GC-308 Heart Disease Prediction and Analysis by Machine Learning

Abstract

Heart disease is one of the primary sources of death in the world. There are 2,380 deaths from heart disease each day, based on 2018 data. Also, heart disease causes the highest number of deaths globally, with approximately 18 million people dying yearly. This prediction can be made using Machine Learning techniques. With machine learning. Combining a prediction model with machine learning correctly classified results for heart disease with classification Zero r, Grip (FURIA) classification, Decision tree J48, Classification MLP, MLR (multinomial logistic regression), Bagging, Boosting, stacking classification. J48 has the highest accuracy. As it is described that early detection of heart disease plays vital role in saving individuals life. Considering the classification method in machine learning is chosen one for diagnosing heart disease and hence there are good outcomes that are came out. This research was based on 2020 survey conducted by CDC on BRFSS. based on these 8 different machine learning methods we have chosen.

Introduction

The heart disease is one of the primary sources of death in the world. Heart disease causes the highest deaths globally, with approximately 18 million people dying yearly. Around 31 percent of deaths are from heart disease, according to the WHO (World Health Organization). Recently, using machine learning techniques have shown promissing results in various domains. Healthcare is one of the domains that received huge benefits from using different ML models for early detection, patient monitoring, etc. In this project, we are exploring the role of ML techniques in heart disease detection and prediction. We used UCI and CDC heart disease datasets and applied various ML techniques.

Research Question(s)

- What part does machine learning play in forecasting heart disease?
- What differences between machine learning models created using weka?
- Which machine learning model predicts heart disease with the highest accuracy?
- What are the machine learning models used to predict heart diseases?

Materials and Methods

- The proposed research considers two main datasets containing machine learning techniques to detect Heart disease predictions. The datasets are including the UCI dataset, and the CDC data collect the last one.
- Preparing the dataset in CDC and data preprocessing
- Using machine learning models such as classification Zero r, Jrip (FURIA) classification, Decision tree J48, Classification MLP, Bagging, Boosting, and stacking classification.

Concept of machine learning

Results

TABLE I
MODEL PERFORMANCE METRICS WITH FEATURE SELECTION FOR CDC

Model	Accuracy	Precision	Recall	F-Measure
ZeroR 10-Fold	64.6623%	64.7%	64.7%	78.5%
ZeroR 15-Fold	64.6623%	64.7%	64.7%	78.5%
ZeroR 80/20	64.36%	64.4%	64.4%	64.4%
Naive Bayes 10-Fold	99.674%	99.7%	99.7%	99.7%
Naive Bayes 20-Fold	99.678%	99.7%	99.7%	99.7%
Naive Bayes 80/20	99.66%	99.7%	99.7%	99.7%
Adaboostm1 10-Fold	91.06%	91.5%	91.11%	90.8%
Adaboostm1 15-Fold	91.03%	91.4%	91.0%	90.8%
Adaboostm1 80/20	91.03%	91.4%	91.0%	90.8%
Bagging 10-Fold	99.6%	99.7%	99.7%	99.7%
Bagging 15-Fold	99.6%	99.7%	99.7%	99.7%
Bagging 80/20	99.6%	99.7%	99.7%	99.7%
Stacking 10-Fold	64.66%	64.7%	64.7%	78.5%
Stacking 15-Fold	64.66%	64.7%	64.7%	68.5%
Stacking 80/20	64.36%	64.4%	64.4%	78.3%

