World Journal of Pharmaceutical Science and Research
www.wjpsronline.com
Research Article
ISSN: 2583-6579
SJIF Impact Factor: 5.111
Year - 2025
Volume: 4; Issue: 2
Page: 613-618

OPTIMIZING MACHINE LEARNING ALGORITHMS FOR HEART DISEASE PREDICTION

Dr. Shiksha Dubey*[1] and Dr. Ashwini Renavikar[2]

[1]Assistant Professor, Thakur Institute of Management Studies, Career Development & Research (TIMSCDR), Mumbai, India.

[2]Professor at Thakur Institute of Management Studies, Career Development & Research (TIMSCDR), Mumbai, India.

Article Received: 26 February 2025 || Article Revised: 15 March 2025 || Article Accepted: 06 April 2025

***Corresponding Author: Dr. Shiksha Dubey**
Assistant Professor, Thakur Institute of Management Studies, Career Development & Research (TIMSCDR), Mumbai, India.
DOI: https://doi.org/10.5281/zenodo.15202324

How to cite this Article: Dr. Shiksha Dubey and Dr. Ashwini Renavikar (2025). OPTIMIZING MACHINE LEARNING ALGORITHMS FOR HEART DISEASE PREDICTION. World Journal of Pharmaceutical Science and Research, 4(2), 613-618. https://doi.org/10.5281/zenodo.15202324

ABSTRACT

This study addresses the pervasive global issue of heart disease, the leading cause of death in both developed and developing nations. The complex interplay of factors contributing to heart disease involves not only individual well-being but also environmental influences. In an effort to improve predictive models, this research employs state-of-the-art machine learning (ML) techniques on a comprehensive dataset, emphasizing hyperparameter tuning as a crucial aspect for enhancing model accuracy. The research demonstrates the effectiveness of hyperparameter tuning in significantly improving the performance of various ML models. By systematically adjusting hyperparameters, the study showcases a tangible advancement in accuracy. The outcomes of this research hold promise for preventing future occurrences of heart attacks, contributing to proactive measures and interventions. The application of advanced ML techniques and strategic hyperparameter tuning emerges as a powerful approach to refining our understanding of heart disease dynamics. This research not only sheds light on the intricate factors involved but also offers a pathway toward the development of more accurate and reliable predictive models. Ultimately, these findings have the potential to influence public health strategies, fostering a proactive stance in mitigating the impact of heart diseases on global well-being.

KEYWORDS: Heart disease, machine learning, hyper-parameter tuning.

I. INTRODUCTION

The circulation in the hearts of embryos is sustained through fundamental mechanisms involving pressure development and blood ejection.[1] Heart diseases includes all abnormal condition related with the functioning of heart. As per the Centres for Disease Control and Prevention (CDC), heart disease stands as the primary cause of death in the United States. Approximately one in four deaths in the U.S. is attributed to heart disease, impacting individuals of all genders and across various racial and ethnic groups.[2] The severity of heart disease on human health is illustrated in below Fig. 1.

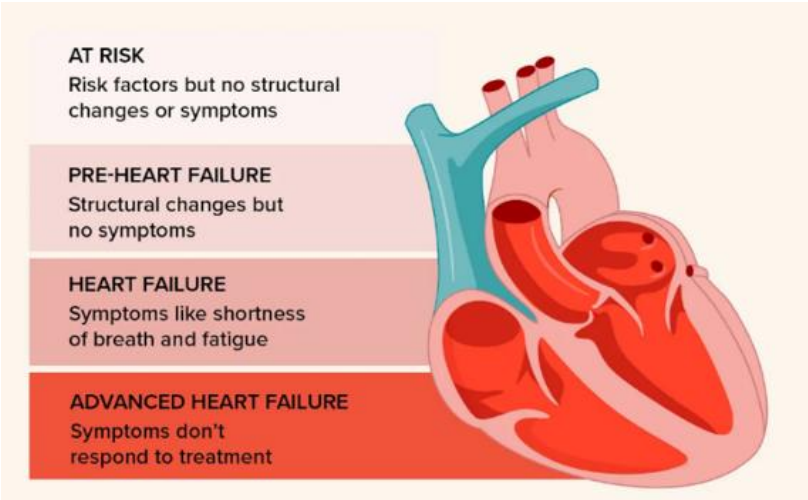

Fig. 1: - Stages of human heart attack

The different stages of heart attacks ranges from initial phase to mild heart failure and ultimately death due to heart failure. It is one of the major causes of mortality and morbidity in human population.[3] Therefore, it is utmost important to detect and prevent the occurrence of heart disease development at an initial stage.

Machine Learning (ML) has become increasingly prevalent in various domains of biomedical science, demonstrating its utility in diverse applications. One particularly impactful application lies in disease prediction, where ML's capabilities can substantially elevate the quality of healthcare for society at large. In the context of this research article, a concise set of machine learning algorithms has been effectively employed to analyze a dataset focused on predicting heart disease. The subsequent evaluation of results is conducted based on accuracy metrics, providing a robust assessment of the algorithms' performance.

Moreover, the research goes beyond mere application, introducing optimization techniques applied to the heart disease prediction dataset. These optimization methods are instrumental in refining the prediction performance of the machine learning algorithms, enhancing their accuracy and effectiveness in forecasting heart-related conditions. To provide a comprehensive overview, the article delves into a discussion of some popular ML algorithms, presenting the findings and insights in Table No. 1. This detailed exploration of machine learning applications in heart disease prediction underscores the potential of these technologies to contribute significantly to the advancement of healthcare practices and outcomes for the broader society.

Table 1: Details of ML algorithms.

Algorithm	Description	Advantages	Limitations
Logistic Regression	Linear regression-based approach	Simple interpretation, low complexity	Limited capturing of complex relationships
Random Forest	Ensemble learning using decision trees	Handles non-linearity, feature importance	May overfit, less interpretable
Gradient Boosting	Ensemble Learning Approach	Provides better performance than traditional models	Requires hyper-parameter tuning to prevent over-fitting
Neural Networks	Deep learning approach with hidden layers	Captures intricate patterns, high performance	Requires large datasets, hyperparameter tuning

II. DATASET DESCRIPTION

A secondary dataset is made available for the prediction of the probability about the occurrence of heart disease based on different feature variable representing the human health. A brief overview of the feature along with their description is illustrated in the below Table No. 2.

Table 2:- Dataset Description.

Name	*Description*
Age	Age of the patient
Sex	Sex of the patient
exang	exercise induced angina (1 = yes; 0 = no)
ca	number of major vessels (0-3)
cp	Chest Pain type chest pain type Value 1: typical angina Value 2: atypical angina Value 3: non-anginal pain Value 4: asymptomatic
trtbps	resting blood pressure (in mm Hg)
chol	cholestoral in mg/dl fetched via BMI sensor
fbs	(fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
rest_ecg	resting electrocardiographic results Value 0: normal Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
thalach	maximum heart rate achieved
target	0= less chance of heart attack 1= more chance of heart attack

An event of heart attack is caused due to multiple parameters as shown in the above table. The target variable is analyzed based on different feature variable such as age and gender. Age and gender are mainly selected due to their impact in increasing the probability of heart attach occurrence. This effect is demonstrated in the following Fig. 2.

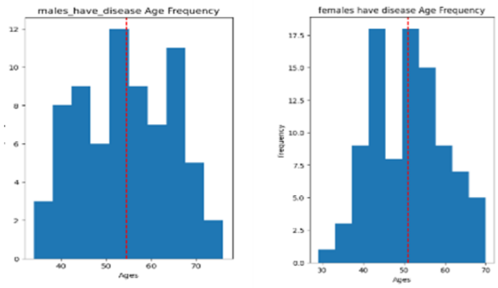

Fig. 2: Impact analysis based on age & gender.

From the above diagram it is understandable the for female the onset of the heart attack occurrence is much earlier than males and the frequency of heart attack is much higher than in males based on the sample size. The below figure Fig 3. shows the effect of only age with respect to the frequency of heart rate.

Fig. 3:- Age vs maximum heart rate.

The above diagram reflects the occurrence of heart attack majorly in the age group between 40 years to 60 years. Although these are not solely responsible for causing heart attack but these do contribute significantly for detection the probability of heart attack prediction.

III. RESULTS

The target feature variable is extracted from the given dataset and further an 80:20 split is done on it. Machine learning algorithms are applied on the given dataset and the results are shown in the below Fig. 4.

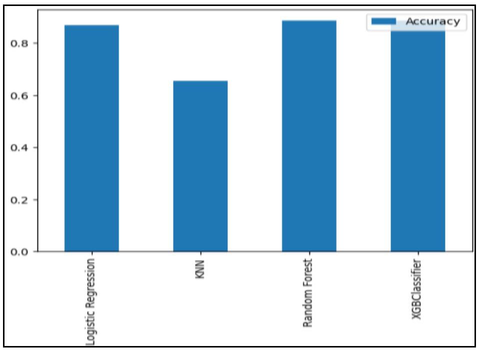

Fig. 4:- Evaluation of ML algorithms based on accuracy evaluation metric.

