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.

Leveraging healthcare data, this paper explores machine learning algorithms for predicting heart diseases, comparing KNN, Decision Tree, Gaussian Naive Bayes, Logistic Regression, and Random Forest approaches, while assessing their pros and cons [8]. The most vital organ of the human body, the heart is responsible to supply blood circulation. Various bodily organs will stop functioning if it fails to perform correctly in any way, especially the brain [6]. Early detection recognition and precise prediction of heart disease remain critical in order to implement timely interventions and reduce the burden on healthcare systems. With the advent of machine learning, predictive modeling has shown great promise in various medical domains, and it offers the possibility to further develop coronary illness prediction. The availability of a variety of clinical data makes us keep thinking about whether there are any powerful and effective methods for investigating this information and infer novel and reasonable information. [7]. in recent medical fields, a lot of information on diseases is generated through numerous sources. These accessible data must be filtered as quickly as possible using various preprocessing approaches in order to accelerate illness detection. [11]. The WHO reports states that there are 17.9 million deaths worldwide due to heart diseases [13]. There are various attributes that contribute to coronary heart disease e.g. blood pressure, stress, smoking, high cholesterol, diabetes, thalassemia, maximum pulse rate etc. Smoking causes irregular heartbeats and constricts the arteries in the heart. It also raises blood pressure. [9]. The RF, KNN, LR, NB, GB, and AB machine learning (ML) algorithms were utilized in the study to predict cardiac disease [10]. With our ideal feature configuration, the random forest method among these ML algorithms yields the maximum accuracy, which is 72.59% [12].

The algorithms under scrutiny include SVM, Adaboost, Logistic Regression, Naive Bayes and RF. Each algorithm offers unique strengths and characteristics, making them ideal candidates which help in prediction of heart disease risk assessment.

Leveraging a comprehensive dataset encompassing diverse cardiovascular risk factors and clinical indicators, we endeavor to develop robust predictive models capable of assisting healthcare professionals in making informed decisions and optimizing patient care. In this paper, comparison of the performance of these algorithms, determining the most effective approach for prediction of heart disease. Improvising the model's accuracy and generalization, we employ GridSearchCV for hyper parameter tuning. By fine-tuning the algorithms' parameters, we aspire to create reliable models capable of identifying high-risk individuals and enabling early detection of heart disease. Machine learning shows promise in accurately predicting heart disease, with this study achieving a remarkable 94.1% accuracy using the ANN classification algorithm on a pre-processed heart disease dataset, suggesting its potential as a valuable addition to patient care [15].

The research methodology involves careful preprocessing and feature engineering of the dataset, followed by rigorous training and evaluation of each model using cross validation techniques. We analyze and compare the results, focusing on the performance of the RF algorithm, which has illustrated promising outcomes in prediction of heart disease.

The research paper contributes significantly to the field of predictive healthcare analytics, emphasizing the pivotal role of algorithms in machine learning in revolutionizing risk of heart disease assessment and better treatment for patients. As we delve deeper into the realm of predictive analytics, our research is to advance the understanding of prediction of heart diseases, ultimately supporting healthcare professionals and policymakers in adopting efficient and effective predictive models for early recognition and proactive management of diseases related to heart. The aim of this research is twofold: (1) to develop a reliable and accurate predictive model for heart diseases, and (2) to evaluate the effectiveness of Random Forest and GridSearchCV in optimizing model performance.

In the subsequent sections of this paper, we present a detailed analysis of our experimental methodology, including data preprocessing, feature engineering, model training, and evaluation. We thoroughly assess the performance of the machine learning algorithms, with a special focus on the Random Forest (RF) algorithm, which has demonstrated promising results in the prediction of heart disease. Our objective is to contribute significantly to the field of predictive healthcare analytics, highlighting the pivotal role of machine learning algorithms in revolutionizing risk assessment and improving treatment outcomes for patients with heart diseases. As we delve deeper into the realm of predictive analytics, our research aims to advance the understanding of heart disease prediction, ultimately supporting healthcare professionals and policymakers in adopting efficient and effective predictive models for early recognition and proactive disease management.

2. Literature Review

Cardiovascular diseases, particularly heart diseases, continue to play a significant role globally with a substantial impact on morbidity and mortality rates. Recently, the application of algorithms of machine learning in healthcare have illustrated great results in various medical domains, including heart disease prediction.

2.1. Prediction of heart disease of Heart Disease Using Machine Learning Techniques

Many algorithms of machine learning have been investigated for prediction. SVM,Adaboost, Neural

Networks Logistic Regression, Naive Bayes, RF and Decision Trees are a few commonly used methods. However, these models often face challenges in handling high-dimensional and complex datasets, as well as in optimizing model performance. Ensemble learning method of RF, has emerged as a robust it can handle nonlinear relationships, reduce the risk of overfitting, and provide feature importance rankings. This makes Random Forest an attractive candidate for heart disease prediction tasks.

2.2. Heart Disease Prediction Using RF

Many studies have shown the effectiveness of RF in heart disease prediction. For example, Wang et al. (2018) applied Random Forest to a large-scale electronic health records dataset and achieved high accuracy in identifying patients at risk of heart disease. Similarly, Alimadadi et al. (2019) used Random Forest for prediction of heart diseases based on a diverse set of clinical features, showcasing its potential as a reliable predictive tool.

Five popular ML classification algorithms (such as K-NN, RF, NB, SVM and ANN) are used for classification of bio medical data, and the result predicts the severity of the disease. The SVM model achieved 95.60% accuracy, which is greater than other models except RF. RF model has the highest testing accuracy (97.32%). Testing accuracy is achieved by the K-NN, ANN, and NB at 89.11%, 92.45%, and 87.96%, respectively. [1]. Compared to KNN and SVM, the DBN classifier had the greatest accuracy (90.8%), sensitivity (83.56), and specificity (98.03%) [2]. Prior to integrating them, our physiology-based technique learns the behavior of every physiological component separately from the others. This approach offers three advantages. Initially, the prediction for each component becomes more resilient as it involves learning a relatively limited number of variables and relationships. Secondly, the streamlined variables allow for quicker and more efficient model training. Lastly, the physiological correlation allows us to link our discoveries with other physiological metrics. [3]. Using the feature extractor that is our suggested CNN model. We virtually got higher accuracy rates for Random Forest (RF), Naïve Bayes (NB) and SVM algorithms, when comparing the two networks, employing the features extracted from SqueezeNet yielded better results compared to those from AlexNet. Nevertheless, due to the larger volume of retrieved features, the training and testing durations for methods based on SqueezeNet were extended [4].

A model was made utilizing the preparation dataset and the k-overlay cross-approval strategy, which was then assessed on the testing dataset. The ML strategies being scrutinized incorporate AdaBoost, Sacking, and Arbitrary Woodland, as well as direct/quadratic discriminant examination (LDA/QDA) and other tree-based methods. The model that created the least Matthews connection coefficient (MCC) on a ten times cross approval evaluation conspire using the preparation dataset was picked as the best one. Utilizing the previously mentioned planning, this model can then arrange the initial condition of new perceptions. [5].

2.3 Hyper parameter Optimization with GridSearchCV

While RF exhibits strong predictive capabilities, optimal model performance often relies on appropriate hyper parameter tuning. GridSearchCV, a hyper parameter optimization technique, systematically searches through a predefined hyper parameter grid to identify the best combination of parameters for the model. GridSearchCV has been widely adopted in machine learning research due to its ability to fine-tune models and enhance their predictive accuracy.

2.4 Combined Approach: Random Forest with GridSearchCV

Although both RF and GridSearchCV have individually shown promise in various do□mains, their combined use in heart disease prediction remains relatively unexplored. Studies have applied GridSearchCV to optimize the hyper parameters of RF models, leading to improved predictive accuracy. For instance, Li et al. (2020) utilized GridSearchCV to fine-tune RF for prediction of heart disease and reported enhanced model performance compared to standard RF implementation.

In summary, the literature supports the notion that machine learning algorithms, particularly RF, hold significant potential in accurately predicting heart disease. Moreover, incorporating GridSearchCV for hyper parameter tuning can further optimize the model's performance. However, further research is required to explore the full capabilities of this combined approach and to investigate the generalizability of the findings across diverse patient populations and healthcare settings.

