iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** VII **Month of publication:** July 2025

DOI: https://doi.org/10.22214/ijraset.2025.73071

Hybrid Learning Model for Cardiovascular Risk Assessment

Sonika Harshitha[1], Dr. Sankhya N Nayak[2]
[1]Research Scholar, Dept. of CS&E, JNNCE,Shimoga, India
[2]Associate Prof., Dept. of CS&E, JNNCE,Shimoga, India

Abstract: Cardiovascular diseases constitute a significant global health concern, demanding accurate predictive tools for timely intervention. This study addresses this imperative by employing a hybrid machine learning approach, integrating Random Forest and Linear Regression techniques for forecasting heart diseases. The complexity of cardiovascular dynamics requires a nuanced methodology, and this research endeavors to enhance forecasting precision without predefining the model structure. Leveraging a diverse patient dataset encompassing age, cholesterol levels, blood pressure, and lifestyle factors, the hybrid framework is trained on historical data, leveraging Random Forest for intricate relationships and Linear Regression for interpretability. Through meticulous feature importance analysis, the study unveils key factors influencing heart disease prognosis, offering valuable insights for personalized healthcare initiatives. This hybrid methodology, without reliance on a predetermined model, holds promise for improving early prediction strategies and contributing to more effective interventions in cardiovascular health.
Keywords: Linear Regression (LR), Random Forest (RF).

I. INTRODUCTION

Cardiovascular diseases, including heart disease, remain a significant global health concern, responsible for a substantial number of deaths and a considerable economic burden. The ability to predict and understand the risk factors associated with heart disease is crucial for early intervention and prevention. Machine learning techniques have emerged as powerful tools in this endeavor, enabling the forecasting of heart disease based on patient characteristics. Two commonly employed methods in this domain are linear regression and random forest, each offering its unique advantages.

Linear Regression: Unveiling Linear Relationships Linear regression is a well-established statistical technique that aims to establish a linear relationship between independent variables, such as age, cholesterol levels, and blood pressure, and the dependent variable, which in case the risk or likelihood of heart disease. Linear regression provides a straightforward and interpretable model for understanding the impact of individual features on heart disease prediction.

Random Forest: Capturing Complexity In contrast, random forest is a machine learning algorithm that takes a more sophisticated approach. It is an ensemble method that combines multiple decision trees to create a more complex and robust predictive model. Decision trees are excellent at capturing non-linear relationships and interactions among features, making them particularly well-suited for healthcare data where such complex associations often exist. By aggregating the predictions of numerous individual decision trees, random forest can provide a comprehensive assessment of heart disease risk, accounting for the intricate relationships that may not be apparent through linear regression.

Early detection and proactive measures can reduce the impact of heart disease and improve overall heart.

In the ever-evolving realm of healthcare, the challenges posed by the detection and management of cardiovascular disease are multifaceted, encompassing an array of risk factors from elevated blood pressure to abnormal pulse rates. The imperative for precise decision-making and optimal treatment in addressing cardiac risk underscores the pressing need for innovative solutions. As the relentless march of machine learning technology unfolds, a paradigm shift looms on the horizon, promising transformative changes in clinical practices within the healthcare industry.

This research emerges as a guiding light in this transformative landscape, acknowledging the pivotal role of machine learning techniques in reshaping the trajectory of cardiovascular disease detection. The primary aim is unequivocal: to propose a highly accurate machine learning-based cardiovascular disease prediction system. Within this pursuit, a diverse array of modern machine learning algorithms, including REP Tree, M5P Tree, Random Tree, Linear Regression [1], Naive Bayes, J48, and JRIP, takes center stage, collectively representing a comprehensive approach to predictive modeling.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VII July 2025- Available at www.ijraset.com

