IJISEM
INTERNATIONAL JOURNAL OF INNOVATIONS IN SCIENCE ENGINEERING AND MANAGEMENT

XGBoost for Heart Disease Prediction: Achieving High Accuracy with Robust Machine Learning Techniques

OPEN ACCESS

Volume: 3

Issue: 3

Month: July

Year: 2024

ISSN: 2583-7117

Published: 19.07.2024

Citation:

Rajni Gandha[1], Dr. Pankaj richhariya[2]
"XGBoost for Heart Disease Prediction: Achieving High Accuracy with Robust Machine Learning Techniques"
International Journal of Innovations In Science Engineering And Management, vol. 3, no. 3, 2024, pp. 07–14.

Rajni Gandha[1], Dr. Pankaj richhariya[2]
[1]Research Scholar , Department of Computer Science, Bhopal institute of technology & science
[2]HOD , Department of Computer Science, Bhopal institute of technology & science

Abstract
Examining the efficacy of the XGBoost algorithm, this study investigates the use of machine learning for the prediction of cardiac illness. We tested XGBoost against Decision Trees and Random Forests using a Kaggle dataset that included more than 1,000 patient records with fourteen important features. With an F1-score of 0.9816, a recall of 1.0, a precision of 96.39%, and an accuracy of 98.04%, our hyperparameter-optimized XGBoost model performed admirably. Prior approaches, such the hybrid Random Forest model, which achieved an accuracy of 88.7 percent with a reduced dataset, are surpassed by this model. There were no missing instances and few false positives, demonstrating the XGBoost model's dependability in predicting heart disease. Its recall and accuracy were also good. Our results show that XGBoost might be a powerful tool for early detection of cardiac disease; it offers major gains over current methods and lays the groundwork for predictive analytics studies in the future.

Keyword: Heart Disease Prediction, Machine Learning, XGBoost, Predictive Modeling.

INTRODUCTION

The need of reliable diagnostic technologies for early identification and intervention is highlighted by the fact that heart disease is still a top killer globally. Although essential, traditional diagnostic approaches sometimes have drawbacks including lower accuracy and intrusive procedures. One potential solution to these problems is the use of machine learning (ML) techniques, which can improve diagnostic results and increase the accuracy of predictions.

The prognosis of cardiac disease seems to benefit greatly from the recent developments in machine learning. Prediction performance has been noticeably enhanced by machine learning algorithms, especially those utilizing ensemble techniques and gradient boosting, as contrasted with more traditional approaches. Examples of machine learning models include Support Vector Machines and Random Forest, which have demonstrated superior accuracy and reliability compared to conventional statistical approaches (Zhou et al., 2016; Alizadehsani et al., 2019).[1][2]

The use of XGBoost, a state-of-the-art gradient boosting algorithm, for the purpose of cardiac disease prediction is the primary emphasis of this research. By using regularization and tree pruning to strengthen models and improve their accuracy and resilience, XGBoost has become well-known for its exceptional performance in a number of prediction tasks (Chen and Geetha, 2016) [2][3]. Although XGBoost has been shown to be useful in other fields, its potential use in predicting the occurrence of heart disease has received little attention.

Using XGBoost on a large dataset obtained from Kaggle, which includes more than a thousand medical records, our research intends to tackle the shortcomings of current models. The dataset provides a solid basis for training and evaluating models, since it contains important variables linked to heart disease. Current techniques, including hybrid Random Forest models, have obtained lower performance metrics (Krishnan et al., 2019) [4], hence improving prediction accuracy and resilience is the main target. This project aims to help develop heart disease prediction by utilizing XGBoost's superior capabilities. It will shed light on how effective contemporary machine learning approaches are in clinical diagnostics. By fixing important problems like accuracy and dataset limits, the results should offer a big improvement over earlier models.

THEORETICAL BACKGROUND

Decision Tree

An essential machine learning tool for regression and classification jobs is the Decision Tree method. It depicts decisions and their potential outcomes in a tree-like structure, with each node standing for an attribute-based decision, each branch for a result of that decision, and each leaf node for a class label or value. For preliminary exploratory research, Decision Trees are often preferred due to their simplicity and interpretability (Quinlan, 1986) [5]. A decision tree is built using a recursive partitioning method, which divides the dataset according to the characteristic that yields the most information gain or the one with the lowest impurity, whether it Gini impurity or entropy. Decision Trees are simple to implement, but they have a serious flaw: they may easily overfit, especially when dealing with complicated datasets, leading to models that are too specialized and unable to adapt to new information (Breiman et al., 1986) [6].

