IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11148

Predictive Analysis for the Detection of Cardio Vascular Disease (CVD) based on Machine Learning Classification Algorithm

Dillip Narayan Sahu[1], Vijay Pal Singh[2*]

[1]Research Scholar, Department of Computer Science, OPJS University, Rajasthan, India

[2]Associate Professor, Department of Computer Science, OPJS University, Rajasthan, India

Abstract: Cardio Vascular Disease (CVD) is a leading cause of death worldwide. It is also called heart disease. As per the WHO report, around 17.9 million people die every year from Cardio Vascular Diseases, an estimated 31% of all deaths globally, and patients die mainly because of non-appropriate and un-affordable treatment. Heart related diseases are a worldwide major health crisis in the present scenario. This disease can be curable with early diagnosis and proper treatment[1][2]. The purpose of this paper is to establish some predictive analytical models using Machine Learning algorithms by taking a real time CVD dataset. In this paper, we have shown some experimental observations with the help of some Machine Learning classification algorithms, and also shown a clear vision on the predictive analysis on medical diagnosis of the Cardio Vascular Disease(CVD) using Machine Learning algorithms using which patients may get benefited by the accurate result for the better diagnose in their early treatment.

Keywords: Algorithm, Cardio Vascular Disease, Classifier, Machine Learning, Predictive Analysis.

INTRODUCTION

According to description asub-umbrella term for a number of conditions that affect the heart itself and/ or the blood vessel system. Especially the modes and highways leading to and from the heart[3][4].
Causes of cardiovascular complaint include diabetes mellitus, hypertension and hyperactive cholesterolemia.

Symptoms

The main symptoms of cardiovascular disease are:
Chest pain, Shortness of breath Pain in the body, Feeling faint, Feeling sick
Diagnosis through Machine Learning

- Electrocardiogram (ECG)
- Echocardiogram (ECHO)
- CT scan
- Magnetic Resonance Imaging (MRI)
- Nuclear stress testing
- Coronary angiogram
- PET/CT scan

Risk Factor

- Age
- Diabetes mellitus
- Tobacco smoking
- Absence of key nutritional elements such as polyphonic antioxidants
- Hyper cholesterolemia
- Higher fibrinogen and PAI-1 blood concentrations

Prevention

- Smoking conclusion (or abstinence) is one of the most effective and fluently adjustable changes
- Regular cardiovascular exercise complements the healthful eating habits Treatment through Machine Learning[5]
- Implantable cardioverter-defibrillator: controlling a chaotic heart
- Pacemakers: Generating regular heartbeats: Angiotensin-II receptor blockers.

Experiments and Observations-1

We have taken Weka-Wekato Environment for Knowledge Analysis Machine Learning tool to preprocess, clean, data prediction and classification of the CVD (Heart) disease based on the real-time Cardio Vascular Disease patient dataset(in .CSV and Attribute Relation File Format .arff)[6].

