International Journal of Scientific Research and Engineering Development-– Volume 6 Issue 5, Sept-Oct 2023

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Prediction of Heart Disease Using ML Models

Almanshara Latif Shaikh*

*(*UG student, Department of Information Technology, B.K.Birla College, Kalyan*

Email: almnashras@gmail.com)

Abstract:

Heart disease is a common type of disease. That includes many types of heart problems. Heart disease is also called cardiovascular Disease (CVD). Cardiovascular disease is a disorder of the heart or blood vessels. Heart disease is the leading cause of death Universally, taking approximately 17.9 million lives each year according to the World Health Organization (WHO). To stave off Heart disease there are various machine learning algorithms to predict Heart disease. In this paper, XG-Boost, Random forest and Extra Trees classifier is used. Among this, all three algorithms are used using Bayesian optimization to find the best model. The Heart Disease dataset which is the Heart Disease Prediction Dataset collected from Kaggle. The Overall research is referred from (1) and future work of this research is completed. In that paper XG-Boost model gave more accuracy and in this research RF model gaveMore accuracy.

Keywords —Heart Disease, XG-Boost(Extreme Gradient Boosting), RF, ET.

I. INTRODUCTION

Heart disease is a big problem or challenge for Healthcare Organizations or the area of clinical data analysis. There are several number of death cases related to heart and their counting increasing fast day by day. Common symptoms/ reasons of heart disease are high blood pressure, unhealthy diet, high cholesterol, diabetes, air pollution, obesity, tobacco use, kidney disease, physical inactivity, harmful use of alcohol, and stress. To solve all problems or to predict the disease the ML model is used. In this research trying to predict a heart Disease using XG-Boost (Extreme Gradient Boosting)Classifier, RF (Random Forest) model classifier , ET(Extra Tree) model classifier .XG-Boost is an improved distributed Gradient boosting library designed for efficient and scalable Training of machine learning models. It is an ensemble Learning method that combines the predictions of multiple Weak models to produce a stronger prediction. In this research,Bayesian Optimization technique is used to XG-Boost hyper-Parameter tuning to improve the prediction results. After the Final proposed model compared with other machine learning Models, the RF(Random Forest) model gives more accuracy Among the both models namely Proposed model and Extra Tree. The results calculated that the RF model shows better Performance among two models.

II. LITERATURE REVIEW

Previous studies for predicting heart disease using machine learning algorithms.

In this paper, (Budholiya et al., 2022) heart disease predict using ml models such as XG-Booster,RF, and ET model using Bayesian Optimization, and the XG-Booster modelsgivesmore accuracy which is 91%.

In this paper, (Sharma et al., 2020)heart disease predict using ML algorithm RF, SVM, Naive Bayes, and DT for the development model. The SVM and RF models give more accuracy.The RF gives 99% SVM model gives 98% accuracy.

In this paper,(Bhuyan & Mishra, 2022)heart disease predict using ML algorithms DT, KNN, LGBN, SCV,LR,RF, andNB. The KNN algorithm is gives more accuracy compared to all other algorithms. The accuracy rating is 85.71 %

III. METHODS

A. *Description About Dataset & Software*

In this research, the Heart Disease Prediction dataset is used. The experiments were carried out on the Heart Disease Prediction dataset from the Kaggle. Attributes Information. There are 4 attributespresent in this dataset such Age, BP, cholesterol, and Heart disease. In this paper, the software tool used is the jupyter notebook the XGBOOSTER model and Random Forest, and Extra Tree all are implemented in Jupyter Notebook and the language used in Python. Table 1 represents the attribute of Heart Disease Prediction dataset.

Table 1
Features of heart disease dataset

Sr. No	Feature Description	Type
1	Age	Numerical
2	BP	Numerical
3	Cholesterol	Numerical
4	HeartDisease	Numerical

B. *Train and test*

In this research,the training set contains 70% of the data whereas the test set contains 30% of the data.

C. *Evaluation Metrics*

In this research the efficiency of the proposed method used 4 evaluation metrics such as accuracy, specificity, F1-Score, and area under the curve of the ROC chart

.

D. Experiment Results of XGBOOSTER (Proposed Model)

In this research, the XG-Boost machine learning algorithm as a classifier is used for training and testing. In this experiment, the hyper parameter method is used for the XG-Boost algorithm. And it's maximized by the Bayesian Optimization algorithm and using this optimization hyperparameter performance analysis. Using the Bayesian Optimization algorithm iterations, getting obtain an optimal set of parameters. In this experiment, 25 parameters are used. The Fig.1 shows the processed hyper-parameters in the hyper-parameter optimization stage. The proposed model evaluation shows in the Table 2.