Random Forest

The idea of Decision Trees is expanded upon in Random Forest, an ensemble learning technique, by combining several trees into a single "forest." To lessen the likelihood of overfitting and decrease correlation between trees, each tree is trained on a bootstrap sample of the data and employs a randomly selected subset of characteristics for each split. This approach eliminates the danger of overfitting seen in individual Decision Trees (Breiman, 2001) [7]. A technique called bagging (Bootstrap Aggregating) is used by the Random Forest algorithm to generate a final conclusion. This involves training many models independently and then combining their predictions. Using the "wisdom of crowds" principle—the idea that a group of models may outperform a single model—this method makes predictions more accurate and resilient (Liaw and Wiener, 2002) [8].

XGboost

A state-of-the-art gradient boosting method, Extreme Gradient Boosting (XGBoost) combines gradient boosting with tree-based models to improve prediction performance. To avoid overfitting and enhance generalization, XGBoost uses regularization techniques, which are not included in classic boosting approaches (Chen and Guestrin, 2016) [9]. In XGBoost's model-building process, each new model iteratively fixes mistakes produced by its predecessors. The gradient of the loss function, which minimizes prediction errors and maximizes the model's performance, guides this process. In addition to its accuracy and resilience, the algorithm's performance is further improved by methods like tree pruning, parallel processing, and managing missing values (Friedman, 2001) [10]. As a result of its outstanding results in both real-world applications and machine learning contests, XGBoost is an effective tool for complicated predicting tasks.

RELATED WORK

Numerous novel strategies to improve diagnostic precision and efficacy have been investigated in current research on cardiac disease prediction. The Heart Disease Prediction System (HDPS) was created by Chen, Austin H., et al. [11] using an ANN that incorporates thirteen important clinical characteristics such as age, sex, and cholesterol levels. Offering a useful tool for heart disease categorization, this system attained an 80% prediction accuracy and contains tools for entering clinical data and visualizing performance indicators. Efficient approaches for detecting cardiac illness are urgently needed, as pointed out by Patel, Jaymin, and Dr. Tejal Upadhyay [12]. Data mining techniques can help in diagnosis and reducing needless procedures. Using the Cleveland heart disease dataset, their study evaluates several algorithms, with an emphasis on the Decision Tree approach. These algorithms include J48, Logistic Model Tree, and Random Forest. Their objective is to find the best algorithm for detecting cardiac problems and finding hidden patterns to improve prediction accuracy. In a similar research, Palaniappan [13] suggested the Intelligent Heart Disease Prediction System (IHDPS) to tackle the problem of unused healthcare data. This system has sophisticated data mining skills to process intricate queries and forecast the probability of heart illness using medical profiles. It was created using Decision Trees, Naïve Bayes, and Neural Networks. The IHDPS, built on the.NET framework, is renowned for its simplicity of use and

scalability; it was created to discover important links and patterns in data related to heart disease.

The significance of precise diagnosis and prognosis in the treatment of cardiovascular illnesses has been highlighted by recent research. In an effort to decrease the number of incorrect diagnoses and deaths caused by them, Bhatt, Chintan M., et al. [14] highlight the growing importance of machine learning in healthcare, particularly pattern recognition. Using GridSearchCV for optimization, their study suggests a model that uses k-modes clustering with Huang as its starting point and tests out a number of techniques, such as XGBoost, decision tree, multilayer perceptron, and random forest. The usefulness of machine learning in cardiovascular disease prediction was demonstrated by the multilayer perceptron model, which achieved the greatest accuracy of 87.28% when tested on a Kaggle dataset. Repaka and Anjan [15] investigate the use of data mining techniques in the detection of cardiac diseases, particularly through the creation of the SHDP system. The SHDP system uses medical records to forecast variables associated with cardiovascular disease using the Naive Bayesian method. The study emphasizes the need of safe data transfer and mobile health apps, demonstrating that the SHDP system is a useful tool for predicting heart disease using big datasets.