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11148

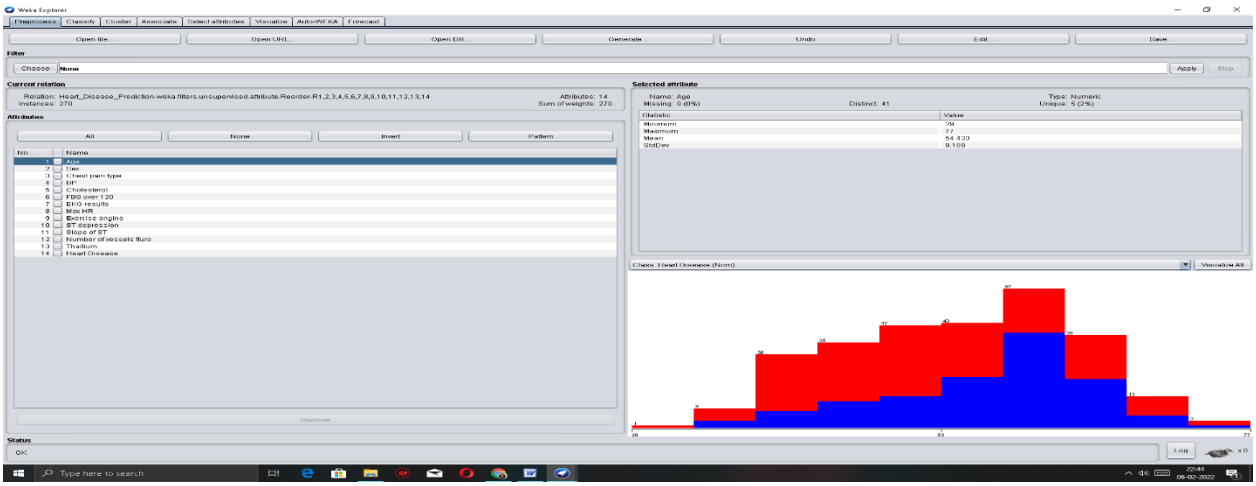

Fig.1 Preprocess of CVD Dataset having 14 Attributes

Bagging- Class for bagging a classifier to reduce variance.

Classifier Output

=== Run information ===

Scheme: weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 10 -W weka.classifiers.trees.REPTree -- -M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0

Relation: Heart_Disease_Prediction-weka.filters.unsupervised.attribute.Reorder-R1,2,3,4,5,6,7,8,9,10,11,12,13,14-weka.filters.AllFilter-weka.filters.MultiFilter-Fweka.filters.AllFilter

Instances: 270 Attributes: 14

Age Sex Chest pain type BP Cholesterol FBS over 120 EKG results

Max HR Exercise angina ST depression Slope of ST Number of vessels fluro

Thallium Heart Disease

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

Bagging with 10 iterations and base learner

weka.classifiers.trees.REPTree -M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0

Time taken to build model: 0.05 seconds

=== Stratified cross-validation ====== Summary ===

Correctly Classified Instances 242 **89.62 %**

Incorrectly Classified Instances 28 10.38 %

Experiments and Observations-2

Stacking- Combines several classifiers using the stacking method.

Classifier Output

=== Run information ===

Scheme: weka.classifiers.meta.Stacking -X 10 -M "weka.classifiers.rules.ZeroR " -S 1 -num-slots 1 -B "weka.classifiers.rules.ZeroR "

Relation: Heart_Disease_Prediction-weka.filters.unsupervised.attribute.Reorder-R1,2,3,4,5,6,7,8,9,10,11,12,13,14-weka.filters.AllFilter-weka.filters.MultiFilter-Fweka.filters.AllFilter

Instances: 270 Attributes: 14

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

Stacking

Base classifiers

ZeroR predicts class value: Absence

Meta classifier

ZeroR predicts class value: Absence

Time taken to build model: 0 seconds

=== Stratified cross-validation ====== Summary ===

Correctly Classified Instances 150 55.5556 %

Incorrectly Classified Instances 120 44.4444 %

Kappa statistic 0

Mean absolute error 0.4939

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11148

Root mean squared error 0.4969
Relative absolute error 100 %
Root relative squared error 100 %
Total Number of Instances 270
=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.000	0.000	?	0.000	?	?	0.500	0.444	Presence
1.000	1.000	0.556	1.000	0.714	?	0.500	0.556	Absence
Weighted Avg. 0.556	0.556	?	0.556	?	?	0.500	0.506	

