1st International Conference on Trends in Engineering Systems and Technologies ICTEST-2024

Comparative Analysis of Machine Learning Algorithms for Cardiovascular Disease Risk Prediction

Paper ID: 129

PRESENTER NAME: Anushka Doiphode

Affiliation: Pune Institute of Computer Technology

Overview

- Introduction
- Literature review
- Motivation and Objective
- Work
- Results
- Conclusion and Future Scope
- Q&A

Introduction

Cardiovascular disease (CVD) is a leading cause of morbidity and mortality globally, posing a significant burden on healthcare systems. Predicting and preventing CVD are crucial in combating these life-threatening conditions. According to the World Health Organization, heart disease is the most common cause of death worldwide, accounting for 17.9 million deaths annually. Current methods for diagnosing and predicting heart problems rely heavily on medical history, symptoms, and physical examinations, often leading to inaccuracies. An automated intelligent system is needed for more accurate determinations. Various machine learning algorithms, such as Artificial Neural Networks, Gradient Boosting machines, K Nearest Neighbor, Decision Tree, Random Forest, and Support Vector Machine, are being used to predict different types of cardiovascular diseases, including CAD, Heart failure, Arrhythmia, hypertension, strokes, and PAD.

Literature Review

Various studies highlight the potential of AI and machine learning (ML) in improving cardiovascular disease (CVD) prediction and management. These technologies can enhance risk assessment, early detection, and personalized treatment. AI models, when combined with data sources like patient demographics and clinical measurements, show promise in accurately predicting CVD events. Additionally, AI can process vast amounts of patient data, improving diagnostic and treatment decisions. However, there are challenges, such as biases in risk assessment models, which require transparent validation processes and interdisciplinary collaboration. Despite these challenges, AI and ML have the potential to transform cardiovascular care, leading to more accurate diagnoses and treatment strategies.