A machine learning model for the prediction of heart illness using the UCI Heart illness dataset is the main emphasis of Sharma, Vijeta et al. [16], which addresses the concerning rise in heart-related disorders. Among the algorithms tested in their study, random forest outperformed SVM, Naive Bayes, and decision trees in terms of accuracy. This research shows that machine learning models can be useful decision support systems, which can help doctors diagnose and anticipate cardiac problems more efficiently. The rising incidence of cardiac disease has recently prompted calls for more sophisticated computational approaches. In order to forecast coronary artery heart disease, Ayon et al. [17] examine and contrast several AI methods, such as Logistic Regression, Support Vector Machines (SVM), Deep Neural Networks (DNN), Decision Tree, Naïve Bayes, Random Forest, and K-NN. With a high sensitivity and precision rate of 98.15% with the Statlog and Cleveland heart disease datasets, they discovered that the DNN was the most successful. Findings from this study support DNN's position as the most promising approach among those tested for coronary artery heart disease prediction.

The significant death rates from cardiovascular illnesses and the diagnostic diversity caused by individual patient patterns are discussed in detail by Dubey and Animesh Kumar [18]. Their research is to examine the worldwide effects of cardiovascular illnesses, the risks associated with various age groups, and the efficacy of current methods for early-stage prediction. They draw the conclusion that the risk of cardiovascular illness rises with age and emphasize the usefulness of data mining and optimization techniques for early-stage prediction after reviewing several sources of data. In order to improve the prediction and diagnosis of cardiac illness, the authors propose a hybrid strategy that combines algorithms inspired by biology with classification and clustering approaches. In their discussion of the difficulty of using clinical data analysis to forecast cardiovascular illness, Mohan et al. [19] draw attention to the function of ML in enhancing the precision of diagnoses. To improve prediction accuracy, they suggest an HRFLM, which combines a random forest with a linear model, to find important characteristics using ML approaches. The HRFLM model outperforms the competition when it comes to predicting cardiovascular illness, with an accuracy of 88.7 percent when tested on large amounts of healthcare data. This research shows that ML methods have great promise for improving clinical decision-making and optimizing cardiac disease prediction systems.

METHODOLGY

Dataset

Using a large dataset obtained from Kaggle, which included medical information pertinent to the prediction of heart disease in the past, we conducted this study. Beginning in 1988, the collection incorporates information from four separate databases: Long Beach V, Cleveland, Hungary, and Switzerland. Building strong prediction models relies on this large collection's varied array of patient information and features.

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
3	61	1	0	148	203	0	1	161	0	0.0	2	1	3	0
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0

Figure 1 Dataset Sample

There are 76 properties in the dataset, and one of them is the target characteristic that indicates if heart disease is present or not. We narrowed our attention to 14 variables for model training and assessment since they have been routinely employed in published trials, guaranteeing relevance and comparability. Among the chosen characteristics are:

Age: The age of the patient.
Sex: The gender of the patient (binary: male or female).
Chest Pain Type: Categorized into four types (e.g., typical angina, atypical angina, non-anginal pain, asymptomatic).
Resting Blood Pressure: Measured in mmHg.
Serum Cholesterol: Measured in mg/dl.
Fasting Blood Sugar: Binary indicator (>120 mg/dl or ≤120 mg/dl).
Resting Electrocardiographic Results: Categorical values (0, 1, 2) representing various electrocardiographic findings.
Maximum Heart Rate Achieved: The highest heart rate recorded during exercise.
Exercise-Induced Angina: Binary indicator (yes or no).
Oldpeak: ST depression induced by exercise relative to rest.
Slope of Peak Exercise ST Segment: Categorized into three types.
Number of Major Vessels: Count of major vessels colored by fluoroscopy (0-3).
Thalassemia: Categorical values (0 = normal, 1 = fixed defect, 2 = reversible defect).

A strong basis for training and testing machine learning models for heart disease prediction is provided by the dataset's broad and rich features, together with thorough preprocessing and suitable splitting. The research findings are further strengthened by the fact that the models built are accurate and generalizable, thanks to this extensive dataset.

