getting a coronary illness utilizing qualities from diabetic's determination and we have demonstrated that it is conceivable to analyze coronary illness powerlessness in coronary illness patients with sensible precision. There by the patients can be forewarned to change their lifestyle. We can easily get the required and accurate result to detect the heart abnormalities using machine learning and data mining algorithms. In future work we can extend this research further in all aspects of learning to the system. We will analyze and learn to the system for more accurate results by measuring more analysis using machine learning algorithms in the current research.

REFERENCES

[1] Choi, S., Jiang, Z., Kim, I. H., & Park, C. W. (2007, October). Cardiovascular abnormality detection method using cardiac sound characteristic waveform with data clustering technique. In Control, Automation and Systems, 2007. ICCAS'07. International Conference on (pp. 1596-1600). IEEE.

[2] Nambiar, V. P., Khalil-Hani, M., Sia, C. W., & Marsono, M. N. (2012, October). Evolvable block-based neural networks for classification of driver drowsiness based on heart rate variability. In Circuits and Systems (ICCAS), 2012 IEEE International Conference on (pp. 156-161). IEEE.

[3] Makki, M. (2013, November). Analyzing Electrocardiograms Via Smartphone To Detect Cardiovascular Abnormalities. In Qatar Foundation Annual Research Conference (No. 2013, pp. BIOSP-014).

[4] Han, C. H., Youn, C. H., & Jung, W. (2008, June). Web-Based System for Advanced Heart Disease Identification Using Grid Computing Technology. In Computer-Based Medical Systems, 2008. CBMS'08. 21st IEEE International Symposium on (pp. 343-348). IEEE.

[5] Wu, W., Lee, J., & Chen, H. (2009, August). Estimation of heart rate variability changes during different visual stimulations using non-invasive continuous ecg monitoring system. In Bioinformatics, Systems Biology and Intelligent Computing, 2009. IJCBS'09. International Joint Conference on (pp. 344-347). IEEE.

[6] Manriquez, A. I., & Zhang, Q. (2007, August). An algorithm for QRS onset and offset detection in single lead electrocardiogram records. In 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 541-544). IEEE.

[7] Masethe, H. D., & Masethe, M. A. (2014, October). Prediction of heart disease using classification algorithms. In Proceedings of the world congress on engineering and computer science (Vol. 2, pp. 22-24)

[8] Sufi, F., Khalil, I., & Tari, Z. (2010, August). A cardiod based technique to identify cardiovascular diseases using mobile phones and body sensors. In 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology (pp. 5500-5503). IEEE.

[9] Bushra, J., Ouadi, B., Eric, F., & Olivier, L. (2010, August). QRS Complex Detection by Non Linear Thresholding of Modulus Maxima. In Pattern Recognition (ICPR), 2010 20th International Conference on (pp. 4500-4503). IEEE

[10] Aardal, Ø., Brovoll, S., Paichard, Y., Berger, T., Lande, T. S., & Hamran, S. E. (2013, April). Detecting changes in the human heartbeat with on-body radar. In 2013 IEEE Radar Conference (RadarCon13) (pp. 1-6). IEEE.

[11] Lee, J., Jung, J., Lee, J., & Kim, Y. T. (2014, May). Diagnostic device for acute cardiac disease using ECG and accelerometer. In 2014 International Conference on Information Science & Applications (ICISA) (pp. 1-3). IEEE.

[12] Khemphila, A., & Boonjing, V. (2010, October). Comparing performances of logistic regression, decision trees, and neural networks for classifying heart disease patients. In Computer Information Systems and Industrial Management Applications (CISIM), 2010 International Conference on (pp. 193-198). IEEE.

[13] Mishra, S. K., Daman, R., Kumar, A., & Singh, S. (2014). Design and Development of Workspaces for Surgical Skill Development. Med-e-Tel 2014, 17.

[14] Kaur, P., & Sharma, R. K. (2014, May). Labview based design of heart disease detection system. In Recent Advances and Innovations in Engineering (ICRAIE), 2014 (pp. 1-5). IEEE.

[15] Polonsky, T. S., McClelland, R. L., Jorgensen, N. W., Bild, D. E., Burke, G. L., Guerci, A. D., & Greenland, P. (2010). Coronary artery calcium score and risk classification for coronary heart disease prediction. Jama, 303(16), 1610-1616.

[16] Zurada, J. M. (1992). Introduction to artificial neural systems (Vol. 8). St. Paul: West.

[17] Quinlan, J. R. (1987). Simplifying decision trees. International journal of man-machine studies, 27(3), 221-234.

[18] Schmid, H. (2013, November). Probabilistic part-ofispeech tagging using decision trees. In New methods in language processing (p. 154). Routledge.

[19] Rish, I. (2001, August). An empirical study of the naive Bayes classifier. In IJCAI 2001 workshop on empirical methods in artificial intelligence (Vol. 3, No. 22, pp. 41-46). IBM New York.

[20] Pattekari, S. A., & Parveen, A. (2012). Prediction system for heart disease using Naïve Bayes. International Journal of Advanced Computer and Mathematical Sciences, 3(3), 290-294.

[21] Soni, J., Ansari, U., Sharma, D., & Soni, S. (2011). Predictive data mining for medical diagnosis: An overview of heart disease prediction. International Journal of Computer Applications, 17(8), 43-48.

[22] Soni, J., Ansari, U., Sharma, D., & Soni, S. (2011). Predictive data mining for medical diagnosis: An overview of heart disease prediction. International Journal of Computer Applications, 17(8), 43-48.

[23] Steinwart, I., & Christmann, A. (2008). Support vector machines. Springer Science & Business Media.

[24] Babaoglu, İ., Findik, O., & Ülker, E. (2010). A comparison of feature selection models utilizing binary particle swarm optimization and genetic algorithm in determining coronary artery disease using support vector machine. Expert Systems with Applications, 37(4), 3177-3183.

[25] Babaoğlu, I., Fındık, O., & Bayrak, M. (2010). Effects of principle component analysis on assessment of coronary artery diseases using support vector machine. Expert Systems with Applications, 37(3), 2182-2185.