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

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 12, December 2021

DOI: 10.17148/IJARCCE.2021.101202

Predictive Analysis of Coronary Heart Disease (CHD) based on Machine Learning Classification Algorithm

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

[1]Lecturer, Department of MCA, School of Computer Science, Gangadhar Meher University (GMU), Odisha, India

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

*Corresponding Author: Dr. Vijay Pal Singh

Abstract*:* Coronary Heart Disease is a major cause of death in worldwide. It is also called ischemic heart disease or coronary artery disease. As per as the WHO statistical report is concern, 17.9 million people die every year from CVDs, an estimated 32% of all deaths worldwide, and patients die mainly because of inappropriate and non-affordable treatment. Heart disease is a worldwide health crisis in the present scenario. This disease can be curable with early diagnosis and proper treatment. The purpose of this paper is to establish some predictive models using Machine Learning algorithms by taking a real time CHD dataset. In this paper, we have shown some real-time experiments and observations with the help of some Machine Learning algorithms, and also shown a clear picture on the predictive analysis on medical diagnosis of the Coronary Heart Disease (CHD) using Machine Learning algorithms using which patients may get accurate data so as to diagnose better for their early treatment.

Keywords: Algorithm, Classifier, Coronary Heart Disease, Machine Learning, Prediction.

I.INTRODUCTION

According to definition in medical science, a disease in which there is a narrowing or blockage of the coronary arteries (blood vessels that carry blood and oxygen to the heart)[1]. An algorithm is a set of instruction for solving a problem or accomplishing a task. Every computerized device uses algorithm to perform its function. Coronary heart disease is a major cause of death in worldwide. It is also called ischemic heart disease or coronary artery disease[2].

II.LITERATURE SURVEY

Coronary heart disease is the usual cause is of a blockade or a buildup of plaque or buildup of cholesterol or fatty substances in the walls that narrows the arteries, limiting blood flow to the heart. Over the time, the walls of arteries become full of cholesterol deposits. This process are called as atherosclerosis and the cholesterol deposits are called atheroma. It can also exacerbate the other condition such as blood pressure and diabetes. A build up of fatty plaques in arteries is the most common cause of coronary heart disease. Unhealthy life style habits, such as poor diet, lack of exercises, being over weight, and smoking are the main causes for aggravating the condition of coronary heart disease[3][4].

Symptoms that coronary heart disease carries are- Frequent chest pain, Feeling of breathless, Pain in overall body, Indigestion and Nausea, Light headedness or sweating, Palpitation etc. The symptoms differ person to person. Some may not have any symptoms. Diagnosis of coronary heart disease can be done by verbal questioning regarding family history, or by tests such as MRI, CT scan, Treadmill test etc[5][6].

Coronary heart disease is a damage or disease in the hearts in major blood vessels.

Risk factor includes-

Ageing, male person are generally at higher risk, however risk for woman increases after menopause. Family history-family history of heart disease associated with a higher risk of CAD.

Smoking, High blood pressure, High blood cholesterol level, Diabetes, Overweight or obesity, Physical inactivity, High stress, Unbalanced diet, High triglycerides, Alcohol use, etc[7].

Treatment:-

There is no cure for CAD but we can manage the condition. Treatment involves- quitting smoking, healthy diet, regular exercise and medication [8]. Medication that reduce the risk includes – Calcium channel blockers to widen the coronary arteries, improving blood flow to the heart and reducing hyper tension. Surgery – surgical procedures can open or replace blocked arteries if they have become very narrow [9].

Prevention:-

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

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 12, December 2021

DOI: 10.17148/IJARCCE.2021.101202

Controlling blood cholesterol levels can help reduce a person's risk of CAD. For prevention it includes – Avoiding tobacco, diet with less sugar, salt and fat, avoid alcohol intake.
A person's risk of developing CAD is increase by these factors – High Blood Pressure or Hypertension, Diabetes, smoking which increases inflammation [10].

III.EXPERIMENTS AND OBSERVATIONS-1

Classifier Output

We have taken Weka as a Machine Learning tool to predict and classify the disease based on the real-time Heart patient dataset(in Attribute Relation File - .arff Format).