Data Pre Processing

To ensure accurate modeling, clean and consistent data, data preparation is an essential step in getting the dataset ready for machine learning applications. In order to prepare the dataset for predictive modeling, this study methodically carried out a number of preprocessing steps. At first, avoiding inaccurate model training was the main focus when dealing with missing variables. Imputation techniques were used for numerical attributes that had missing entries. In these cases, the mean or median of the attribute was used to replace the missing values, depending on the data's distribution and characteristics. In order to keep the dataset intact, missing values for categorical characteristics were either filled in via mode imputation or a placeholder category was assigned. The next step was to identify and handle outliers, as they might have a major effect on the accuracy of the model. These outliers were located using statistical tools such as z-scores and IQR analysis. To keep outliers from throwing off the model's accuracy, we used transformation or capping to keep their impact to a minimum. The next step was to convert the categorical features into a numerical representation that machine learning algorithms could understand. To make the variables more like binary dummy variables, we used one-hot encoding on the categories ones. For instance, characteristics like 'Sex,' 'Chest Pain Type,' 'Resting Electrocardiographic Results,' 'Exercise-Induced Angina,' and 'Slope of Peak Exercise ST Segment' were quantified using binary variables. The program was able to better understand categorical data because of this encoding procedure.

Another important stage in the preprocessing was feature scaling. To guarantee that numerical characteristics contribute equally to the model, they were standardized. For algorithms like gradient boosting that are sensitive to feature scales, it is crucial to normalize features such that they have a mean of zero and a standard deviation of one. A later 80:20 split was used to divide the dataset into training and testing sections. With this setup, we could train the model and fine-tune its hyperparameters on the training set, and then assess its efficacy and generalizability on the testing set. In order to

get a good gauge of the model's prediction power, this method makes use of unseen data for evaluation. To further prepare the data for use by machine learning techniques, dummy variables were also constructed for category categories. To make sure the data was encoded correctly for model training, binary columns were used to represent each category of the categorical characteristics.

Proposed Model

Selecting, training, and evaluating algorithms to construct predictive models is known as model building, and it is an essential part of the machine learning workflow. To find the best way for predicting cardiac problems, we tested various machine learning algorithms in this study. Random Forest, Decision Tree, and Extreme Gradient Boosting (XGBoost) are among the algorithms that are being investigated.

Decision Tree One of the most basic algorithms, Decision Tree builds a decision model similar to a tree using the values of attributes. Decision Trees are easy to use and provide interpretability, but they can overfit, particularly with complicated datasets. So, they weren't the main focus of our model selection, but they did give a baseline for comparison.

Random Forest, Combining several trees trained on separate data subsets, Random Forest expands upon the idea of Decision Trees as an ensemble approach based on bagging. By averaging the predictions of several trees, this method enhances prediction accuracy and helps alleviate the overfitting issue that individual Decision Trees may have. In order to evaluate Random Forest's performance in comparison to other algorithms, a comparative model was utilized.

XGBoost, Out of all the algorithms that were tested, XGBoost (Extreme Gradient Boosting) showed the most promise. When it comes to a number of different types of prediction tasks, XGBoost—an advanced gradient boosting technique—is unrivaled. The process of building models is sequential, with each new model making an effort to fix the mistakes made by the ones before it. The accuracy and resilience of the model are improved by combining this approach with regularization techniques and tree pruning. XGBoost is an excellent tool for predictive modeling since it efficiently computes, handles missing variables, and employs advanced optimization algorithms.

A set of optimized hyperparameters was used to the XGBoost classifier during model construction. A random state of 55, 500 estimators, and a learning rate of 0.1 were the parameters that were selected. Each tree's contribution to the final model is controlled by the learning rate, and the overall number of trees in the ensemble is defined by the number of estimators. The random state initializes the algorithm's random number generator, which assures that results can be reproduced.