=== Confusion Matrix ===

a	b	<-- classified as
0	120	a = Presence
0	150	b = Absence

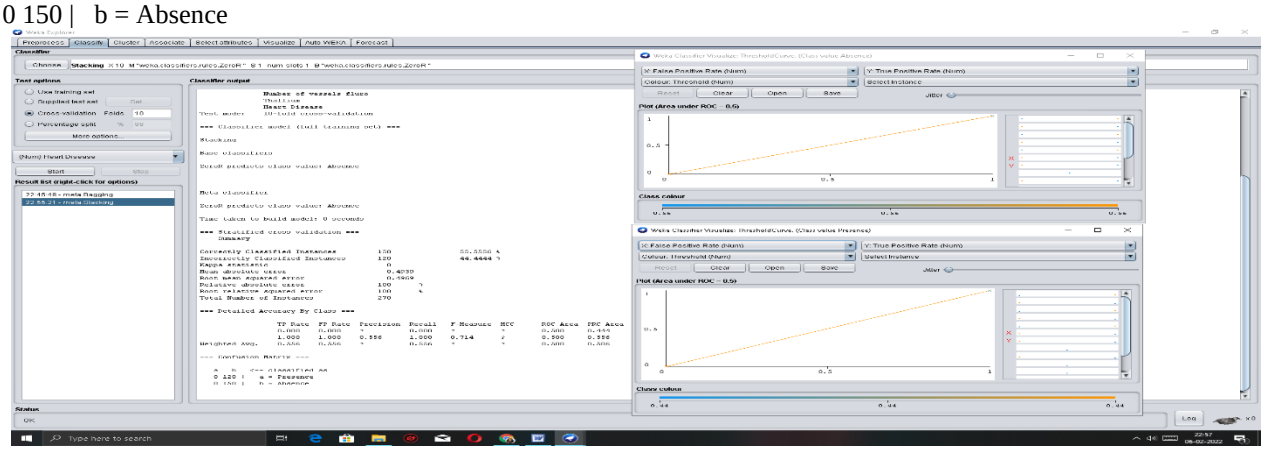

Fig.2 Stacking Classifier with Visualize curve

Experiments and Observations-3

UltraBoost- UltraBoost adaptively boosts (AdaBoosts) heterogeneous classifiers: a different classifier can be boosted at each stage.

Classifier Output

=== Run information ===
Scheme: weka.classifiers.meta.UltraBoost Relation: Heart_Disease_Prediction-weka.filters.unsupervised.attribute.Reorder-R1,2,3,4,5,6,7,8,9,10,11,12,13,14-weka.filters.AllFilter-weka.filters.MultiFilter-Fweka.filters.AllFilter
Instances: 270 Attributes: 14
Test mode: 10-fold cross-validation
=== Classifier model (full training set) ===
UltraBoost
Base classifiers
FilteredClassifier using weka.classifiers.bayes.NaiveBayes on data filtered through weka.filters.unsupervised.attribute.RemoveType -V -T nominal
Filtered Header
@relation 'Heart_Disease_Prediction-weka.filters.unsupervised.attribute.Reorder-R1,2,3,4,5,6,7,8,9,10,11,12,13,14-weka.filters.AllFilter-weka.filters.MultiFilter-Fweka.filters.AllFilter-weka.filters.unsupervised.attribute.RemoveType-V-Tnominal'
@attribute 'Heart Disease' {Presence,Absence}
@data
Classifier Model
Naive Bayes Classifier

	Class	
Attribute	Presence	Absence
	(0.45)	(0.55)

FilteredClassifier using weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4 on data filtered through weka.filters.unsupervised.attribute.RemoveType -V -T numeric
Filtered Header

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11148

@relation 'Heart_Disease_Prediction-weka.filters.unsupervised.attribute.Reorder-R1,2,3,4,5,6,7,8,9,10,11,12,13,14-weka.filters.AllFilter-weka.filters.MultiFilter-Fweka.filters.AllFilter-weka.filters.unsupervised.attribute.RemoveType-V-Tnumeric'

Classifier Model

Logistic Regression with ridge parameter of 1.0E-8

Coefficients...

Variable	Class Presence
Age	-0.0169
Sex	1.5578
Chest pain type	0.6981
BP	0.0253
Cholesterol	0.0073
FBS over 120	-0.8046
EKG results	0.3002
Max HR	-0.0212
Exercise angina	0.8217
ST depression	0.3418
Slope of ST	0.4437
Number of vessels fluro	1.168
Thallium	0.3412
Intercept	-8.4283

Odds Ratios...

Variable	Class Presence
Age	0.9832
Sex	4.7483
Chest pain type	2.0099
BP	1.0256
Cholesterol	1.0074
FBS over 120	0.4473
EKG results	1.3501
Max HR	0.979
Exercise angina	2.2743
ST depression	1.4075
Slope of ST	1.5584
Number of vessels fluro	3.2156
Thallium	1.4066

Time taken to build model: 0.01 seconds

=== Stratified cross-validation === === Summary ===

Correctly Classified Instances	225	83.3333 %
Incorrectly Classified Instances	45	16.6667 %
Kappa statistic	0.6599	
Mean absolute error	0.3364	
Root mean squared error	0.3723	
Relative absolute error	68.1079 %	
Root relative squared error	74.9264 %	
Total Number of Instances	270	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.775	0.120	0.838	0.775	0.805	0.661	0.903	0.884	Presence
	0.880	0.225	0.830	0.880	0.854	0.661	0.903	0.919	Absence
Weighted Avg.	0.833	0.178	0.834	0.833	0.833	0.661	0.903	0.903	

