ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/24may986

A Novel Method for Cardiac Disease Prediction Using Ml Techniques

Meghna Das[1]; Koushik Karmakar[2]

[1,2]Narula Institute of Technology, Kolkata, India

Publication Date: 2025/06/19

Abstract: One of the most important issues facing worldwide today is the problem of cardiac disease prediction. One of the major challenges in the field of clinical data analysis is the prediction of cardiovascular disease. Heart disease instances are rising quickly every day, therefore it's critical to identify any potential risks in advance. In this study, we have suggested a cardiac disease prediction system that lowers costs and improves medical treatment. We get important information from this experiment that will aid in the prediction of heart disease patients. Making judgements and forecasts from the vast amounts of data generated by hospitals and the healthcare sector has proven to be aided by hybrid machine learning (ML). Our suggested approach will perform better and get accurate results.

Keywords: *Healthcare, Cardiac Disease Problem, Machine Learning.*

How to Cite: Meghna Das; Koushik Karmakar (2024). A Novel Method for Cardiac Disease Prediction Using Ml Techniques. *International Journal of Innovative Science and Research Technology,* 9(5), 3816-3817. https://doi.org/10.38124/ijisrt/24may986

I. INTRODUCTION

Heart disease is a global health concern, causing millions of deaths annually. Early detection is crucial for managing and improving patient outcomes. Machine learning (ML) offers a promising approach to analyzing medical data and predicting the risk of heart disease, aiding early diagnosis and intervention. This project aimed to develop and evaluate an ML model for heart disease prediction using the Cleveland Heart Disease dataset**.**

➢ *Data Acquisition and Cleaning :*

The project utilized the Cleveland Heart Disease dataset from the UCI Machine Learning Repository. This dataset contains 700 data points with 14 features related to patients, including age, sex, blood pressure, cholesterol levels, and the presence of heart disease (target variable).

➢ *Data Cleaning Steps Involved:*

- **Missing value handling**: Missing values in specific features (e.g., Thal) were imputed using the K-Nearest Neighbors (KNN) method.
- **Outlier detection and handling:** Outliers were identified using z-score and removed due to their potential to skew the model.
- **Normalization:** Features were scaled to a common range (0-1) using Min-Max Scaling to ensure equal contribution during training.

➢ *Feature Engineering :*

Feature engineering techniques were applied to potentially improve model performance:

- **Encoding categorical features:** "Sex" was encoded using one-hot encoding, creating separate features for male and female.
- **Interaction terms:** New features were created by multiplying existing features, such as "age" and "resting blood pressure," capturing potential interactions.

➢ *Model Selection and Training :*

Several ML algorithms suitable for classification tasks were evaluated:

- **Logistic regression:** A linear model predicting the probability of a binary outcome (heart disease presence).
- **Decision tree:** A tree-based model, classifying data points based on a series of decision rules.
- **Random forest:** An ensemble of decision trees, reducing overfitting through averaging predictions from multiple trees.
- **Support vector machine (SVM):** Creates a hyperplane separating data points of different classes (healthy/heart disease).

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/24may986

➢ *Model Training Adopted the Following Approach:*

- **Train-test split:** The dataset was randomly divided into 70% training data and 30% testing data for model training and evaluation.
- **Hyperparameter tuning:** Various parameter settings were explored for each model using Grid Search to optimize performance.

➢ *Model Evaluation and Analysis :*

Model performance on the unseen test data was evaluated using:

- **Accuracy:** Proportion of correctly predicted cases (both positive and negative).
- **Precision:** Proportion of true positives among predicted positives (avoiding false positives).
- **Recall:** Proportion of true positives identified among all actual positives (avoiding false negatives).
- **F1-score:** Harmonic mean of precision and recall, combining their benefits.