A. Optimizing ML algorithms based on hyper-parameter tuning

Optimizing hyperparameters is a vital aspect of training any machine learning model. Unlike parameters, which are learned from the data during the training process, hyperparameters play a crucial role in governing the learning process itself.[4] These hyperparameters are inherent in the mathematical formulation of machine learning models. For instance, while the weights in a linear regression model are parameters learned during training, the learning rate in gradient descent is considered a hyperparameter.[5] Achieving optimal performance on a dataset hinges on effectively tuning these model hyperparameters to identify the most favorable combination. Many techniques are there for hyperparameter optimization, like as Randomized Search, Grid Search, Bayesian Optimization, etc.

Grid Search and Randomized Search are two popular methods used in Hyper Parameter Tuning. Grid Search thoroughly examines every possible combination of given hyper parameter values, whereas Randomized Search takes a different approach. Instead of testing all provided parameter values, Randomized Search selects a set number of parameter combinations by sampling from the given distributions. When hyper parameters are listed, sampling occurs without replacement, but if at least one parameter is defined as a distribution, sampling with replacement is used. The sampling process in Randomized Search can be preconfigured. Each hyper parameter can be linked to either a distribution of possible values or a list of specific values, sampled evenly. For hyper parameters with continuous values, continuous distributions are recommended to maximize the advantages of randomization. A key benefit of Randomized Search is that it significantly cuts down on processing time.

Randomized Search has been applied on this dataset for hyper-parameter tuning and the results obtained are shown in the Fig. 5 below.

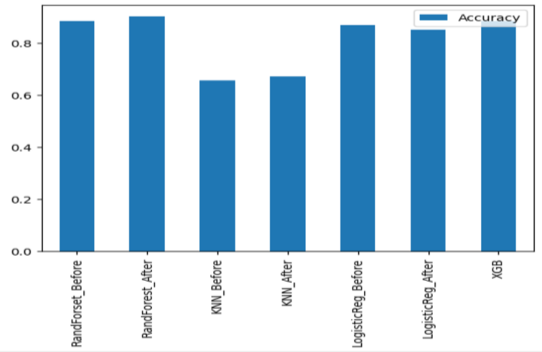

Fig. 5:- ML algorithms performance after applying hyper-parameter tuning.

From the diagram above clearly shows the performance improvement after hyper-parameter tuning the dataset. The performance of XGBoost was already optimal before application of hyper-parameter tuning therefore optimization was not applied to it.

IV. CONCLUSION

Heart diseases stand out as the leading cause of mortality in both developed and developing nations. The factors contributing to the occurrence of heart diseases in the human population are intricate and multifaceted. They extend beyond individual well-being to encompass the environmental impact on individuals. This research endeavors to leverage emerging machine learning (ML) techniques on a comprehensive dataset with the aim of enhancing model performance through hyperparameter tuning.

The study highlights that hyperparameter tuning plays a pivotal role in refining the accuracy of various ML models. By systematically adjusting hyperparameters, the researchers observed significant improvements in model performance. The findings of this research hold promise in contributing to the prevention of future heart attack occurrences. The application of advanced ML techniques, coupled with effective hyperparameter tuning, could prove instrumental in developing more accurate and reliable models for identifying and mitigating the risk factors associated with heart diseases. Ultimately, this research serves as a valuable step toward enhancing our understanding of heart disease dynamics and implementing proactive measures to reduce their impact on individuals' health.

REFERENCES

1. Anderson, P. A., The heart and development. In *Seminars in perinatology*, WB Saunders, 1996, December; 20(6): 482-509.
2. Cooper, R., Cutler, J., Desvigne-Nickens, P., Fortmann, S. P., Friedman, L., Havlik, R., ... & Thom, T., Trends and disparities in coronary heart disease, stroke, and other cardiovascular diseases in the United States: findings of the national conference on cardiovascular disease prevention. Circulation, 2000; 102(25): 3137-3147.
3. S. Mohan, C. Thirumalai and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," in IEEE Access, 2019; 7: 81542-81554.
4. Yu, T., & Zhu, H., Hyper-parameter optimization: A review of algorithms and applications, 2020; arXiv preprint arXiv:2003.05689.
5. Baydin, A. G., Cornish, R., Rubio, D. M., Schmidt, M., & Wood, F., Online learning rate adaptation with hypergradient descent, 2017; arXiv preprint arXiv:1703.04782.
6. Bergstra, J., & Bengio, Y., Random search for hyper-parameter optimization. Journal of machine learning research, 2012; 13(2).
7. Osband, I., Aslanides, J., & Cassirer, A., Randomized prior functions for deep reinforcement learning. Advances in Neural Information Processing Systems, 2018; 31.