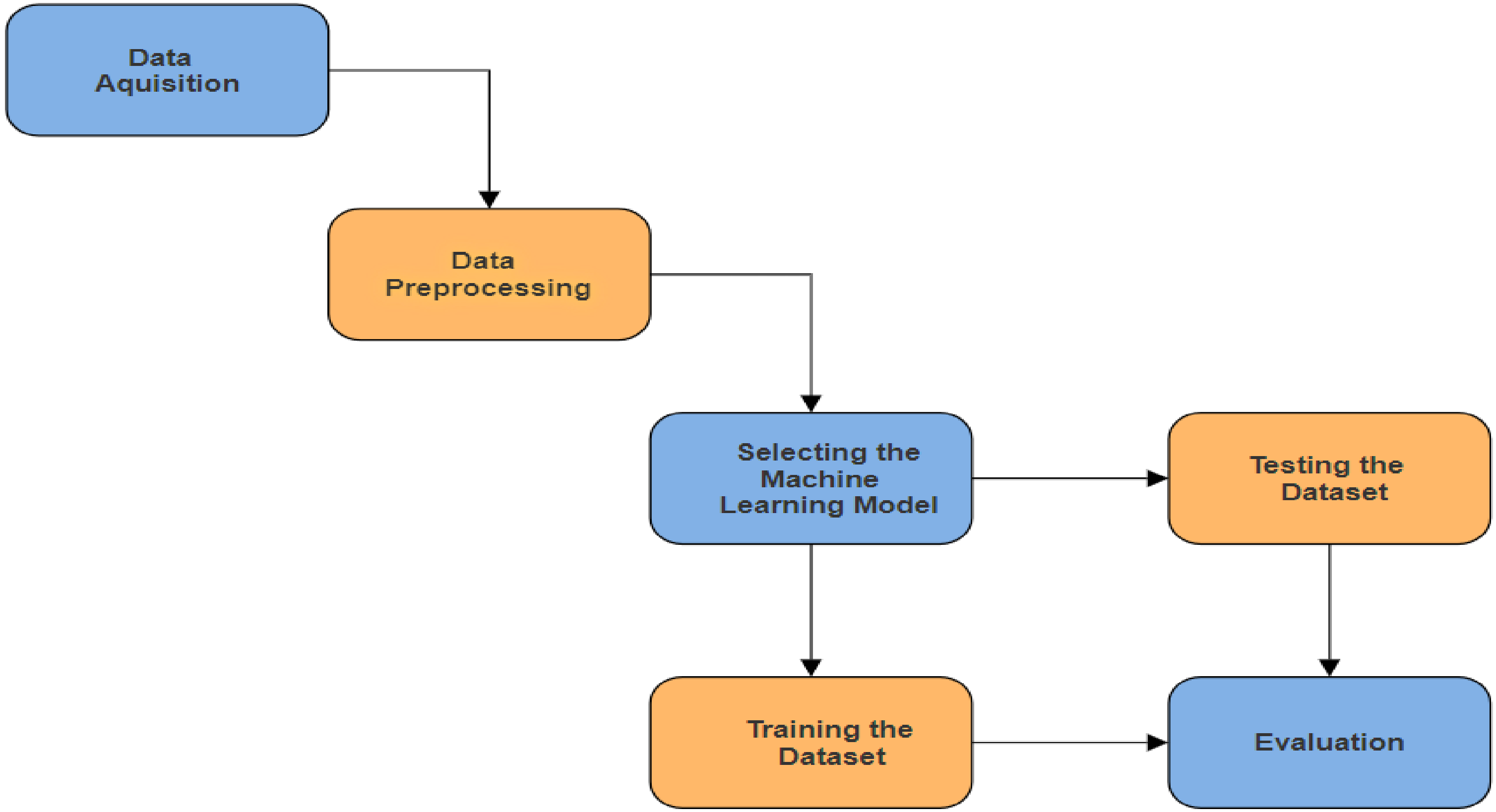
Motivation

This paper was motivated by the identified gaps in existing research on the application of AI and machine learning in cardiovascular disease (CVD) prediction and management. Therefore, this paper aims to fill these gaps by providing a comprehensive review and analysis of current literature, comparing four algorithms to provide a view of the efficiency of each one on a dataset, and emphasizing the need for more focused research and practical solutions in utilizing AI to enhance CVD prediction and management.

Objective

The objective of this paper is to address the gaps in existing research on the application of AI and machine learning in cardiovascular disease (CVD) prediction and management. Specifically, the paper aims to compare the efficiency of four algorithms on a dataset to provide insights into their performance.

Proposed Work/ Design etc.



Proposed Work/ Design (Contd..)

The dataset, sourced from Kaggle, undergoes crucial data preprocessing to enhance the accuracy of the machine learning model. This includes cleaning the data by addressing missing, noisy, and inconsistent data, removing duplicates, and selecting features for training and testing. The four algorithms K Nearest Neighbor (KNN), Decision Tree, Random Forest, and Support Vector Machine (SVM) are compared to ascertain the optimal accuracy among various options. The dataset is split into distinct training and testing sets, and the sci-kit learn library is used to implement and fine-tune machine learning models. This ensures the dataset is refined and prepared for the application of classification algorithms, aiming to improve the precision and effectiveness of the model in making accurate predictions.