=== Confusion Matrix ===

a	b	<-- classified as
93	27	a = Presence

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11148

18 132 | b = Absence

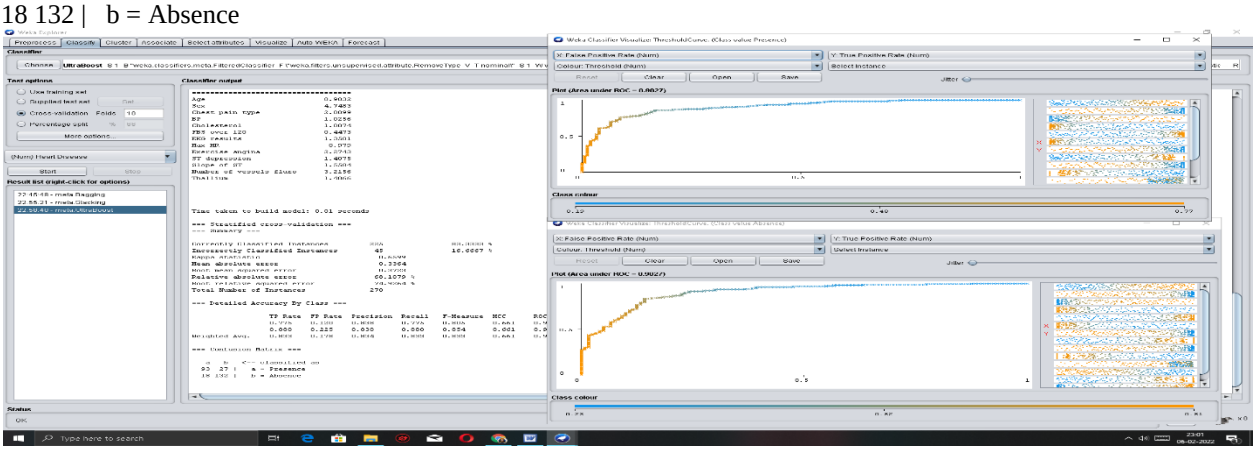

Fig.3 UltraBoost Classifier with Visualize curve

Experiments and Observations-4

Vote- Class for combining classifiers..

Classifier Output

=== Run information ===

Scheme: weka.classifiers.meta.Vote -S 1 -B "weka.classifiers.rules.ZeroR " -R AVG

Relation: Heart_Disease_Prediction-weka.filters.unsupervised.attribute.Reorder-R1,2,3,4,5,6,7,8,9,10,11,12,13,14-weka.filters.AllFilter-weka.filters.MultiFilter-Fweka.filters.AllFilter

Instances: 270 Attributes: 14 Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

Vote combines the probability distributions of these base learners:

weka.classifiers.rules.ZeroR using the 'Average' combination rule

All the models:

ZeroR predicts class value: Absence

Time taken to build model: 0 seconds

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.000	0.000	?	0.000	?	?	0.500	0.444	Presence
	1.000	1.000	0.556	1.000	0.714	?	0.500	0.556	Absence
Weighted Avg.	0.556	0.556	?	0.556	?	?	0.500	0.506	

=== Confusion Matrix ===

```
  a   b   <-- classified as
  0 120 |   a = Presence
  0 150 |   b = Absence
```