The proposed Cardiovascular Disease Prediction System (CDPS) undergoes meticulous evaluation, employing various metrics to discern the most effective machine learning model. In this rigorous assessment, the Random Tree model emerges as a standout performer in predicting cardiovascular disease patients, boasting a remarkable accuracy of 100%, the lowest Mean Absolute Error (MAE) at 0.0011, the lowest Root Mean Squared Error (RMSE) at 0.0231, and an impressive prediction time of 0.01 seconds. This success not only underscores the potential of machine learning to redefine cardiovascular disease prediction but also envisions a future where precision and expediency in diagnosis take center stage. As the healthcare industry stands at the threshold of this technological evolution, researchers and clinicians must recognize and harness the transformative power of machine learning for the enhancement of cardiovascular care. The Random Tree model's admirable performance in this dynamic landscape heralds a new era in predictive healthcare analytics, promising more accurate, timely, and effective management of cardiovascular disease.

Heart disease, a pervasive global health concern, faces a transformative prospect with the integration of machine learning. This paper delves into the profound impact of machine learning on cardiovascular disease, capitalizing on the abundant health data flooding modern healthcare sectors. Leveraging this technological shift, the research pioneers a method utilizing machine learning techniques to pinpoint significant features, enhancing predictive accuracy. The heart of this innovation lies in a prediction model employing diverse machine learning algorithms, achieving an impressive 92% accuracy through the collaborative prowess of random forest [2] and a linear model. This union showcases machine learning's potential to revolutionize our comprehension and management of heart disease.

Beyond numerical success, this groundbreaking approach promises to reshape cardiovascular disease prediction, representing a significant advance at the intersection of machine learning and healthcare. It signifies a stride toward a future where technology not only enhances our understanding but also empowers us to combat pervasive health concerns with newfound hope and precision. As we navigate this innovative crossroads, the amalgamation of machine learning and healthcare emerges as a beacon, refining predictions and ushering in a new era in our ongoing battle against cardiovascular disease.

II. COMPARISON AMONG MODELS

We compared the work based on metrics, algorithms used and the accuracy on several datasets used by the authors, The work is summarized as shown in Table 1

Reference	Dataset	Tools used	Machine Learning Algorithms Used	Evaluation Metrics	Outcomes of Outperformed Algorithms
Rajkumar Gangappa et al. [1] 2023	Cleveland Heart Dataset	WEKA (implied)	SMO, Bagging + Logistic Regression, others	Accuracy, ROC	SMO – Accuracy: 86.47%; Bagging + LR – ROC: 0.91
Kompella Sri Charan et al. [2] 2022	UCI Heart Dataset	Python (implied)	Random Forest	Accuracy	Random Forest – Best accuracy (exact value not given); good early prediction
Karna Vishnu Vardhana Reddy et al. [3] 2021	Combined dataset (CVD related)	Not specified	RFBM, DTBM, KNNBM, ABBM, GBBM	Accuracy, Sensitivity, Precision, F1-score, Error rate	RFBM with Relief – Highest accuracy (value not given)
M. Snehith Raja et al. [4] 2021	Cleveland Heart Dataset	Python + QPSO (optimization)	SVM + QPSO	Accuracy	QPSO-SVM – Accuracy: 96.31%; avoids local minima

Pronab Ghosh et al. [5] 2021	Korean NHIS Dataset (9,398 records)	Python, XGBoost package	XGBoost, Gradient Boosting, Random Forest	Accuracy	XGBoost / GBoost / RF – High accuracy; strong for national data
E.I Elsidimy et al. [6] 2023	Korean NHIS Dataset	Python	Same as above: XGBoost, Gradient Boosting, RF	Accuracy	Confirms 2.5 findings; similar results
Nadikatla Chandrashekar et al. [7] 2023	Not specified (multi-centre)	MATLAB / Python (implied)	ANFIS, SVR, Neural Networks	Accuracy, Precision, F1	ANFIS – Accuracy: 96.56% (training phase)
Joung Oak Kim et al. [8] 2021	UCI Benchmark Dataset (14 features)	Python (sklearn)	RF, SVM, NB, DT	Accuracy	Random Forest – Best performer; fast & efficient
Osman Taylan et al. [9] 2023	Kaggle CVD Dataset (70,000 records)	Python + GridSearchCV	MLP, XGBoost, RF, DT	Accuracy, AUC	MLP – Accuracy: 87.28%, AUC: 0.94–0.95
Vijetha Sharma et al. [10] 2021	Large patient datasets (not specified)	Python	Multiple ML models	Accuracy, Interpretability	Highlights ML's potential but notes bias, interpretability, and privacy issues

