Heart Disease Prediction Using XAI

1st Shimul Mandal Tamo
Department of Computer Science and Engineering
BRAC University
66 Mohakhali, Dhaka 1212, Bangladesh
shimul.mondol.tamo@g.bracu.ac.bd

2nd Md. Fazle Rabbi Linkon
Department of Computer Science and Engineering
BRAC University
66 Mohakhali, Dhaka 1212, Bangladesh
fazle.rabbi.linkon@g.bracu.ac.bd

3rd Sharia Hoque
Department of Computer Science and Engineering
BRAC University
66 Mohakhali, Dhaka 1212, Bangladesh
shahria.hoque@g.bracu.ac.bd

4th Annajiat Alim Rasel
Senior Lecturer
Department of Computer Science and Engineering
BRAC University
66 Mohakhali, Dhaka 1212, Bangladesh
annajiat@bracu.ac.bd

***Abstract*—Machine learning is a way of manipulating and extracting implicit, previously unknown/known and potential useful information about the data. Machine learning Includes various classifiers Supervised, unsupervised, and ensemble learning that are used to predict and search for accuracy given data file. We can use this knowledge in our heard diseases prediction project. Cardiovascular diseases are very common nowadays, they describe a number of conditions that could affect them our heart The World Health Organization estimates that 17.9 million global deaths from CVD(cardiovascular disease). It is the leading cause of death in adults. Our project can help predict who is likely to diagnose with heart disease using their medical history. It recognizes who all has some symptoms heart disease, such as chest pain or high blood pressure, and can help diagnose disease with less medical means tests and effective treatments so that they can be cured accordingly.**

***Index Terms*—machine learning, heart diseases, KNN, Random forest, Logical Rigression**

I. Introduction

Our work mainly focuses on three data mining techniques, namely: (1) Logistic Regression, (2) KNN and (3) Random Forest Classifier. The accuracy of our project is 87.5%, which is better than the previous one system where only one data mining technique is used. So using more data mining techniques has increased HDPS accuracy and efficiency. Logistic regression falls under the category of supervised learning. Only discrete values are used in logistic regression. The aim of this project is to verify whether a patient is likely to be diagnosed with one cardiovascular heart disease based on their medical attributes such as gender, age, chest pain, fasting sugar level etc. A data-set with patient history and attributes is selected from the UCI repository. According to we use this data set to predict whether a patient may have heart disease or not. To predict this, we use 12 medical characteristics of the patient and classify it if the patient is likely to have heart disease. These medical attributes are trained according to three algorithms: **Logistic Regression**, **KNN** and **Random Forest Classifier**. Most the efficient one among these algorithms is KNN, which gives us an accuracy of 90.63%. And finally we classify patients who are at risk of heart disease or not and also this method is completely cost effective.

II. Literature Review

Lots of work has been carried out to predict heart disease using UCI Machine Learning data-set. Different levels of accuracy have been attained using various data mining techniques which are explained as follows. Avinash Golande studies various different ML algorithms that can be used for classification of heart disease. Research was carried out to study Decision Tree, KNN and K-Means algorithms that can be used for classification and their accuracy were compared[1]. This research concludes that accuracy obtained by Decision Tree was highest further it was inferred that it can be made efficient by combination of different techniques and parameter tuning. T.Nagamani have proposed a system [2] which deployed data mining techniques along with the MapReduce algorithm. The accuracy obtained according to this paper for the 45 instances of testing set, was greater than the accuracy obtained using conventional fuzzy artificial neural network. Here, the accuracy of algorithm used was improved due to use of dynamic schema and linear scaling. Fahd Saleh Alotaibi has designed a ML model comparing five different algorithms [3]. Rapid Miner tool was used which resulted in higher accuracy compared to Matlab and Weka tool. In this research the accuracy of Decision Tree, Logistic Regression, Random forest, Naive Bayes and SVM classification algorithms were compared. Decision tree algorithm had the highest accuracy. Anjan Nikhil Repaka, proposed a system in [4] that uses NB (Naïve Bayesian) techniques for classification of dataset and AES (Advanced Encryption Standard) algorithm for secure data transfer for prediction of disease. Theresa Princy. R, executed a survey including different classification algorithm used for predicting heart disease. The classification techniques used were Naive Bayes, KNN (K-Nearest Neighbour), Decision tree, Neural network and accuracy of the classifiers was analyzed for different number of attributes [5]. Nagaraj M Lutimath, has performed the heart disease prediction using Naive

bayes classification and SVM (Support Vector Machine). The performance measures used in analysis are Mean Absolute Error, Sum of Squared Error and Root Mean Squared Error, it is established that SVM was emerged as superior algorithm in terms of accuracy over Naive Bayes [6]. The main idea behind the proposed system after reviewing the above papers was to create a heart disease prediction system based on the inputs as shown in Table 1. We analysed the classification algorithms namely Decision Tree, Random Forest, Logistic Regression and Naive Bayes based on their Accuracy, Precision, Recall and f-measure scores and identified the best classification algorithm which can be used in the heart disease prediction.

III. METHODOLOGY

REFERENCES

[1] Avinash Golande, Pavan Kumar T, "Heart Disease Prediction Using Effective Machine Learning Techniques", International Journal of Recent Technology and Engineering, Vol 8, pp.944-950,2019.

[2] T.Nagamani, S.Logeswari, B.Gomathy," Heart Disease Prediction using Data Mining with Mapreduce Algorithm", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-3, January 2019.

[3] Fahd Saleh Alotaibi," Implementation of Machine Learning Model to Predict Heart Failure Disease", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 6, 2019.

[4] Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin, "Design And Implementation Heart Disease Prediction Using Naives Bayesian", International Conference on Trends in Electronics and Information(ICOEI 2019).

[5] Theresa Princy R,J. Thomas,'Human heart Disease Prediction System using Data Mining Techniques', International Conference on Circuit Power and Computing Technologies,Bangalore,2016.

[6] Nagaraj M Lutimath,Chethan C,Basavaraj S Pol.,'Prediction Of Heart Disease using Machine Learning', International journal Of Recent Technology and Engineering,8,(2S10), pp 474-477, 2019.