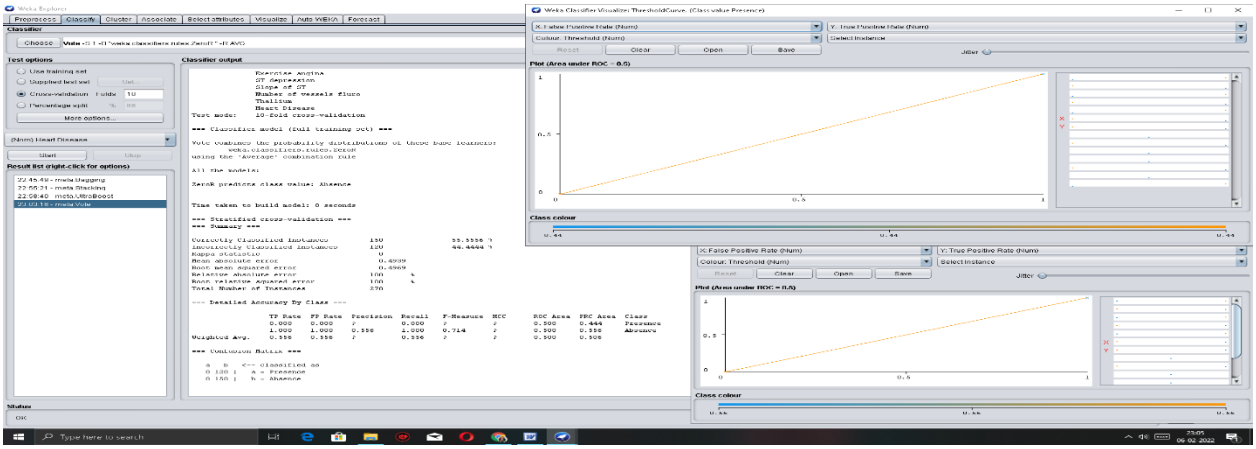

Fig.4 Vote Classifier with Visualize curve

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11148

Experiments and Observations-5

NaiveBayes- Class for a Naive Bayes classifier using estimator classes.

Classifier Output

=== Run information ===

Scheme: weka.classifiers.bayes.NaiveBayes

Relation: Heart_Disease_Prediction-weka.filters.unsupervised.attribute.Reorder-R1,2,3,4,5,6,7,8,9,10,11,12,13,14-weka.filters.AllFilter-weka.filters.MultiFilter-Fweka.filters.AllFilter

Instances: 270 Attributes: 14

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

Naive Bayes Classifier Class

Attribute Presence Absence

(0.44) (0.56)

=== Stratified cross-validation ====== Summary ===

Correctly Classified Instances	228	84.4444 %
Incorrectly Classified Instances	42	15.5556 %
Kappa statistic	0.6834	
Mean absolute error	0.1814	
Root mean squared error	0.356	
Relative absolute error	36.7302 %	
Root relative squared error	71.6335 %	
Total Number of Instances	270	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.800	0.120	0.842	0.800	0.821	0.684	0.901	0.875	Presence
	0.880	0.200	0.846	0.880	0.863	0.684	0.901	0.917	Absence
Weighted Avg.	0.844	0.164	0.844	0.844	0.844	0.684	0.901	0.898	