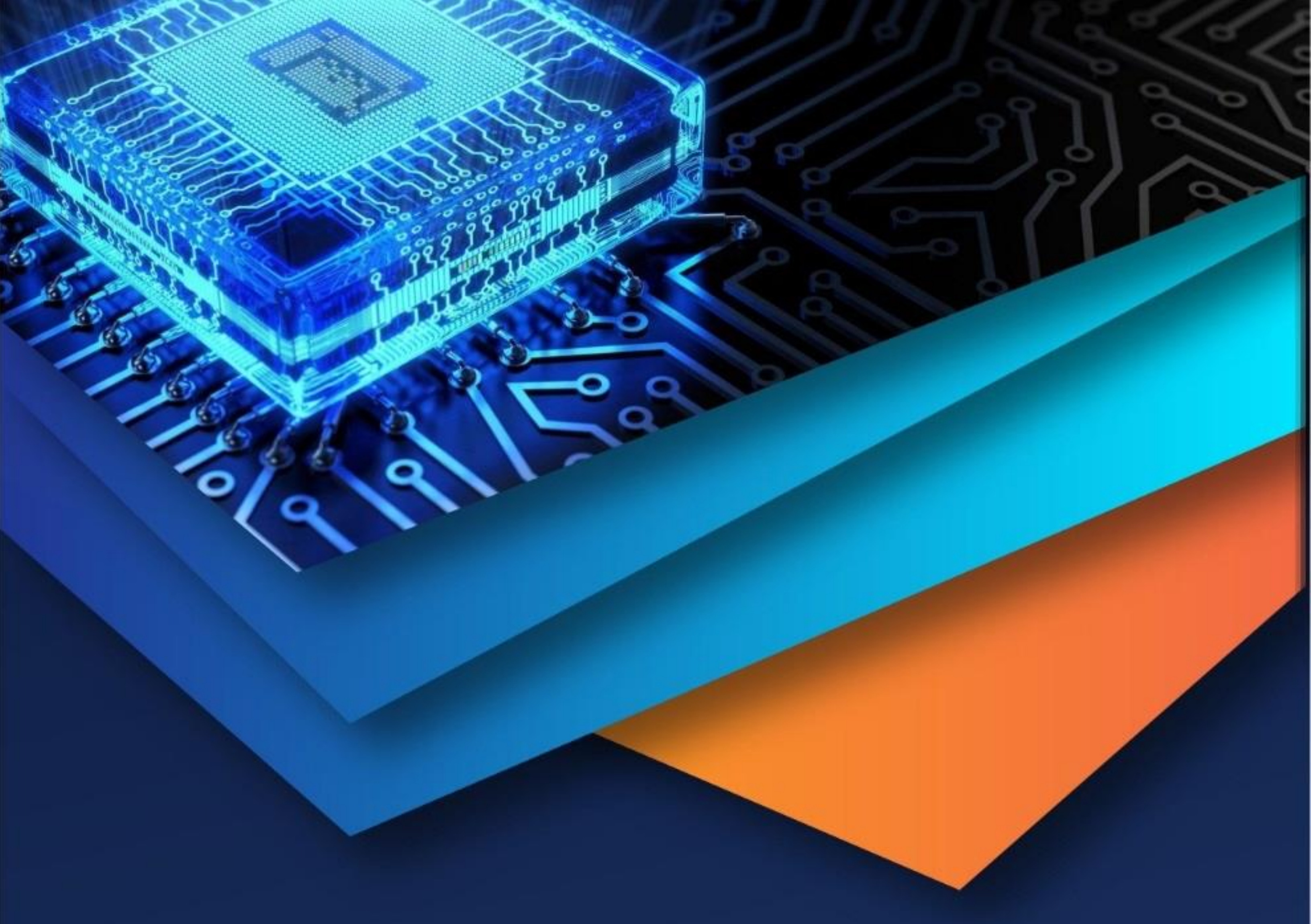
III. CONCLUSION

The Prediction of heart disease presents a critical challenge given its complex nature and the involvement of multiple risk factors with nonlinear relationships. While algorithms like random forest can address complexity, there is a risk of overfitting, and linear regression might oversimplify the issue. The solution suggests leveraging the strengths of different algorithms through a combination approach. This strategy aims to enhance predictive accuracy, offering a more nuanced understanding of the intricate relationships in heart diseases while mitigating the limitations of individual models. Ultimately, this integrative approach holds promise for improving early detection and treatment of heart conditions, marking a step forward in addressing the multifaceted challenges posed by cardiovascular diseases.

REFERENCES

[1] Rajkumar Gangappa Nadakunamani, Reyana, Sandeep Kautish,A.S. Vibith,Yogita Gupta,Sayed F.Abdelwahab and Wagdy Mohamed "Clinical Data Analysis for Prediction of Cardiovascular Disease Using Machine Learning Techniques".2023
[2] Kompella Sri Charan and Kolluru S S N S Mahendra Nath,"Heart Disease Prediction Using Random Forest Algorithm".vol 09, 2022
[3] Karna Vishnu Vardhana Reddy, Irraivan Elamvazuthi, Azrina Abd Aziz, Sivajothi Paramasivam, Hui Na Chua and S. Pranavanand, "Heart Disease Risk Prediction Using Machine Learning Classifiers with Attribute Evaluators",MDPI, 2021.
[4] M. Snehith Raja, M. Anurag, Ch. Prachetan Reddy, Nageswara Rao Sirisala. "Machine Learning Based Heart Disease Prediction System". In proceeding of International Conference on Computer Communication and Informatics, IEEE, pp. 12-28, 2021.
