INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 11, Issue 2 - V11I2-1364)
Available online at: https://www.ijariit.com

Heart Disease Prediction

Aniruddha Ambre
aniruddha.ambre@somaiya.edu
SK Somaiya University, Mumbai, Maharashtra

ABSTRACT

Heart disease remains one of the leading causes of mortality globally. new diagnosis gets importantly better endurance rates and cuts discourse costs. In the research, we explore Machine learning techniques to predict heart disease based on clinical Information. exploitation associate in nursing open-source dataset we apply and value respective sorting Procedures, including logistical regression, decision trees, support vector machines (SVM). Our results demonstrate that machine learning can effectively identify potential heart disease cases, providing a promising tool for healthcare Uses.

Keywords: *Heart Disease Prediction, Machine Learning, Clinical Data, Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), Feature Selection, Data Pre-processing, Hyperparameter Tuning, UCI Heart Disease Dataset, Model Evaluation, Accuracy / Precision / Recall / F1-Score, ROC-AUC, Healthcare Analytics, Non-Invasive Diagnosis, Risk Assessment, Predictive Modeling, Ensemble Methods, Cross-Validation*

INTRODUCTION

Heart disease encompasses a range of conditions that affect the heart's , Including coronary artery disease arrhythmias and heart failure. Identifying high-risk patients early is crucial for effective management. Traditional diagnostic methods rely on clinical expertise and invasive Checks but they can be time-consuming and costly. Machine learning offers amp non-invasive data-driven approach to predict heart disease using patient data.

This report investigates the feasibleness of Machine learning Procedures inch predicting Harth disease and compares their operation along amp publically free data set

Background and Related Work

Machine learning has gained traction in healthcare due to its ability to Examine Complicated data sets and uncover Layouts. respective studies bear explored its diligence inch predicting heart disease:

Framingham Heart study: researchers old logistical regress to call cardiovascular risks founded along Goal and demographic Information

Deep learning approaches: Recent works employ neural networks for high-dimensional ekg point Method comparative studies: Procedures such as Random Forest and SVM have shown promise in detecting heart conditions with high accuracy.

Model and important Concepts

1.Model Used

Logistic Regression: A simple yet effective bas for binary classification.

Decision Trees: A rule-based Representation that splits Information iteratively to classify samples.

Random Forest: An ensemble of decision trees reducing overfitting and improving Precision.

Support Vector Machine (SVM): Finds the optimal hyperplane to separate classes in high-dimensional space.

IMPORTANT CONCEPTS

Characteristic Selection: Identifying difficult Characteristics such as age cholesterol level and blood pressure.

Data Pre-processing: Handling missing values scaling and encoding categorical Information.

Evaluation Metrics: Using accuracy ,precision recall and F1-score to compare Representations.

Implementation

Dataset

The Data set used is sourced from the UCI Heart Disease Storage. it includes 303 instances with cardinal attributes such as age, chest pain type resting blood pressure, and cholesterol levels.

information pre processing handled lost values away mean/mode imputation normalized perpetual variables for break Check Effectiveness one-hot encoded flat Characteristics

model training and evaluation

Split information into 80% education and 20% examination subsets performed hyperparameter tuning exploitation gridiron look for Random Forest and support vector machine(svm)
Evaluated models using a confusion matrix, ROC-AUC curve, and cross-validation.

RESULTS

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	85.7%	83.0%	84.2%	83.6%
Decision Tree	82.1%	80.2%	81.4%	80.8%
Random Forest	**88.3%**	85.4%	86.7%	86.0%
SVM	84.5%	82.7%	83.9%	83.3%

Logistic Regression
Accuracy: 85.7%
Logistic Regression is a baseline classification Model that calculates the probability of a binary outcome (heart disease or not). inch this suit it right expected 857% of the instances
precision: 830%
precision is the part of bold predictions that are extremely right. For Logistic Regression 83.0% of the time when it predicted heart disease the prediction was correct.
Recall: 84.2%
Recall measures the ability of the Representation to correctly identify positive cases (true positives). inch this suit it right known 842% of complete the fast cases of hearth disease
f1-score: 836%
f1-score is the sympathetic base of preciseness and think. Logistic Regression strikes a good balance between precision and recall with an F1-score of 83.6%.

Decision Tree

Accuracy: 82.1%
Decision Trees work by splitting the data into subsets based on the feature that best separates the classes. Here, it achieved 82.1% accuracy, meaning that around 82% of the predictions were correct.
Precision: 80.2%
This model predicted heart disease correctly 80.2% of the time when it made a positive prediction.
Recall: 81.4%
Recall indicates that the Decision Tree model identified 81.4% of the actual heart disease cases.
F1-Score: 80.8%
The F1-score of 80.8% shows that while the Decision Tree is effective, it's slightly less balanced than Logistic Regression.

Random Forest

Accuracy: 88.3%
Random Forest, an ensemble model based on multiple decision trees, provided the highest accuracy at 88.3%. It aggregates the results of many decision trees, making it more robust and less prone to overfitting.
Precision: 85.4%
Precision for Random Forest is 85.4%, meaning that when it predicted heart disease, it was correct 85.4% of the time.
Recall: 86.7%
Recall for this model is 86.7%, indicating it identified 86.7% of the actual heart disease cases.
F1-Score: 86.0%
The F1-Score is the highest among all models, suggesting that Random Forest is the most balanced in terms of precision and recall.

Support Vector Machine (SVM)

Accuracy: 84.5%
SVM is a powerful model that finds the optimal hyperplane separating classes. It achieved 84.5% accuracy, correctly predicting heart disease in most cases.
Precision: 82.7%
When SVM predicted heart disease, it was correct 82.7% of the time.
Recall: 83.9%
The model identified 83.9% of all true heart disease cases.
F1-Score: 83.3%
With an F1-score of 83.3%, SVM performs well in balancing precision and recall, though it slightly lags behind Random Forest in overall performance.

CONCLUSION

This research demonstrates the potential of Machine learning in predicting heart disease using clinical Information. Random Forest proved to be the most effective model, offering high accuracy and interpretability. While Machine learning provides a promising tool for healthcare further

work is needed to Combine these methods into clinical workflows. prospective search might search advance sound acquisition techniques and corroborate Representations along big different data sets

REFERENCES

[1] **Seaman, A. (n.d.). "What You Need to Know About References During a Job Search."** LinkedIn. https://www.linkedin.com/pulse/what-you-need-know-references-during-job-search-andrew-seaman

[2] **Staffing Advisors. (n.d.). "Job Searching 101: How to Choose and Prepare Your References."** Staffing Advisors Blog. https://www.staffingadvisors.com/blog/job-searching-101-how-to-choose-and-prepare-your-references

[3] **Job-Hunt.org. (n.d.). "How to Manage Your References to Close – Not Kill – Job Opportunities."** Job-Hunt.org. https://www.job-hunt.org/effective-job-search-references

[4] **Dickinson College Career Center. (n.d.). "Alumni Job Search Methods."** Dickinson College. https://www.dickinson.edu/info/20211/career_center/524/alumni_-_job_search_methods

[5] **Northern Vermont University Career Services. (n.d.). "Job Search Strategies."** PDF handout. https://www.northernvermont.edu/sites/default/files/job-search-strategies.pdf