Results

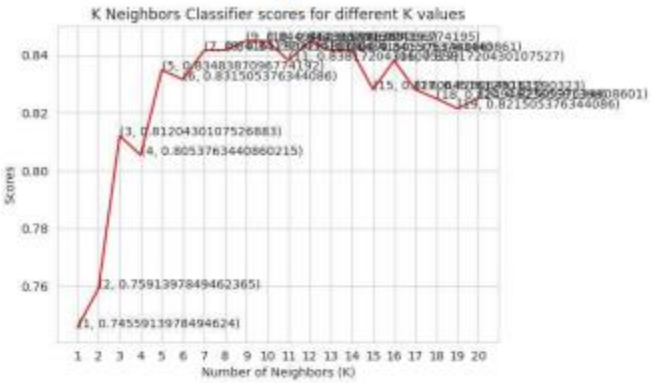


Fig. 7. KNN Distribution

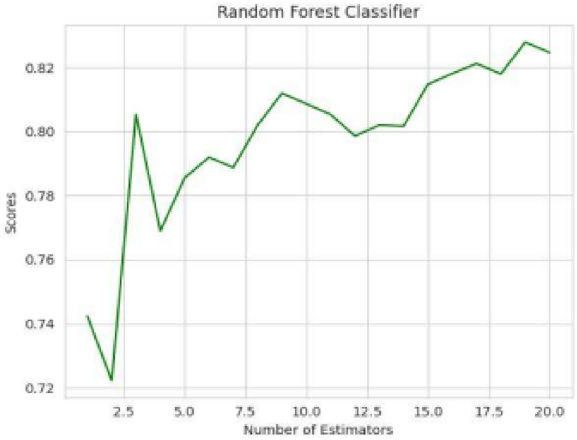


Fig. 8. Random Forest Distribution

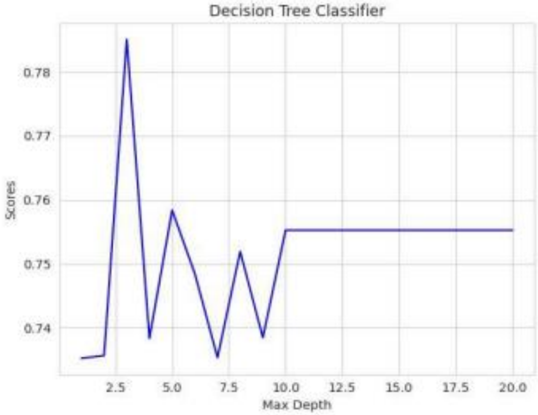


Fig. 9. Decision Tree Distribution

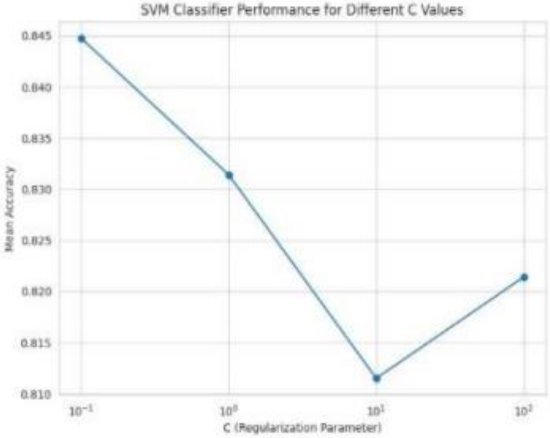


Fig. 10. SVM Distribution

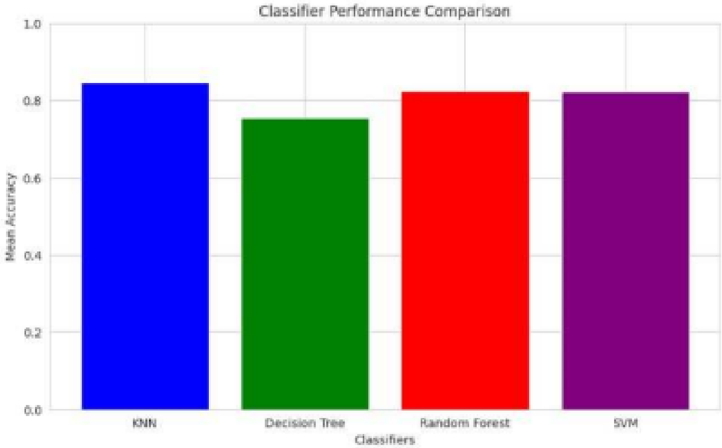


Fig. 12. Bar Graphs

After comparing the four algorithms, it is seen that KNN is the most efficient algorithm giving an 84.4% efficiency.