[5] Pronab Ghosh, Sami Azam, Mirjam Jonkman, Member, IEEE, Asif Karim, F.M.Javed Mehedi Shamrat, Eva Ignatious, Shahana Shultana, Abhijith Reddy Beeravolu and Friso De Boer, "Efficient prediction of cardiovascular disease using machine learning algorithms with Relief and LASSO feature selection techniques", vol.9, pp. 19304-19326, IEEE, 2021
[6] E. I. Elsedimy1 · Sara M. M. AboHashish1 · Fahad Algarni ,"New cardiovascular disease prediction approach using support vector machine and quantum behaved particle swarmoptimization",Springer,2023
[7] Nadikatla Chandrashekar and Samineni Peddakrishna, "Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization", MDPI,2023.

[8] Joung ouk(Ryan) Kim, Yong-Suk Jeong, Jin Ho Kim, Jong Weon Lee, Dougho Park and Hyoung-Seop Kim, “Machine Learning-Based Cardiovascular Disease Prediction Model: A Cohort Study on the Korean National Health Insurance Service Health Screening Database”,MDPI,2021.
[9] Osman Taylan, Abdulaziz S. Alkabaa, Hanan S. Alqabbaa, Esra Pamukcu and Victor Leiva, “Early Prediction in Classification of Cardiovascular Diseases with Machine Learning, Neuro-Fuzzy and Statistical Methods”,MDPI,2023.
[10] Vijeta Sharma, Shrinkhala yadav,Manjari Gupta "Heart disease prediction using Machine Learning Techniques", International Conference on Advances in Computing, Communication Control and Networking (ICACCCN),IEEE, 9362842,2021.



10.22214/IJRASET

INDEX COPERNICUS
45.98

ISRA JIF
IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY