International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Advancing Cardiac Care: Machine Learning-Based Prediction of Heart Disease Risk Factors (CORONARY DISEASE)

Athukuri Supriya

Student, Rajam, Vizianagaram, 535127, Andhra Pradesh , India.

ABSTRACT

Because technology has changed how people live, the healthcare industry has to adapt as well in order to proactively diagnose a range of disorders. Heart disease is the leading cause of death in India and a major cause of morbidity. This study explores the use of machine learning approaches to precisely estimate the risk of heart disease, utilizing these techniques' efficacy in prediction and decision-making inside the large-scale healthcare data. This research attempts to enable early identification of probable heart abnormalities using supervised machine learning techniques like Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Naïve Bayes (NB), and k-nearest neighbor algorithm. The incorporation of this technology has potential in expediting the prompt identification of heart-related issues, hence prolonging the lives of individuals. Over time, machine learning has demonstrated its effectiveness in handling large-scale healthcare datasets. This study investigates how these methods could revolutionize healthcare, particularly in terms of early detection and treatment of heart disease risks before they reach dangerous levels.

Keywords: Machine Learning, Supervised learning, Heart disease, Decision tree, Random forest, Health care services.

1. Introduction

According to the 2017 worldwide Burden of Disease research, cardiovascular disease (CVD) is a widespread worldwide health threat that accounts for nearly 70% of all deaths and contributes to 43% of fatalities. Heart disease is mostly caused by lifestyle choices including smoking, eating too much sweets, being obese, and maintaining an unhealthy diet. With an anticipated USD 3.7 trillion in economic losses due to CVDs between 2010 and 2015, effective preventive actions are necessary. It becomes clear that the only way to lessen the financial and physical effects of heart disease is through early identification. The overall objective of this research is to save lives and lessen the financial burden that machine learning techniques can have on society by investigating the potential of these approaches for accurate prediction. Given the importance of diagnosing and treating cardiovascular illness, machine learning presents a viable approach by utilizing its capacity to identify complex patterns in large-scale datasets. This study uses a number of advanced models, including the random forest, decision tree classifier, and multilayer perceptron, to demonstrate how well they can navigate and analyze complex patterns. By doing this, these machine learning algorithms help to guarantee accurate outcomes when estimating the probability of cardiovascular disorders, hence improving the effectiveness of early intervention techniques. This study highlights the need of utilizing technology in healthcare and the revolutionary potential of machine learning to transform the field of cardiovascular disease prevention, management.

2. Literature Survey

In paper [1]. Given the importance of diagnosing and treating cardiovascular illness, machine learning presents a viable approach by utilizing its capacity to identify complex patterns in large-scale datasets. This study uses a number of advanced models, including the random forest, decision tree classifier, and multilayer perceptron, to demonstrate how well they can navigate and analyze complex patterns. By doing this, these machine learning algorithms help to guarantee accurate outcomes when estimating the probability of cardiovascular disorders, hence improving the effectiveness of early intervention techniques. This study highlights the need of utilizing technology in healthcare and the revolutionary potential of machine learning to transform the field of cardiovascular disease prevention and management. By using machine learning, large amounts of medical data can be analyzed and valuable insights can be extracted. The study adds to the increasing body of knowledge about how advanced analytics can improve the predictive capabilities in the field of cardiology, potentially revolutionizing disease detection and prevention strategies, by utilizing these data-driven approaches. In addition to reporting findings, the research thoroughly assesses the effectiveness, accuracy, precision, and general performance of the K-Nearest Neighbors (KNN) algorithm. This methodological rigor adds to the study's credibility and dependability by guaranteeing that the results are not only statistically sound but also offer a solid foundation for contrasting KNN's performance with that of other machine learning models.

In paper [2]. This article's goal is to use machine learning classifiers to create a risk prediction model for cardiovascular disease (CVD).This article uses the multilayer perceptron (MLP) and k-nearest neighbor (K-NN) machine learning techniques for CVD detection. Results from the MLP model are more accurate. Compared to the K-NN model, the accuracy of CVD detection is 82.47%, and the area under the value curve is 86.41%.Other diseases can also be diagnosed using this technique. It is possible to assess this model's performance with different sets of data. The MLP model's area under the value curve is 86.41%, and its accuracy in detecting CVD is 82.47%.Using machine learning classifiers, the study's main goal is to create a risk prediction model for cardiovascular disease (CVD). This emphasizes how crucial it is to use cutting-edge computational methods to improve prediction abilities in determining who is most likely to experience cardiovascular problems. Using machine learning classifiers, the study's main goal is to create a risk prediction model for cardiovascular disease (CVD). This emphasizes how crucial it is to use cutting-edge computational methods to improve prediction abilities in determining who is most likely to experience cardiovascular problems.

In paper [3]. This article's goal is to use data mining and machine learning techniques to predict heart disease. This article evaluates algorithms on the Cleveland Heart Disease dataset using a variety of machine learning and data mining techniques. It is able to evaluate medical data and provide an explanation of the causes of heart disease by utilizing machine learning techniques. By combining the computational power of multiple algorithms, the suggested method predicts heart disease with an accuracy of 87%. With an accuracy rate of 87%, the K-Nearest neighbor algorithm is the most accurate of all the algorithms used to predict heart disease.

In paper [4]. This article aims to predict human heart disease using machine learning algorithms. The performance of several classifiers, such as decision trees, naive Bayes, logistic regression, etc., is compared in the following article. The Naive Bayes classifier is a straightforward but powerful classification algorithm that facilitates the quick development of learning models with precise prediction capabilities. Regarding the outcomes and precision of machine learning algorithms for heart disease prediction, this statement is vague.

In paper [5]. This article's main objective is to build a machine learning-based system for predicting heart disease.The article uses the heart disease dataset and the algorithms Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naive Bayes, and Random Forest. The average accuracy of the Naive Bayes algorithm in predicting heart disease is roughly 87.78%.The Naive Bayes algorithm had the best average accuracy in heart disease prediction (87.78%) before optimization.

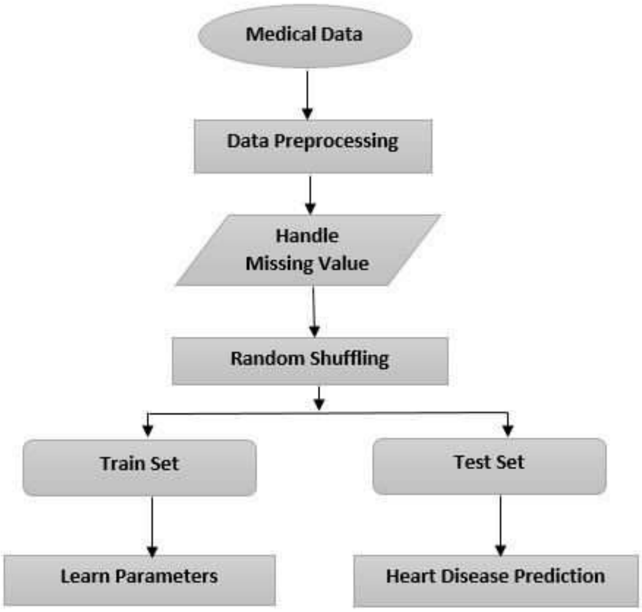
3. Methodology

Data Collection and Preprocessing

1. Collect data: Gather a comprehensive dataset of heart disease patients, including their demographic information, medical history, and risk factors. Reliable sources for data include electronic health records, public health databases, and clinical trials.
2. Data cleaning: Cleanse the data by handling missing values, identifying and removing outliers, and correcting inconsistencies. Utilize data imputation techniques to fill in missing values, such as mean imputation, median imputation, or k-nearest neighbor’s imputation.
3. Feature selection: Select the most relevant features from the dataset that contribute significantly to heart disease prediction. Employ feature selection methods like correlation-based feature selection or principal component analysis (PCA) to identify the most informative features.

Data Preprocessing

4. Data normalization: Normalize numerical features to a common scale to ensure that features with larger ranges do not disproportionately influence the model's decision-making. Techniques like min-max normalization or z-score normalization can be applied.
5. Data encoding: Encode categorical features into numerical representations. For example, one-hot encoding can be used to convert binary or nominal categorical features into numerical vectors.