Table 2
Performance of XG-Booster model

Data	Accuracy	Specificity	F1Score
Train	53	100	53
Test	56	100	60

Fig. 1simulation result of XG-Boost model

Performing hyperparameter tuning using Bayesian optimization...

iter	target	colsam...	learni...	max_depth	subsample
1	0.5926	0.4753	0.363	3.001	0.3721
2	0.5926	0.2321	0.05525	4.118	0.411
3	0.5926	0.4571	0.274	5.515	0.7167
4	0.5926	0.284	0.4403	3.164	0.7034
5	0.5926	0.4756	0.2838	3.842	0.2783
6	0.5926	0.9891	0.4356	8.988	0.2043
7	0.5926	0.4983	0.2126	6.287	0.7698
8	0.5926	0.661	0.1194	8.755	0.8242
9	0.5926	0.9767	0.434	8.997	0.8428
10	0.5926	0.9073	0.04513	8.999	0.313
11	0.5926	0.9342	0.06588	3.003	0.1494
12	0.5926	0.9831	0.1425	7.853	0.5019
13	0.5926	0.3924	0.3215	5.403	0.7764
14	0.5926	0.3753	0.3416	6.308	0.3843
15	0.5926	0.4575	0.1051	3.317	0.4105
16	0.5926	0.8648	0.3552	8.999	0.9876
17	0.5926	0.1283	0.2464	3.001	0.2255
18	0.5926	0.1174	0.03767	8.989	0.9834
19	0.5926	0.2061	0.2798	5.949	0.2314
20	0.5926	0.3364	0.0207	4.005	0.6775
21	0.5926	0.2765	0.2062	6.092	0.1631
22	0.5926	0.2627	0.1432	7.821	0.6832
23	0.5926	0.2203	0.2588	8.335	0.6104
24	0.5926	0.2365	0.1101	7.035	0.8873
25	0.5926	0.9183	0.06345	3.017	0.9638

Fig2 Simulation result of RF model

iter	target	max_fe...	max_sa...	n_esti...
1	0.5508	0.668	0.6104	153.4
2	0.5508	0.6829	0.6636	186.1
3	0.5508	0.7406	0.9496	197.5
4	0.5508	0.8405	0.7086	166.8
5	0.5508	0.7236	0.7763	147.8
6	0.5508	0.6024	0.6906	189.8
7	0.5508	0.9502	0.5185	139.2
8	0.5508	0.6197	0.9363	191.6
9	0.5508	0.5776	0.7049	195.1
10	0.5508	0.9604	0.776	141.2
11	0.5508	0.8022	0.5833	120.3
12	0.5508	0.6318	0.841	193.5
13	0.5508	0.8559	0.8085	133.3
14	0.5508	0.8162	0.8799	163.8
15	0.5508	0.9871	0.9284	180.5
16	0.5508	0.507	0.5573	103.9
17	0.5508	0.8013	0.8372	194.3
18	0.5508	0.8933	0.8356	121.2
19	0.5508	0.7647	0.943	125.5
20	0.5508	0.695	0.6327	100.9
21	0.5508	0.6684	0.5568	100.0
22	0.5508	0.8838	0.8184	100.0
23	0.5508	0.6167	0.5306	100.0
24	0.5508	0.5718	0.8842	200.0
25	0.5508	0.5823	0.5447	100.0

Fig3 simulation of ET model

iter	target	max_fe...	max_sa...	n_esti...
1	0.5508	0.668	0.6104	153.4
2	0.5508	0.6829	0.6636	186.1
3	0.5508	0.7406	0.9496	197.5
4	0.5508	0.8405	0.7086	166.8
5	0.5508	0.7236	0.7763	147.8
6	0.5508	0.6024	0.6906	189.8
7	0.5508	0.9502	0.5185	139.2
8	0.5508	0.6197	0.9363	191.6
9	0.5508	0.5776	0.7049	195.1
10	0.5508	0.9604	0.776	141.2
11	0.5508	0.8022	0.5833	120.3
12	0.5508	0.6318	0.841	193.5
13	0.5508	0.8559	0.8085	133.3
14	0.5508	0.8162	0.8799	163.8
15	0.5508	0.9871	0.9284	180.5
16	0.5508	0.507	0.5573	103.9
17	0.5508	0.8013	0.8372	194.3
18	0.5508	0.8933	0.8356	121.2
19	0.5508	0.7647	0.943	125.5
20	0.5508	0.695	0.6327	100.9
21	0.5508	0.6684	0.5568	100.0
22	0.5508	0.8838	0.8184	100.0
23	0.5508	0.6167	0.5306	100.0
24	0.5508	0.5718	0.8842	200.0
25	0.5508	0.5823	0.5447	100.0

E. *Co*mparison of XG-Booster with Random Forest , and Extra Tree

In this research, the RF and ET models are compared with theXG-Boost model. In both the RF and ET models, the same methodis applied. That the hyper parameter method is use for RF and ET algorithm, and it's maximized by the Bayesian Optimization algorithm and using this optimization hyper parameter performance analysis. Using the Bayesian Optimization iterations, getting obtain an optimal set of parameters. In this experiment, 25 sets of parameterswere used in both RF and ET algorithms. Figure 2 shows the simulation result of the RF model and the Fig.3 shows the simulation result of the ET model. All 3 AUC(ROC) graphs of all models are shown in Fig. 4, Fig.5, Fig.6 respectively.

