EAI Endorsed Transactions
on Pervasive Health and Technology
Research Article EAI.EU

Heart Disease Prediction Using GridSearchCV and Random Forest

Shagufta Rasheed[1*], G Kiran Kumar[2], D Malathi Rani[3], MVV Prasad Kantipudi [4] and Anila M[5]

[1,2,5]Department of Computer Science and Engineering, Chaitanya Bharathi Institute of Technology, Gandipet, Hyderabad, Telangana State, India
[3]Department of Electronics and Communication Engineering, Marri Laxman Reddy Institute of Technology and Management, Dundigal, Hyderabad, Telangana State, India
[4]Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune 412115, Maharastra, India

Abstract

INTRODUCTION: This study explores machine learning algorithms (SVM, Adaboost, Logistic Regression, Naive Bayes, and Random Forest) for heart disease prediction, utilizing comprehensive cardiovascular and clinical data. Our research enables early detection, aiding timely interventions and preventive measures. Hyperparameter tuning via GridSearchCV enhances model accuracy, reducing heart disease's burdens. Methodology includes preprocessing, feature engineering, model training, and cross-validation. Results favor Random Forest for heart disease prediction, promising clinical applications. This work advances predictive healthcare analytics, highlighting machine learning's pivotal role. Our findings have implications for healthcare and policy, advocating efficient predictive models for early heart disease management. Advanced analytics can save lives, cut costs, and elevate care quality.
OBJECTIVES: Evaluate the models to enable early detection, timely interventions, and preventive measures.
METHODS: Utilize GridSearchCV for hyperparameter tuning to enhance model accuracy. Employ preprocessing, feature engineering, model training, and cross-validation methodologies. Evaluate the performance of SVM, Adaboost, Logistic Regression, Naive Bayes, and Random Forest algorithms.
RESULTS: The study reveals Random Forest as the favored algorithm for heart disease prediction, showing promise for clinical applications. Advanced analytics and hyperparameter tuning contribute to improved model accuracy, reducing the burden of heart disease.
CONCLUSION: The research underscores machine learning's pivotal role in predictive healthcare analytics, advocating efficient models for early heart disease management.

Keywords: AdaBoost Classifier (AB), Cross-Validation Methods, Data Preprocessing Techniques, Early Diagnosis Models, Healthcare Analytics, Logistic Regression (LR), Naïve Bayes Classifier (NB), Random Forest Algorithm (RF), Support Vector Machines (SVM)

Received on 13 December 2023, accepted on 17 March 2024, published on 22 March 2024

doi: 10.4108/eetpht.10.5523

*Corresponding author. Email: shaguftarasheed21@gmail.com

1. Introduction

Cardiovascular diseases, including heart disease, keep on being a leading reason of morbidity and death worldwide. Heart disease, a leading global cause of death, is strongly linked to risk factors like smoking, high blood pressure, and cholesterol, affecting nearly half of the US population. Machine learning plays a pivotal role in predicting cardiovascular diseases based on personal indicators, with this paper presenting six models, including Xgboost, Adaboost, Random Forest, Decision Tree, Logistic Regression, and Naïve Bayes, achieving an impressive 91.57% accuracy using the logistic regression model [14]. In recent years, cardiovascular diseases have become a leading global cause of death, driven by lifestyle changes, dietary habits, and work culture. Early detection and continuous medical monitoring can mitigate this issue, but limited resources necessitate technological solutions.