=== Confusion Matrix ===

a	b	<-- classified as
96	24	a = Presence
18	132	b = Absence

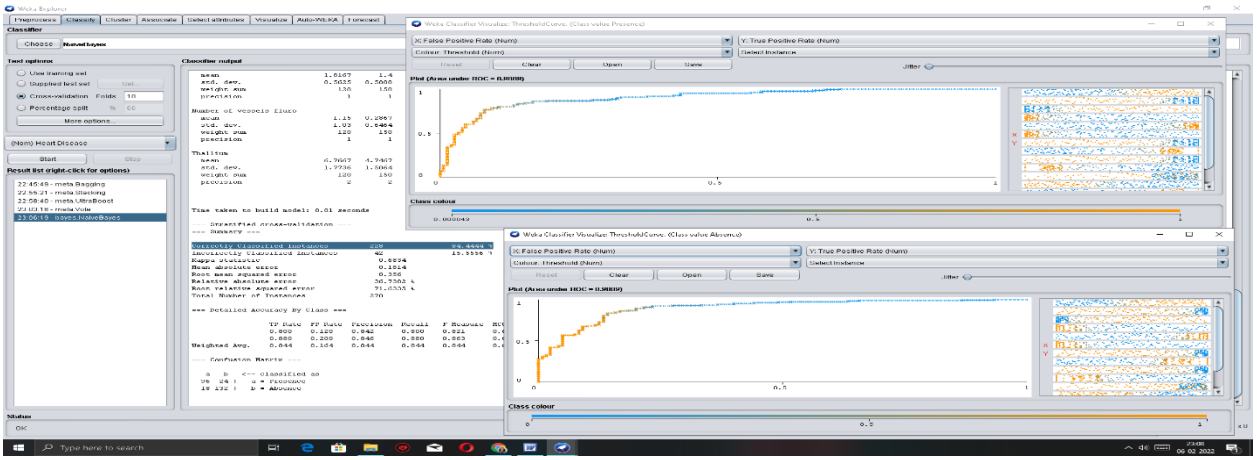

Fig.5 NaiveBayes Classifier with Visualize curve

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 1, January 2022

DOI: 10.17148/IJARCCE.2022.11148

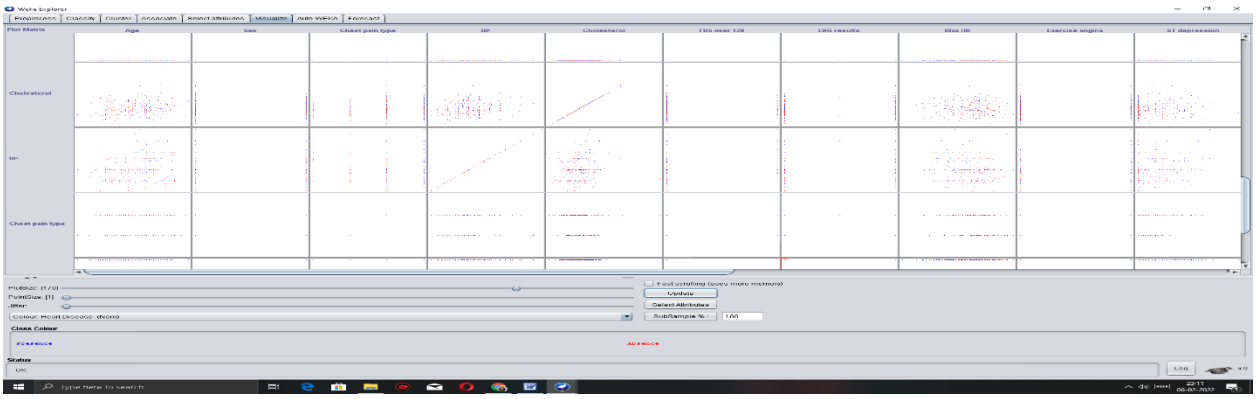

Fig.6 Visualize Plot Matrix for all Attributes

CONCLUSION

We have taken 5 different machine learning classifier algorithms and with the observations to decide the acceptability of a particular domain in the machine learning model. In the study of the above real time medical dataset implementation and in different observations, it is found that the accuracy level using the machine learning classification model Bagging is very much satisfactory, having good accuracy rate of 89.62 % and so will be a good option in the field of medical sciences to predict early diagnosis of Cardio Vascular Disease. As we have taken 5 different experimental observations using the machine learning tool to clearly analyze, detect and predict for the Cardio Vascular Disease. In the study of the above experimental observations, it is found that, machine learning tools are no doubt an excellent way to predict and detect the Cardio Vascular disease (Heart) at an early stage prior to the satisfiability of the conditions of the early stage patient. It is found that the accuracy level using different algorithm in Machine Learning is an excellent option for detection and prediction of Cardio Vascular disease, having good accuracy rate and so will be efficient and acceptable.

REFERENCES

[1] Jha, V., Garcia-Garcia, G., Iseki, K., Li, Z., Naicker, S., Plattner, B. & Yang, C. W. (2013). Chronic kidney disease: global dimension and perspectives. The Lancet, 382(9888), 260-272.

[2] Ali, S., Dave, N., Virani, S. S., & Navaneethan, S. D. (2019). Primary and secondary prevention of cardiovascular disease in patients with chronic kidney disease. Current Atherosclerosis Reports, 21(9), 1-9.

[3] Levey, A. S., Coresh, J., Bolton, K., Culleton, B., Harvey, K. S., Ikizler, T. A. & Briggs, J. (2002). K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. American Journal of Kidney Diseases, 39(2 SUPPL. 1), i-ii+.

[4] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260.

[5] Gunarathne, W. H. S. D., Perera, K. D. M., & Kahandawaarachchi, K. A. D. C. P. (2017, October). Performance evaluation on machine learning classification techniques for disease classification and forecasting through data analytics for chronic kidney disease (CKD). In 2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE) (pp. 291-296). IEEE.

[6] Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining (pp. 785-794).

[7] Noble, W. S. (2006). What is a support vector machine?. Nature Biotechnology, 24(12), 1565-1567. [8] Pregibon, D. (1981). Logistic regression diagnostics. Annals of Statistics, 9(4), 705-724.

[9] Lakshmi, K. R., Nagesh, Y., & Krishna, M. V. (2014). Performance comparison of three data mining techniques for predicting kidney dialysis survivability. International Journal of Advances in Engineering & Technology, 7(1), 242.

[10] Vijayarani, S., Dhayanand, S., & Phil, M. (2015). Kidney disease prediction using SVM and ANN algorithms. International Journal of Computing and Business Research (IJCBR), 6(2), 1-12. [11] Baby, P. S., & Vital, T. P. (2015). Statistical analysis and predicting kidney diseases using machine learning algorithms. International Journal of Engineering Research and Technology, 4(7), 206–210.