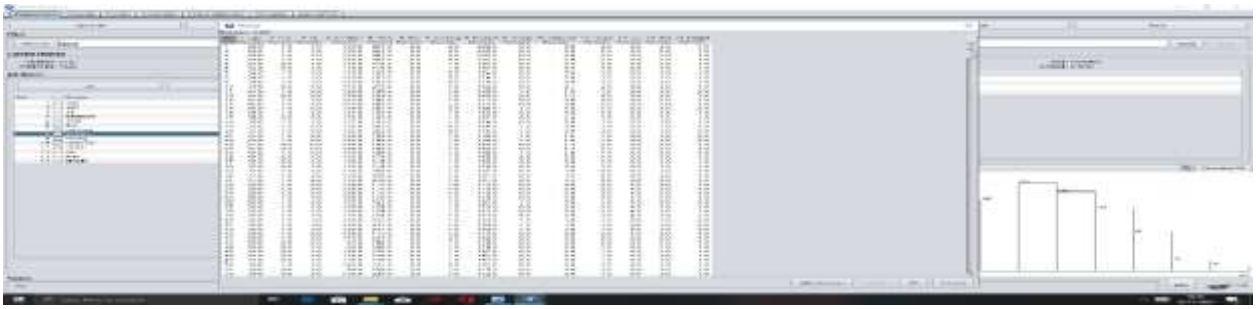

Fig.1 Coronary Heart Disease Dataset (in .arff File Format)

=== Run information ===Scheme: weka.classifiers.rules.ZeroR
Relation: CHD Instances: 1025 Attributes: 14
Age sex cp trestbps chol fbs restecg thalach exang
Oldpeak slope ca thal target
Test mode: 10-fold cross-validation
=== Classifier model (full training set) ===
ZeroR predicts class value: 0.5131707317073171
Time taken to build model: 0 seconds
=== Cross-validation ====== Summary ===
Correlation coefficient -0.0773
Mean absolute error 0.5
Root mean squared error 0.5002
Relative absolute error 100 %
Root relative squared error 100 %
Total Number of Instances 1025

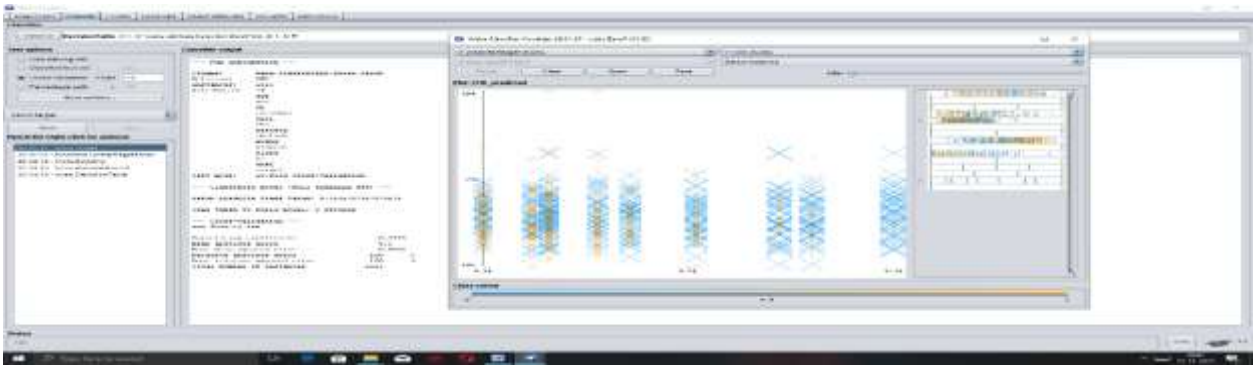

Fig.2 ZeroR Classifier Output with Visualize Error

IV.EXPERIMENTS AND OBSERVATIONS-2

Classifier Output

=== Run information ===Scheme: weka.classifiers.functions.LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4
Test mode: 10-fold cross-validation=== Classifier model (full training set) ===

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

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 12, December 2021

DOI: 10.17148/IJARCCE.2021.101202

Linear Regression Model taken to build model: 0.08 seconds
=== Cross-validation ====== Summary ===
Correlation coefficient 0.7084
Mean absolute error 0.2853
Root mean squared error 0.3529
Relative absolute error 57.0633 %
Root relative squared error 70.5457 %
Total Number of Instances 1025

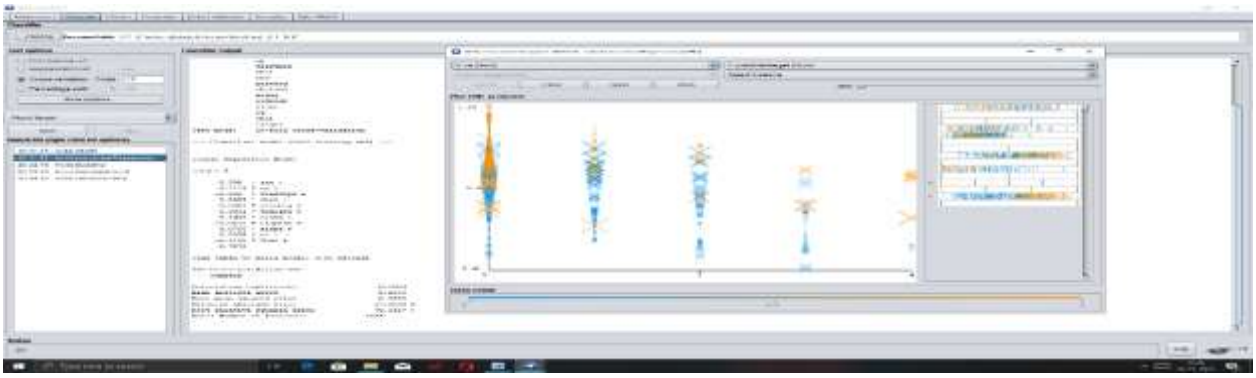

Fig.3 LinearRegression Classifier Output with Visualize Error

V.EXPERIMENTS AND OBSERVATIONS-3

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: CHD Instances: 1025 Attributes: 14
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: 0.13 seconds
=== Cross-validation ====== Summary ===
Correlation coefficient 0.9222
Mean absolute error 0.1219
Root mean squared error 0.2
Relative absolute error 24.3761 %
Root relative squared error 39.9806 %
Total Number of Instances 1025

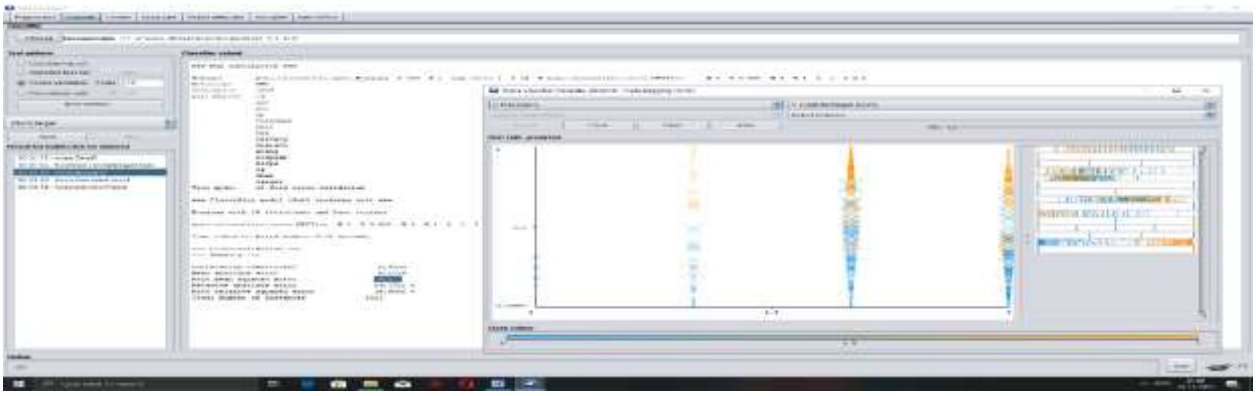

Fig.4 Bagging Classifier Output with Visualize Error

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

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 12, December 2021

DOI: 10.17148/IJARCCE.2021.101202

VI.EXPERIMENTS AND OBSERVATIONS-4

Classifier Output

=== Run information ===Scheme: weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Relation: CHD Instances: 1025 Attributes: 14
Test mode: 10-fold cross-validation
=== Classifier model (full training set) ===RandomForest
Bagging with 100 iterations and base learner
weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities
Time taken to build model: 0.25 seconds
=== Cross-validation ====== Summary ===

Correlation coefficient	0.9927
Mean absolute error	0.0375
Root mean squared error	0.0675
Relative absolute error	7.5066 %
Root relative squared error	13.4894 %
Total Number of Instances	1025

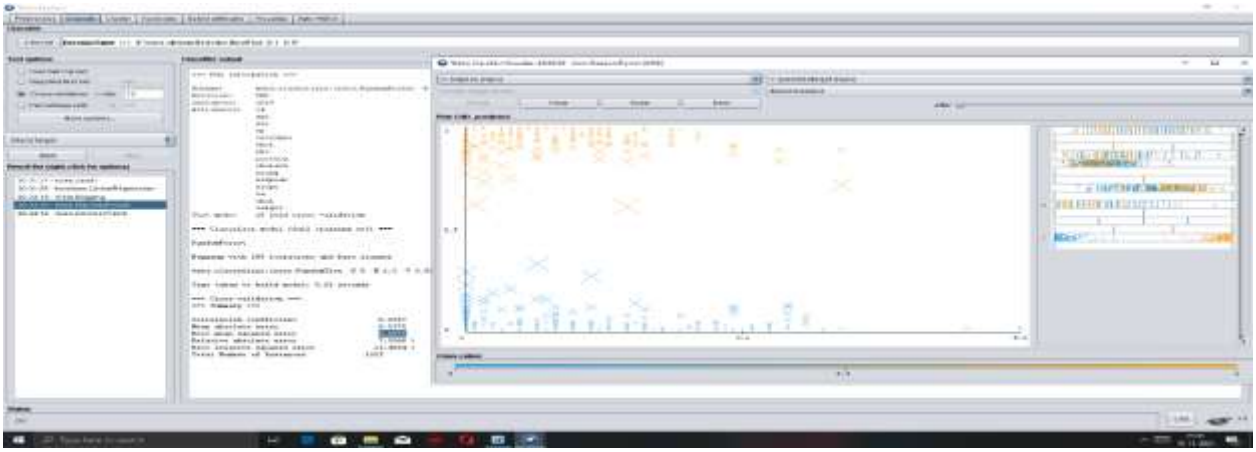

Fig.5 RandomForest Classifier Output with Visualize Error

VII.EXPERIMENTS AND OBSERVATIONS-5

Classifier Output

=== Run information ===Scheme: weka.classifiers.rules.DecisionTable -X 1 -S "weka.attributeSelection.BestFirst -D 1 -N 5"
Relation: CHD Instances: 1025 Attributes: 14
Test mode: 10-fold cross-validation
== Classifier model (full training set) ===Decision Table:
Number of training instances: 1025 Number of Rules : 293
Non matches covered by Majority class.
Best first. Start set: no attributes Search direction: forward
Stale search after 5 node expansions Total number of subsets evaluated: 90
Merit of best subset found: 0
Evaluation (for feature selection): CV (leave one out)
Feature set: 1,3,4,7,8,12,13,14 Time taken to build model: 0.08 seconds
=== Cross-validation ====== Summary ===

Correlation coefficient	0.9986
Mean absolute error	0.0014
Root mean squared error	0.0265
Relative absolute error	0.287 %
Root relative squared error	5.3025 %
Total Number of Instances	1025

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

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 12, December 2021

DOI: 10.17148/IJARCCE.2021.101202

Fig.5 Visualize all Instance

VIII.DISCUSSION

Coronary heart disease continues to take a tool on the number of lives lost and on the quality of life for survivors of cardiac events. Coronary heart disease is not curable but treatment can help to alleviate the condition such as good lifestyle, surgery etc.

IX.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 and so will be a good option in the field of medical sciences to predict early diagnosis of Coronary Heart Diseases.

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.