Conclusion

In our comprehensive study focused on cardiovascular disease risk prediction, we rigorously evaluated the performance of diverse machine learning classifiers, including K-Nearest Neighbors (KNN), Random Forest, Decision Tree, and Support Vector Machine (SVM). Notably, the results revealed that KNN consistently outperformed its counterparts, exhibiting the highest accuracy. This conspicuous success can be attributed to KNN's exceptional ability to discern and exploit intricate patterns inherent in our dataset. By leveraging the proximity of data points, KNN achieved a level of precision in risk prediction that surpassed our expectations. The algorithms mentioned above have some limitations, the future scope includes some advanced ML techniques such as Gradient Boosting Machines (GBM) which uses multiple decision trees sequentially, CNN’s and RNN’s for image and text analysis, SVM with kernels, Bayesian methods, etc.

Future scope

Future research in the application of AI and machine learning in cardiovascular disease (CVD) prediction and management could focus on exploring deep learning models for improved feature extraction, ensemble methods for combining algorithm strengths, data augmentation for increased dataset diversity, and methods to enhance model interpretability for clinician trust.

References

[1] Heart disease prediction using machine learning techniques: A survey (2018). International Journal of Engineering & Technology.
[2] Reddy Prasad,Pidaparthi Anjali, S.Adil, N.Deepa. Heart Disease Prediction using Logistic Regression Algorithm using Machine Learning. ISSN: 2249 – 8958, Volume8, Issue-3S, February.
[3] Physician-Friendly Machine Learning: A Case Study with Cardiovascular Disease Risk Prediction. Journal of Clinical Medicine, 8(7), 1050.
[4] Kagiyama, N., Shrestha, S., Farjo, P. D., & Sengupta, P. P. (2019). Artificial Intelligence: Practical Primer for Clinical Research in Cardiovascular Disease. Journal of the American Heart Association, 8, e012788.
[5] Prasadgouda B Patil, Dr. P Mallikarjun Shastry, Dr Ashokumar P S (2020). Machine Learning Based Algorithm for Risk Prediction of Cardiovascular Disease (CVD).
[6] Maarten van Smeden, Georg Heinze, Ben Van Calster, Folkert W Asselbergs, Panos E Vardas, Nico Bruining, Peter de Jaegere, Jason H Moore, Spiros Denaxas, Anne Laure Boulesteix, Karel G M Moons (2022). Critical appraisal of artificial intelligence-based prediction models for cardiovascular disease.
[7] Pankaj Mathur, Shweta Srivastava, Xiaowei Xu, and Jawahar L Mehta (2020). Artificial Intelligence, Machine Learning, and Cardiovascular Disease.
[8] Jasjit S. Suri, Mrinalini Bhagawati, Sudip Paul, Athanasios Protogeron, Petros P. Sfikakis, George D. Kitas, Narendra N. Khanna, Zoltan Ruzsa, Aditya M. Sharma, Sanjay Saxena, Gavino Faa, Kosmas I. Paraskevas, John R. Laird, Amer M. Johri, Luca Saba, Manudeep Kalra (2022). Understanding the bias in machine learning systems for cardiovascular disease risk assessment: The first of its kind review.
[9] Yahia Baashar, Gamal Alkawsi, Hitham Alhussian, Luiz Fernando Capretz, Ayed Alwadain, Ammar Ahmed Alkahtani, Malek Almomani (2022). Effectiveness of Artificial Intelligence Models for Cardiovascular Disease Prediction: Network Meta-Analysis.
[10] Mohammad Moshawrab, Mehdi Adda, Abdenour Bouzouane, Hussein Ibrahim, Ali Raad (2022). Cardiovascular Events Prediction using Artificial Intelligence Models and Heart Rate Variability.
[11] Madhumita Pal, Smita Parija, Ganapati Panda, Kuldeep Dhama and Ranjan K. Mohapatra (2022). Risk prediction of cardiovascular disease using machine learning classifiers.
[12] Chayakrit Krittanawong, Hafeez Ul Hassan Virk, Sripal Bangalore, Zhen Wang, Kipp W. Johnson, Rachel Pinotti, HongJu Zhang, Scott Kaplin, Bharat Narasimhan, Takeshi Kitai, Usman Baber, Jonathan L. Halperin & W. H. Wilson Tang (2020). Machine learning prediction in cardiovascular diseases: a meta-analysis.

Thank You

Questions?