Fig4. ROC of XG-Boost

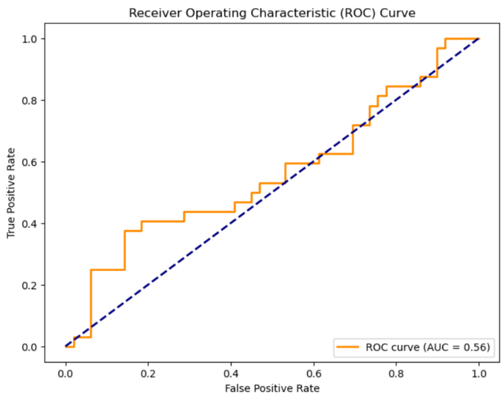


Fig5.ROC chart of RF

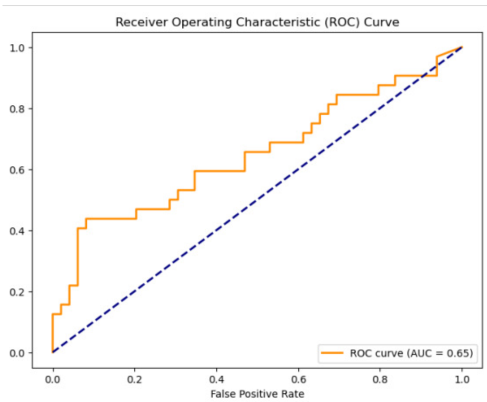
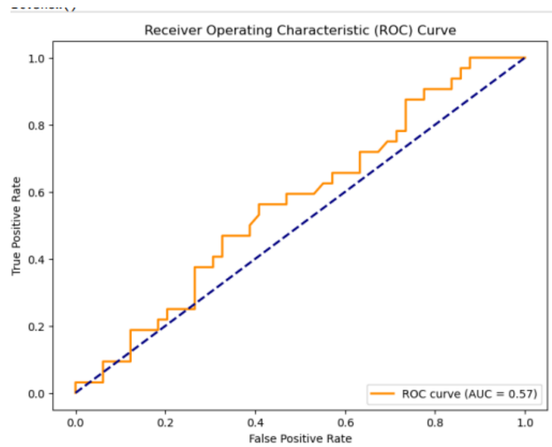


Fig6.ROC chart of ET model



IV. Result

When compared with this experiment in (1) the XG-Boost model compared with the other two models such as RF and ET and applied to theCleveland heart disease dataset, and XG-Booster gave more accuracy compared to RF and ET in that research. In this research, the XG-Boost model and the other two models such as RF and ET applied to the dataset which name is the Heart Disease Dataset. In that experiment same XG-Booster model was compared with the other two models such as RF and ET but the RF model gives more accuracy compared to the XG-Boost and ET models. The accuracy model is shown in Table 3.

Table3.

Model	Accuracy
RF	65
XG-Booster	56
ET	57

V. Conclusion and Future Work

In this research,there are 4 efficiency metrics are used such as accuracy, Specificity, F1Score, and AUC (ROC).According to all the experiments the RF model gives more accuracy compared to XG-Boost and ET models. On the other hand, there is a chance that this models accuracy can be improved.

Reference

1. Bhuyan, S. S., & Mishra, A. K. (2022). Machine Learning Algorithms for Heart Disease Prediction. Asian Journal of Convergence in Technology, 8(1), 87. https://doi.org/10.33130/AJCT.2022v08i01.013

2. Budholiya, K., Shrivastava, S. K., & Sharma, V. (2022). An optimized XGBoost based diagnostic system for effective prediction of heart disease. Journal of King Saud University - Computer and Information

Sciences, 34(7), 4514–4523. https://doi.org/10.1016/j.jksuci.2020.10.013

3. Prediction of Heart Disease using Machine Learning. (2019). International Journal of Recent Technology and Engineering, 8(2S10), 474–477. https://doi.org/10.35940/ijrte.B1081.0982S1019

4. Sharma, V., Yadav, S., & Gupta, M. (2020). Heart Disease Prediction using Machine Learning Techniques. 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), 177–181. https://doi.org/10.1109/ICACCCN51052.2020.9362842

5. Prediction of Heart Disease using Machine Learning. (2019). International Journal of Recent Technology and Engineering, 8(2S10), 474–477. https://doi.org/10.35940/ijrte.B1081.0982S1019

6. Salhi, D. E., Tari, A., & Kechadi, M.-T. (2021). Using Machine Learning for Heart Disease Prediction. In M. R. Senouci, M. E. Y. Boudaren, F. Sebbak, & M. Mataoui (Eds.), Advances in Computing Systems and Applications (Vol. 199, pp. 70–81). Springer International Publishing. https://doi.org/10.1007/978-3-030-69418-0_7