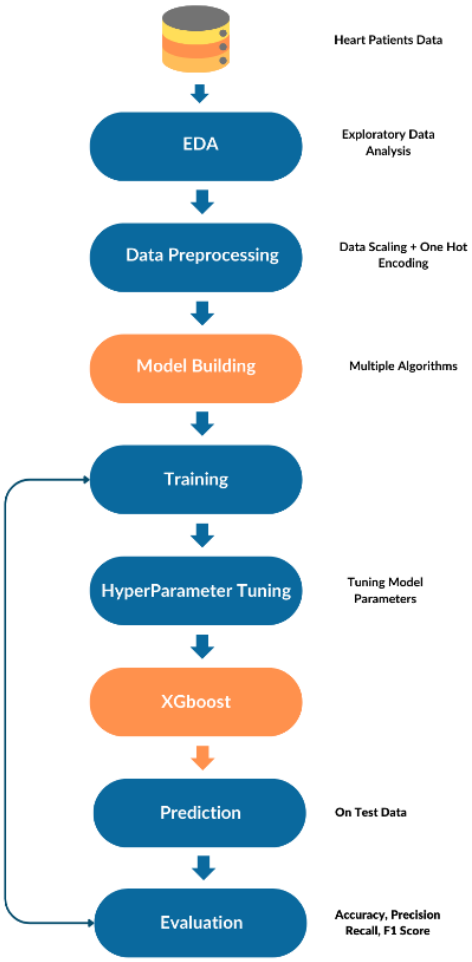

Figure 2 Proposed Workflow

With 80% of the data coming from the training set, the model was trained using the preprocessed dataset. In order to maximize the model's efficiency, hyperparameter tuning was employed. This entailed modifying parameters like learning rate and number of estimators according to the outcomes of cross-validation. The testing set, which made up the last 20% of the data, was used to evaluate the final model.

RESULTS AND DISCUSSION

Because of XGBoost's proven superiority in prediction accuracy and resilience, we've decided to incorporate it into

our suggested model. Thanks to its fine-tuned settings, the XGBoost classifier was able to achieve an impressive 98.04% accuracy, 96.39% precision, 1.0 recall, and 0.9816 F1-score. Compared to other approaches, this one is light years ahead of the competition, even the 88.7 percent accurate hybrid Random Forest model that was utilized in earlier research.

The suggested XGBoost model outperforms the competition across a number of important parameters in the prediction of cardiac illness. Impressive findings highlight the efficiency of the XGBoost classifier in predicting the existence of heart disease when assessed on the preprocessed dataset.

Accuracy: 98.04% was the accuracy attained by the XGBoost model. According to the model's accuracy, which is the percentage of occurrences properly predicted relative to the total number of instances, it successfully categorized 98.04% of the dataset samples. The model's capacity to accurately differentiate between individuals with and without cardiac disease is demonstrated by its high level of accuracy.

Precision: The XGBoost model achieved an precision of 96.39%. Precision is the ratio of the number of correct forecasts to the total number of correct predictions produced by the model. More specifically, the accuracy rate for cases that were deemed positive (meaning the existence of cardiac disease) was 96.39%. The dependability of the model in reducing false positives is demonstrated by its excellent accuracy; this is especially important in clinical settings, where patients may experience needless anxiety and further testing due to false positives.

Recall: The model's recall was 1.0, which is 100%. The fraction of true positives that the model properly detected is measured by recall, which is also called sensitivity. When the recall is 1.0, it means that the XGBoost model found every single patient with cardiac illness. Because of the grave consequences that might result from misdiagnosing a patient with heart disease, this measure is vital in medical diagnosis.

F1-Score: An F1-score of 0.9816 was recorded by the XGBoost model. A fair evaluation of the model's accuracy in avoiding false positives and negatives may be found in the F1-score, which is the harmonic mean of recall and precision. With an F1-score of 0.9816, the model reliably and effectively predicts the occurrence of heart disease, indicating a high degree of accuracy.

Table 1 Results of XGboost on Test Data

Metrics	Value
Accuracy	0.9804
Precision	0.9639
Recall	1.0
F1 Score	0.9816

Confusion Matrix: The confusion matrix for the XGBoost model is as follows:

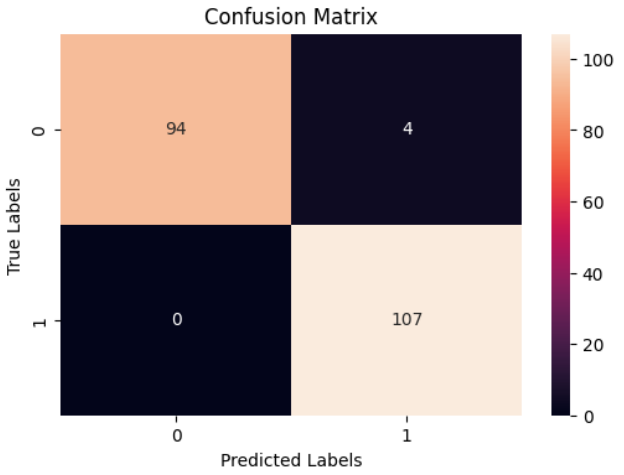

Figure 3 Confusion Matrix of XGboost

In this matrix, the rows represent the actual classes (with 0 indicating no disease and 1 indicating the presence of disease), while the columns represent the predicted classes. The matrix shows:

- **True Positives (TP)**: 107 instances were correctly predicted as having heart disease.
- **True Negatives (TN)**: 94 instances were correctly predicted as not having heart disease.
- **False Positives (FP)**: 4 instances were incorrectly predicted as having heart disease when they did not.
- **False Negatives (FN)**: 0 instances were missed by the model, meaning no cases of heart disease were incorrectly predicted as not having the condition.