TABLE II
MODEL PERFORMANCE METRICS WITH ALL ATTRIBUTES FOR UCI

Model	Accuracy	Precision	Recall	F-Measure
MLP 10-Fold	77.55%	77.6%	77.6%	77.6%
MLP 15-Fold	80.19%	80.2%	80.20%	80.20%
MLP 80/20	83.60%	83.8%	83.60%	83.60%
J-48 10-Fold	78.54%	78.5%	78.5%	78.5%
J-48 15-Fold	79.8%	79.9%	79.9%	79.8%
J-48 80/20	77.04%	77.6%	77.0%	77.0%
Logistic 10-Fold	82.17%	82.3%	82.2%	82.1%
Logistic 15-Fold	82.5%	82.8%	82.5%	82.3%
Logistic 80/20	80.32%	80.5%	80.3%	80.3%
JRip 10-Fold	77.55%	77.7%	77.6%	77.4%
JRip 15-Fold	80.52%	81.0%	80.5%	80.3%
JRip 80/20	85.24%	86.3%	85.2%	85.2%
Logitboost 10-Fold	80.58%	80.8%	80.9%	80.8%
Logitboost 15-Fold	81.18%	81.2%	81.2%	81.1%
Logitboost 80/20	83.60%	84.2%	83.6%	83.6%
Bagging 10-Fold	79.53%	79.5%	79.5%	79.5%
Bagging 15-Fold	81.51%	81.5%	81.5%	81.4%
Bagging 80/20	81.96%	82.9%	82.2%	82.19%
Stacking 10-Fold	54.45%	54.5%	54.5%	54.5%
Stacking 15-Fold	54.45%	54.5%	54.5%	54.5%
Stacking 80/20	49.18%	49.2%	49.2%	49.2%

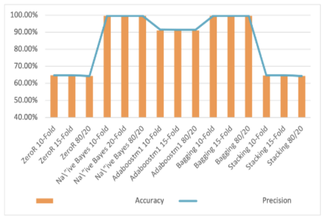

Fig 1. Accuracy of Experiments with UCI data set attributes compared to baseline.

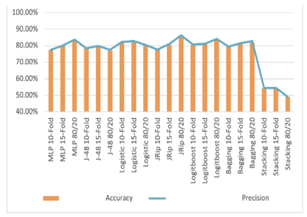

Fig 2.UCI data folding experiment

Conclusions

As the number of deaths from heart disease has increased, it has become necessary to develop a system to predict heart disease. Using the UCI machine learning repository dataset and CDC data set, this study compares the accuracy score of MLP, FURIA, Naive Bayes and C4.5 algorithms for investigating heart disease. Also, classification results for heart disease with zero r, Jrip (FURIA) classification, Decision tree J48, Classification MLP, Bagging, Boosting and stacking classification. According to these findings, better prediction results are more accurate.

Acknowledgments

Special thanks to Dr. Jack Zheng and Dr. Seyedamin Pouriyeh for their expert advice and support throughout this research.

The results will be submitted to a suitable journal.

Contact Information

Lead: Venkata Sabarinath Nekkalapu, vnekkalapu@students.Kennesaw.edu
Project CO-Ordinator, gzheng@kennesaw.edu
Project Owner, spouriye@kennesaw.edu

References

[1] Md Mamun Ali et al. “Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison”. In: *Computers in Biology and Medicine* 136 (2021), p. 104672.
[2] Vincy Cherian and MS Bindu. “Heart disease prediction using Naive Bayes algorithm and Laplace Smoothing technique”. In: *Int. J. Comput. Sci. Trends Technol* 5.2 (2017), pp. 68–73.
[3] Ibrahim M El-Hasnony et al. “Multi-label active learning-based machine learning model for heart disease prediction”. In: *Sensors* 22.3 (2022), p. 1184.
[4] Mohammed Khalid Hossen. “Heart Disease Prediction Using Machine Learning Techniques”. In: *American Journal of Computer Science and Technology* 5.3 (2022), pp. 146–154.
[5] M Nikhil Kumar, KVS Koushik, and K Deepak. “Prediction of heart diseases using data mining and machine learning algorithms and tools”. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology* 3.3 (2018), pp. 887–898.
[6] T Mythili et al. “A heart disease prediction model using SVM-Decision Trees-Logistic Regression (SDL)”. In: *International Journal of Computer Applications* 68.16 (2013).
[7] Faria Rahman, Md Mahmood, et al. “A Dynamic Approach to Identify the Most Significant Biomarkers for Heart Disease Risk Prediction Utilizing Machine Learning Techniques”. In: *International Conference on Bangabandhu and Digital Bangladesh*. Springer. 2022, pp. 12–22.
[8] S Vijayarani and S Sudha. “An Effective Classification Rule Technique for Heart Disease Prediction”. In: *International Journal of Engineering Associates* (2013).

KENNESAW STATE UNIVERSITY
COLLEGE OF COMPUTING AND SOFTWARE ENGINEERING

Authors-Venkata Sabarinath Nekkalapu, Suman Pudota, Sai Sindhu Ponugoti, Venkata Anusha Yammada, Mrinal Kapoor
Advisors(s)-Jack Zheng, Dr.Seyedamin Pouriyeh