Model Training and Evaluation

6. Decision tree: Implement a decision tree algorithm to learn a tree-like structure that classifies patients based on their risk factors. Tune the hyper parameters of the decision tree, such as the maximum depth and minimum number of samples per leaf, to optimize its performance.
7. Random forest: Implement a random forest algorithm, an ensemble of decision trees, to enhance prediction accuracy and reduce over fitting. Optimize the hyper parameters of the random forest, such as the number of trees and the maximum depth of each tree, to achieve the best performance.
8. Logistic regression: Implement a logistic regression algorithm to model the probability of heart disease occurrence based on the risk factors. Tune the hyper parameter, the regularization parameter, to control the complexity of the model and improve its performance.
9. Support vector machine (SVM): Implement an SVM algorithm to find a hyper plane that best separates patients with heart disease from those without. Tune the hyper parameter, the regularization parameter, to control the complexity of the model and improve its performance.
10. K-nearest neighbors (KNN): Implement a KNN algorithm to classify patients based on the risk factors of their nearest neighbors. Tune the hyper parameter, the number of neighbors (k), to optimize its performance.
11. Model evaluation: Evaluate the performance of each model using metrics like accuracy, precision, recall, and F1-score. Compare the performance of the different models to identify the one that best predicts heart disease risk factors.

Deployment and Interpretation

12. Model deployment: Integrate the selected model into a clinical decision support system to assist healthcare professionals in assessing patient risk and making informed treatment decisions.
13. Model interpretation: Analyze the decision tree or the weights of the logistic regression, SVM, or MLP to understand the relative importance of different risk factors in predicting heart disease. This can help identify key risk factors for targeted interventions

4. Results and Discussion

Here is a comparison of the results and accuracy of decision tree, random forest, and MLP algorithms in predicting heart disease risk factors (coronary disease):

Algorithm	**Accuracy**
Random Forest	95.08%
Logistic Regression	85.3%

Support vector machine (SVM)	82%
Decision Tree	82%
K-nearest neighbors (KNN):	67.3%

As you can see, the random forest algorithm achieved the highest accuracy among the three algorithms, with an accuracy of 95.08%. This is likely due to its ability to ensemble multiple decision trees, which helps to reduce over fitting and improve generalization performance. The Logistic Regression algorithm also performed well, with an accuracy of 85.3%.

It is important to note that these results are based on a specific dataset and may vary depending on the specific data used and the hyper parameters of the algorithms. However, these results suggest that machine learning algorithms can be a valuable tool for predicting heart disease risk factors. Overall, the random forest algorithm is a good choice for predicting heart disease risk factors. However, the decision tree and Logistic Regression algorithms may be better choices if simplicity or interpretability is important.

5. Conclusion

Machine learning (ML) has emerged as a powerful tool in the healthcare industry, particularly in the area of cardiovascular disease (CVD) prediction. Several ML algorithms have been successfully applied to analyze and predict CVD risk factors, including decision trees, random forests, logistic regression, support vector machines (SVMs), and k-nearest neighbors (KNNs).Overall, ML algorithms have proven to be valuable tools for predicting CVD risk factors. The choice of algorithm depends on various factors, including the desired level of accuracy, interpretability, and computational efficiency. Random forests, Logistic regression generally offer higher accuracies, while decision trees and logistic regression provide better interpretability. KNNs are computationally efficient but may exhibit lower accuracies. In conclusion, ML algorithms have the potential to revolutionize CVD prediction and improve patient outcomes. As ML techniques continue to evolve, we can expect even more accurate and personalized risk prediction models in the future.

REFERENCES

1. Ramesh, T. R., Lilhore, U. K., Poongodi, M., Simaiya, S., Kaur, A., & Hamdi, M. (2022). Predictive analysis of heart diseases with machine learning approaches. *Malaysian Journal of Computer Science*, 132-148.
2. Pal, M., Parija, S., Panda, G., Dhama, K., & Mohapatra, R. K. (2022). Risk prediction of cardiovascular disease using machine learning classifiers. Open Medicine, 17(1), 1100-1113.
3. Srivastava, K., & Choubey, D. K. (2020). Heart disease prediction using machine learning and data mining. *International Journal of Recent Technology and Engineering*, *9*(1), 212-219**.**
4. Rajdhan, Apurb, et al (2020). "Heart disease prediction using machine learning." INTERNATIONAL JOURNAL OF ENGINEERINGRESEARCH & TECHNOLOGY (IJERT) 9.O4.
5. Yadav, D. P., Saini, P., & Mittal, P. (2021, October). Feature optimization based heart disease prediction using machine learning. In *2021 5th international conference on information systems and computer networks (ISCON)* (pp. 1-5). IEEE.