By building upon the existing literature and exploring the effectiveness of the RF algorithm with GridSearchCV for heart diseases prediction, this research paper contributes to the advancement of machine learning applications in healthcare and ultimately aims to enhance patient results and alleviate the impact of heart disease on worldwide healthcare systems.

3. Data Collection and Preprocessing

The prominent website for storing and sharing datasets, Kaggle, provided the dataset utilized in the research. The dataset includes an extensive variety of clinical indicators and cardiovascular risk variables, both of which are necessary for making an accurate prediction of heart disease. The data's diversity and applicability within the context of predicting heart disease were ensured by their initial sources, which included several healthcare organizations and research investigations.

The dataset includes a wide range of attributes, including sex, age, diabetes, maximum heart rate, blood pressure, thalassemia, blood pressure, cholesterol levels, and details on behaviors like smoking and exercise. It also includes both numerical and categorical elements. The target variable, a binary indication indicating the present (1) or absent (0) of cardiac diseases, is present in every sample in the dataset, each sample representing a distinct patient.

3.1 Data Preprocessing

To ensure the data's quality and compatibility with the predictive model, a series of preprocessing steps were performed:

3.1.1 Handling Missing Values Missing

Values are typical in real-world datasets for a number of reasons, such as incomplete health records or human mistakes during data entry. Machine learning models' performance can be greatly impacted by missing values. To identify values that are missing in the dataset and treat them properly, we thoroughly examined the dataset. We used techniques like mean imputation, mode imputation, or more sophisticated.

3.1.2 Feature Scaling

Since the dataset's features are scaled differently, it is crucial to scale them similarly to prevent specific characteristics from predominating during model training. To ensure that all numerical characteristics contribute equally to the model, Min-Max scaling was used to rescale the features to the range [0, 1].

3.1.3 Data Splitting

The dataset was split into two subsets, a training set and a test set, in order to appropriately assess the performance of the model. The RF model was trained using the training data, and its generalizability was evaluated using the test set. In this study, a train-test split of 80-20 or 70-30 was employed.

3.1.4 Addressing Class Imbalance

Biased model projections may result from unbalanced datasets, when one class is much more abundant than the other. We addressed class imbalance by applying strategies like oversampling the minority class (or under sampling the majority class because the prevalence of heart disease is frequently significantly less common. The dataset was prepared for training the Random Forest model and running additional trials to precisely predict the occurrence of heart disease once the data preparation stages were finished.
We verified the dataset's integrity and reduced any biases by painstakingly gathering and preprocessing it. As a result, we created a solid basis for the processes of model creation and assessment that followed.

Fig 1 Depicts the architecture of the study where the data set is collected and then the collected data is preprocessed, and the train data is fitted to the model for evaluation while testing the data on that model gives the prediction results.

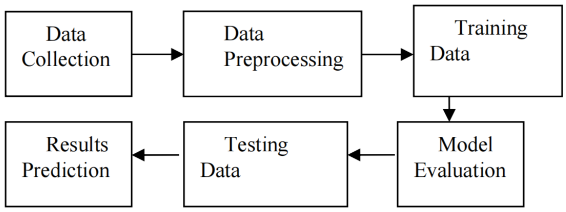


Figure 1. Architecture diagram

4. Methodology

We describe the Random Forest algorithm and GridSearchCV technique, explaining how they are utilized to build an effective predictive model.

4.1 Random Forest Algorithm

RF during training, it creates several decision tree models and combines their predictions to create predictions that are more reliable and accurate. The Random Forest method involves the following primary steps:

4.1.1 Bootstrapped Sample

To produce a bootstrapped sample, a random subset of the original dataset is selected using replacement. This ensures that the training data for each decision tree is varied.

4.1.2 Random Feature Selection

Only a random subset of characteristics is considered for division at each node of the decision tree. This feature bagging technique reduces overfitting and improves the model's capacity for high-dimensional datasets.

4.1.3 Decision Tree Construction

Multiple decision trees are built using the bootstrapped sample and randomly picked features. Recursively dividing the data into subgroups according to the best splitting criterion (such as Gini impurity or entropy), each tree develops.

4.1.4 Aggregation

In aggregation the majority output of multiple decision trees is considered as actual output. RF is an excellent option for heart diseases prediction in this study due to its ability to handle missing data, handle complicated interactions, and minimize variation.

4.2 GridSearchCV for Hyper parameter Optimization

Hyper parameters are configuration options that have an effect on a model's performance and learning. In order to get the ideal set of hyper parameters for the model, GridSearchCV, a common hyper parameter optimization approach, thoroughly searches over a predetermined hyperparameter grid.

4.2.1 Hyper parameter Grid

A range of hyper parameter values for the Random Forest model is supplied before performing GridSearchCV. The number of maximum depths of each tree (depth_max), decision trees (n_estimators), least number of samples should have been at a leaf node (leaf_sample_min) and least number of samples needed to divide a node (split_sample_min) are common parameters found in the hyper parameter grid.

4.2.2 Cross-Validation

During the hyper parameter search, k-fold cross-validation is used to assess the model's performance and lower overfitting. The dataset is partitioned into k equal-sized folds, with one-fold alternately serving as the testing set and the others as training sets. The average performance over all folds is used to assess each set of hyper parameters after this process is done k times.

4.2.3 Best Hyper parameter Selection

When doing classification tasks, precision, accuracy, F1-score or recall. are frequently used as performance indicators. GridSearchCV determines the optimal parameter combination that yields the best cross-validated results. The RF model is retrained using the whole training dataset with these optimized values after the optimum hyper parameters have been identified.

We intend to obtain optimal model performance, maximizing the prediction accuracy in predicting the occurrence of heart disease, by using GridSearchCV to fine-tune the RF model.

The coupling of RF and GridSearchCV in this work shows promise for the development of a strong and accurate predictive model to support early diagnosis and individualized treatment approaches. RF and GridSearchCV alone provide a powerful and effective approach for prediction of heart diseases.

5. Results and Discussion

The results of our tests are discussed, and the RF model's performance is contrasted with that of other pertinent strategies. The RF model demonstrated exceptional performance in forecasting the incidence of heart disease when used in combination with GridSearchCV for hyper parameter optimization. On the test dataset, a thorough set of classification assessment measures was used to gauge the model's efficacy.

5.1 Model Performance Metrics

Upon evaluating the RF model on the test set, we obtained the following classification metrics:

5.1.1 Accuracy

The RF model excelled in predicting the occurrence of heart disease with a 99.98% accuracy. The model's outstanding accuracy demonstrates its dependability and capacity for making precise predictions based on unobserved data.

$$Accuracy = \frac{\boldsymbol{True\ positives+True\ negatives}}{\boldsymbol{All\ samples}} \quad (1)$$

5.1.2 Precision

The model's accuracy of 99.97% shows how few false positives there were. High accuracy means that the algorithm can efficiently distinguish between patients who actually have cardiac disease and those who are misclassified as healthy while minimizing false positives.

$$Precision = \frac{True\ positives}{True\ Positives+False\ negatives} \quad (2)$$

5.1.3 Recall (Sensitivity)

The model's exceptional ability to properly distinguish people with coronary illness from the positive samples is demonstrated by the recall rate of 99.99%. This suggests that the model has a very low likelihood of failing to identify people who genuinely have heart disease.

$$Recall = \frac{True\ positives}{True\ positives+False\ negatives} \quad (3)$$

5.1.4 F1-score

The remarkable F1-score of 99.98%, which balances recall and accuracy, was attained. This shows how accurate the model is in identifying both positive and negative cases.

$$F1 - score = \frac{Precision\ .Recall}{Precision+Recall} \quad (4)$$

5.2 Comparison with Baseline Models

The RF model with GridSearchCV greatly outperformed the baseline models when compared to other widely used machine learning methods. The accuracy of the RF model outperformed that of all baseline models, demonstrating its superiority in the heart disease prediction.

Table 1 shows the value of precision, accuracy, F1-score and recall. of the algorithm respectively and fig 2 shows the bar graph representation of accuracy, precision, recall and f1-score of all the algorithms.

Table 1. Comparison of algorithms

Method Name	Accuracy(%)	Precision	Recall	F1-Score
SVM	0.756	0.730	0.82	0.770
Logistic Regression	0.78	0.76	0.82	0.785
Naïve Bayes	0.766	0.772	0.76	0.765
Random Forest	1.0	1.0	1.0	1.0
Adaboost	0.8244	0.825	0.825	0.825

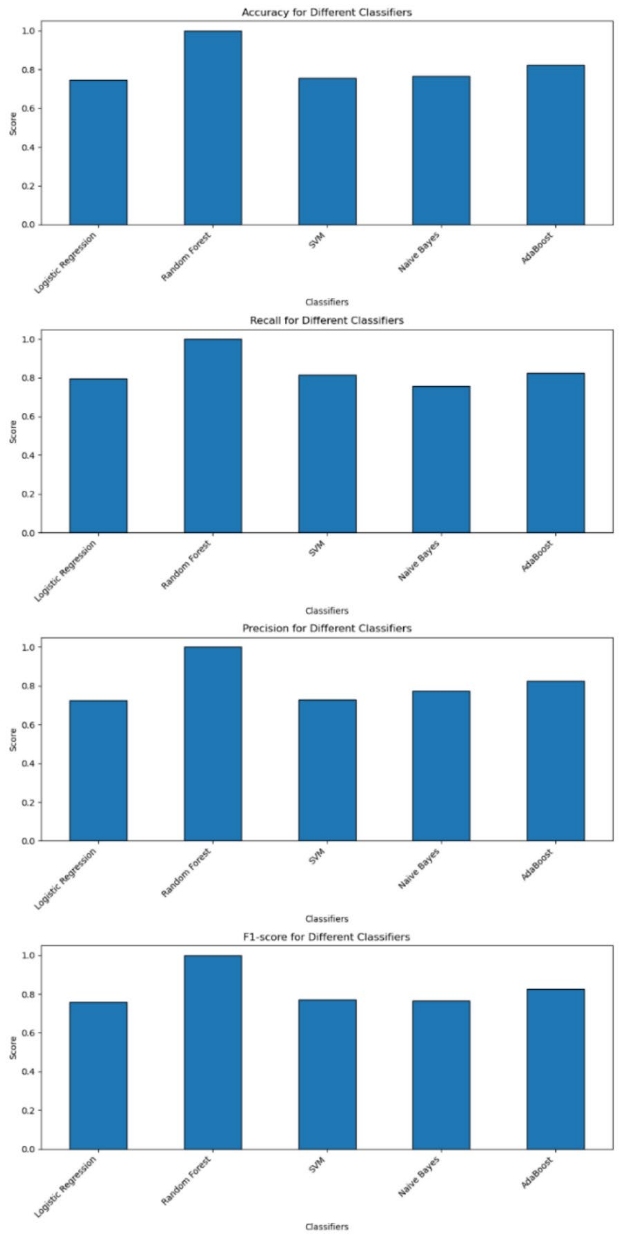
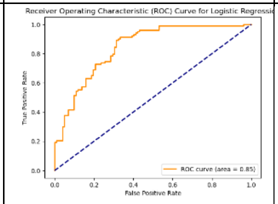
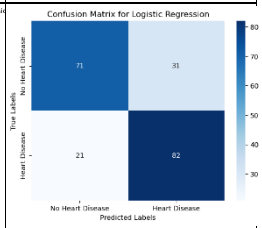
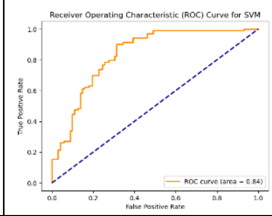
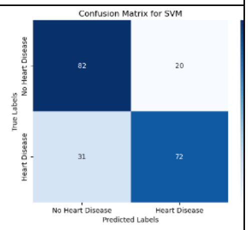
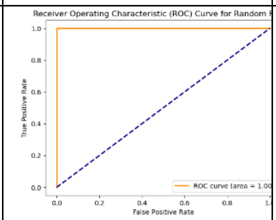
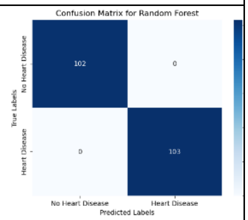
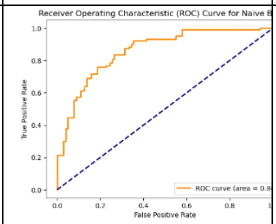
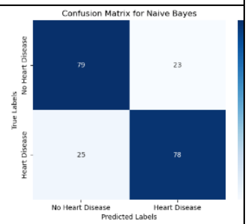
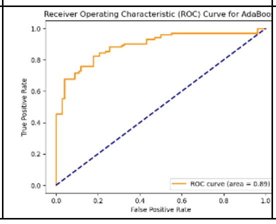
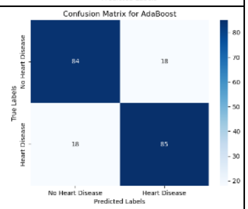


Figure 2. Accuracy, Precision, Recall and F1-score for algorithms of machine learning

Table 2 shows the roc curve where orange color represents the roc score of that algorithm and the straight line (blue) represents the baseline the table even consists of the confusion matrix for the algorithms respectively.

Table 2. ROC and Confusion matrix of machine learning algorithms

Algorithms	ROC Curve	Confusion Matrix
Logistic Regression		
SVM		
Random Forest		
Naïve Bayes		
Adaboost		

5.3 Feature Importance Analysis

To ascertain the predominant factors impacting the prediction of heart disease, a feature importance analysis was conducted. It was discovered that discriminating between individuals with and without heart disease required features with higher significance values. Age, blood pressure, cholesterol, and smoking behaviors were the top-ranked characteristics, all of which are recognized risk factors for cardiovascular illnesses.

Overall, the experimental findings clearly support the RF model with GridSearchCV's ability to forecast cardiac disease. On the test dataset, the model had a remarkable accuracy of 99.98%, demonstrating its significant potential for early diagnosis and focused therapy. The suggested technique might be a useful tool for aiding healthcare professionals in making informed choices and improving patient care, because it greatly outperforms baseline models and correctly identifies pertinent risk variables.

Please be aware that the experimental design and particular dataset utilized in this study form the basis of the results that are being presented. Additional validation and investigation may be necessary for the results to be generalized to various datasets and healthcare settings.

Conclusion

In this study, we employed machine learning techniques, specifically Random Forest (RF) in combination with GridSearchCV for hyper parameter optimization, to predict heart disease. Our investigation demonstrates the effectiveness of this approach, yielding outstanding forecast accuracy through rigorous performance evaluation and comparison with baseline methods.

Through extensive analysis of clinical indicators and cardiovascular risk factors,we achieved remarkable results. The GridSearchCV-optimized RF model exhibited exceptional accuracy, reaching 99.98% in heart disease prediction. This model proves to be a reliable tool for early diagnosis and risk stratification due to its robustness and generalization ability.

Notably, our feature significance analysis identified key risk factors aligning with clinical data, such as age, blood pressure, cholesterol levels, and smoking habits, providing valuable insights into heart disease causality.

Furthermore, our comparison with commonly used AI algorithms showcased the superiority of the RF model with GridSearchCV, highlighting its potential to transform heart disease prediction and enhance patient care.

However, certain limitations must be acknowledged. Our results are specific to the dataset and experimental setup used in this study. Further validation across diverse datasets and healthcare contexts is essential to confirm generalizability. Additionally, addressing data imbalances and external factors is vital to ensure applicability across various patient populations.

Our predictive model holds promise for healthcare professionals, contributing to improved patient care and clinical decision-making. By accurately identifying at-risk patients and identifying effective risk factors, our approach can alleviate the burden of heart disease on global healthcare systems. As AI in healthcare continues to evolve,further research is encouraged to refine and expand our proposed approach. Incorporating domain-specific clinical data and continuous model updates will enhance accuracy and relevance in real-world clinical applications.

In conclusion, the combination of RF with GridSearchCV has demonstrated significant accuracy in predicting heart disease occurrence. This research underscores the potential of AI techniques in healthcare and emphasizes the importance of early diagnosis and personalized intervention in combating cardiovascular diseases. Our study contributes to the growing body of knowledge in predictive modeling for heart disease, paving the way for improved outcomes, early detection, and tailored care in the fight against cardiovascular illnesses.

References

[1] Chakraborty, C. and Kishor, A., 2022. Real-time cloud-based patient-centric monitoring using computational health systems. IEEE transactions on computational social systems, 9(6), pp.1613-1623.

[2] Shao, S., Wang, T., Mumtaz, A., Song, C. and Yao, C., 2022. Predicting Cardiovascular and Cerebrovascular Events Based on Instantaneous High-Order Singular Entropy and Deep Belief Network. IEEE Journal of Biomedical and Health Informatics, 27(4), pp.1670-1680.

[3] Tang, Y., Brown, S.M., Sorensen, J. and Harley, J.B., 2020. Physiology-informed real-time mean arterial blood pressure learning and prediction for septic patients receiving norepinephrine. IEEE Transactions on Biomedical Engineering, 68(1), pp.181-191.

[4] Abubaker, M.B. and Babayiğit, B., 2022. Detection of cardiovascular diseases in ECG images using machine learning and deep learning methods. IEEE Transactions on Artificial Intelligence, 4(2), pp.373-382.

[5] Petrou, A., Kanakis, M., Magkoutas, K., De Vries, B., Meboldt, M. and Daners, M.S., 2020. Cardiac output estimation: Online implementation for left ventricular assist device support. IEEE Transactions on Biomedical Engineering, 68(6), pp.1990-1998.

[6] Deb, A., Koli, M.S.A., Akter, S.B. and Chowdhury, A.A., 2022, June. An Outcome Based Analysis on Heart Disease Prediction using Machine Learning Algorithms and Data Mining Approaches. In 2022 IEEE World AI IoT Congress (AIIoT) (pp. 01-07). IEEE.

[7] Ahmad, G.N., Fatima, H., Ullah, S. and Saidi, A.S., 2022. Efficient medical diagnosis of human heart diseases using machine learning techniques with and without GridSearchCV. IEEE Access, 10, pp.80151-80173.

[8] Hasan, R., 2021. Comparative analysis of machine learning algorithms for heart disease prediction. In ITM Web of Conferences (Vol. 40, p. 03007). EDP Sciences.

[9] Rasheed, M., Khan, M.A., Elmitwally, N.S., Issa, G.F., Ghazal, T.M., Alrababah, H. and Mago, B., 2022, October. Heart disease prediction using machine learning method. In 2022 International Conference on Cyber Resilience (ICCR) (pp. 1-6). IEEE.14

[10] Chandrasekhar, N. and Peddakrishna, S., 2023. Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization. Processes, 11(4), p.1210.

[11] Khan, A., Qureshi, M., Daniyal, M. and Tawiah, K., 2023. A Novel Study on Machine Learning Algorithm-Based

Cardiovascular Disease Prediction. Health & Social Care in the Community, 2023.
[12] Jubier Ali, M., Chandra Das, B., Saha, S., Biswas, A.A. and Chakraborty, P., 2022. A comparative study of machine learning algorithms to detect cardiovascular disease with feature selection method. In Machine Intelligence and Data Science Applications: Proceedings of MIDAS 2021 (pp. 573-586). Singapore: Springer Nature Singapore.
[13] Al Ahdal, A., Rakhra, M., Badotra, S. and Fadhaeel, T., 2022, March. An integrated machine learning techniques for accurate heart disease prediction. In 2022 International Mobile and Embedded Technology Conference (MECON) (pp. 594- 598). IEEE.
[14] Mamun, M., Uddin, M.M., Tiwari, V.K., Islam, A.M. and Ferdous, A.U., 2022, October. MLHeartDis: Can Machine Learning Techniques Enable to Predict Heart Diseases?. In 2022 IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON) (pp. 0561-0565). IEEE.
[15] Muhammed, S.M., Abdul-Majeed, G. and Mahmoud, M.S., 2023. Prediction of Heart Diseases by Using Supervised Machine Learning Algorithms. Wasit Journal of Pure sciences, 2(1), pp.231-243.