With zero false negatives and almost no false positives, the confusion matrix provides more proof of the XGBoost model's outstanding performance. The model has successfully recognized all patients with heart disease, as shown by its high recall, and its resilience and dependability are emphasized by its excellent precision and accuracy.

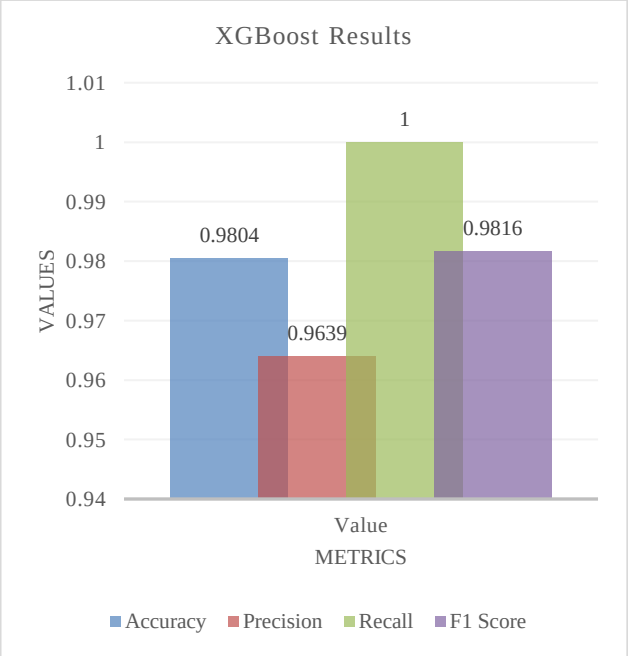

Figure 4 Results of XGboost on Test Data

Our work presents a new method for predicting the occurrence of heart disease using the XGBoost algorithm, which outperforms other studies in this area. We trained our proposed XGBoost model on a larger dataset exceeding 1,000 patient records and achieved an impressive accuracy of 98.04%, recall of 1.0, and an F1-score of 0.9816. This is in contrast to the existing study, which used a hybrid Random Forest with a linear model and achieved an accuracy of 88.7% with a smaller dataset of 297 records. This improvement not only makes predictions more accurate, but it also makes the model more reliable and robust, so it can identify all cases of heart disease without missing a beat. This solves the problems with older methods and makes early diagnosis much better.

Table 2 Work Comparison of Existing and Proposed work

Aspect	Existing Work [20]	Proposed XGBoost Model
Algorithm	Hybrid Random Forest with Linear Model	XGBoost
Accuracy	88.7%	98.04%
Precision	87.5%	96.39%
Recall	92.8%	1.0
F1-score	90%	0.9816
Dataset Size	297 patient records	Over 1000 patient records
Confusion Matrix	-	[[94 4] [0 107]]
Model Robustness	Limited by hybrid model's complexity	Advanced gradient boosting with robust optimization techniques
Novelty	Hybrid approach combining Random Forest and Linear Model	State-of-the-art XGBoost algorithm with advanced features

CONCLUSION

With an emphasis on the XGBoost algorithm's efficacy, this study offers a thorough evaluation of machine learning for the prediction of cardiac illness. In comparison to more conventional models like Decision Trees and Random Forests, our findings show that the sophisticated gradient boosting method XGBoost provides far better performance. We obtained impressive results: a recall of 1.0, an F1-score of 0.9816, an accuracy of 98.04%, and a precision of 96.39% when training and testing the XGBoost classifier on a large dataset consisting of more than 1,000 patient records. These measures demonstrate that the model can consistently and precisely predict the occurrence of heart disease with very few false positives and no false negatives. Our method is unique because we employ XGBoost, which has very good performance metrics that outperform those of other research, including ones that used hybrid Random Forest models. XGBoost's strengths lie in its built-in optimization and regularization algorithms, its efficiency with big datasets, and its capacity to manage complicated data linkages and interactions. The XGBoost model's confusion matrix demonstrated a very high level of classification accuracy, with almost no false positives and zero false negatives, demonstrating its efficacy in predicting the existence of cardiac illness.Our suggested approach is successful, as shown by the vast evaluation across many performance indicators and the gains in prediction accuracy. The improved generalizability and reliability of the model are also due to the large dataset and careful preprocessing.