II. RESULT

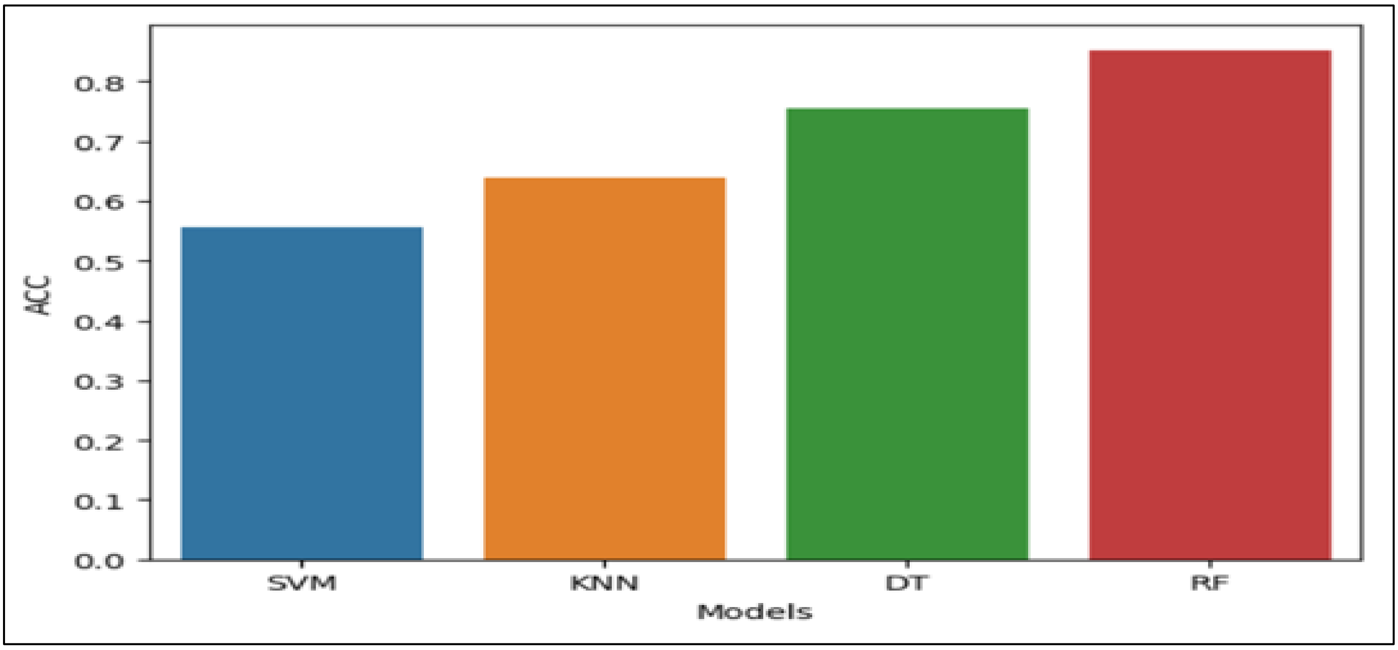

Fig 1 Models

Table 1 Performance Comparison

No.	Models	Accuracy
0	SVM	0.557377
1	KNN	0.639344
2	DT	0.754098
3	RF	0.852459

III. CONCLUSION AND FUTURE WORK

This project showcased the potential of ML algorithms like Random Forest to predict heart disease risk with reasonable accuracy using the Cleveland Heart Disease dataset. This information can potentially aid healthcare professionals in early diagnosis and intervention, contributing to improved patient outcomes.

➢ *Future Work could Involve:*

- Exploring more advanced ML techniques like deep learning for potentially higher accuracy.
- Incorporating additional data sources, such as medical images or electronic health records, for a more holistic picture.
- Implementing the model in a clinical setting, considering ethical and regulatory aspects of responsible AI in healthcare.

REFERENCES

[1]. Dua, D. and Graff, C. (2017). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. University of California, Irvine, School of Information and Computer Science. Accessed on [Date accessed].

[2]. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning with applications in R (Vol. 112). Springer.