Finally, the study confirms that XGBoost may forecast the occurrence of cardiac disease, which could lead to better patient outcomes through earlier detection. Additional characteristics, more sophisticated algorithms, and validation on varied datasets might be investigated in future study to improve prediction accuracy and clinical usefulness. This discovery lays the groundwork for future studies that want to improve the prediction of cardiac disease using advanced machine learning methods.

REFERENCE

[1] Alizadehsani, Roohallah, et al. "Machine learning-based coronary artery disease diagnosis: A comprehensive

review." Computers in biology and medicine 111 (2019): 103346.

[2] Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016..

[3] Zhou, Zhi-Hua. "Ensemble methods." Combining pattern classifiers. Wiley, Hoboken (2014): 186-229.

[4] Krishnan, Santhana, and S. Geetha. "Prediction of heart disease using machine learning algorithms." 2019 1st international conference on innovations in information and communication technology (ICIICT). IEEE, 2019.

[5] Quinlan, J.R. “Induction of Decision Trees.” Machine Learning, vol. 1, no. 1, 1986, pp. 81-106.

[6] Breiman, Leo, et al. “Classification and Regression Trees.” Monographs in Statistics and Applied Probability, vol. 1, Wadsworth International, 1986.

[7] Breiman, Leo. "Random forests." Machine learning 45 (2001): 5-32.

[8] Liaw, Andy, and Matthew Wiener. “Classification and Regression by RandomForest.” R News, vol. 2, no. 3, 2002, pp. 18-22.

[9] Chen, Tianqi, and Carlos Guestrin. “XGBoost: A Scalable Tree Boosting System.” Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785-794.

[10] Friedman, Jerome H. “Greedy Function Approximation: A Gradient Boosting Machine.” Annals of Statistics, vol. 29, no. 5, 2001, pp. 1189-1232.

[11] Chen, Austin H., et al. "HDPS: Heart disease prediction system." 2011 computing in Cardiology. IEEE, 2011.

[12] Patel, Jaymin, Dr TejalUpadhyay, and Samir Patel. "Heart disease prediction using machine learning and data mining technique." Heart Disease 7.1 (2015): 129-137.

[13] Palaniappan, Sellappan, and Rafiah Awang. "Intelligent heart disease prediction system using data mining techniques." 2008 IEEE/ACS international conference on computer systems and applications. IEEE, 2008.

[14] Bhatt, Chintan M., et al. "Effective heart disease prediction using machine learning techniques." Algorithms 16.2 (2023): 88.

[15] Repaka, Anjan Nikhil, Sai Deepak Ravikanti, and Ramya G. Franklin. "Design and implementing heart disease prediction using naives Bayesian." 2019 3rd International conference on trends in electronics and informatics (ICOEI). IEEE, 2019.

[16] Sharma, Vijeta, Shrinkhala Yadav, and Manjari Gupta. "Heart disease prediction using machine learning techniques." 2020 2nd international conference on advances in computing, communication control and networking (ICACCCN). IEEE, 2020.

[17] Ayon, Safial Islam, Md Milon Islam, and Md Rahat Hossain. "Coronary artery heart disease prediction: a comparative study of computational intelligence techniques." IETE Journal of Research 68.4 (2022): 2488-2507.

[18] Dubey, Animesh Kumar, and Kavita Choudhary. "A systematic review and analysis of the heart disease prediction methodology." International Journal of Advanced Computer Research 8.38 (2018): 240-256.

[19] Mohan, Senthilkumar, Chandrasegar Thirumalai, and Gautam Srivastava. "Effective heart disease prediction using hybrid machine learning techniques." IEEE access 7 (2019): 81542-81554.

[20] Mohan, Senthilkumar, Chandrasegar Thirumalai, and Gautam Srivastava. "Effective heart disease prediction using hybrid machine learning techniques." IEEE access 7 (2019): 81542-81554.[2] Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016.