iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VII **Month of publication:** July 2023

DOI: https://doi.org/10.22214/ijraset.2023.54629

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VII Jul 2023- Available at www.ijraset.com

Heart Disease Prediction using Deep Learning

K. Tejaswini[1], K. V. Varun[2], K. Narsa Raju[3], K. Vinay Kumar[4]
Artificial Intelligence & Machine Learning, Malla Reddy University

Abstract: ***Heart disease is a common, potentially fatal disorder that affects a lot of people all over the world. A thorough and timely diagnosis is essential for efficient treatment and management. By analyzing large datasets, deep learning algorithms have demonstrated considerable promise in the diagnosis of cardiac disease.***
Given the alarming prevalence of heart diseases and their substantial impact on mortality rates, researchers worldwide have dedicated substantial efforts to address this issue. They have approached heart diagnosis as a classification problem, employing data mining techniques to identify meaningful patterns.
This project specifically focuses on supervised learning, employing various algorithms such as regression (including linear regression, support vector machine, and Poisson regression) and classification (including logistic regression, decision tree, random forest, and naive Bayes).
The Python library SKLEARN is a prerequisite for making predictions in this context. By utilizing SKLEARN, data can be preprocessed by segregating attributes and labels, as well as splitting the data into test and train subsets. Additionally, SKLEARN allows the importation of algorithms for further analysis and prediction purposes.
Keywords: heart disease, analyzing, large datasets, preprocessing, classification problem, linear regression, support vector machine, logistic regression, random forest, navie bayes.

I. INTRODUCTION

For early intervention and efficient therapy, accurate and rapid heart disease prediction is essential. The use of these algorithms for heart disease prediction has gained popularity as a result of improvements in machine learning methods, notably in the area of supervised learning.
The accuracy and effectiveness of diagnosis can be improved by using machine learning models to analyse large volumes of data and spot complicated patterns that may be challenging for human experts to spot.
The science of heart disease prediction could be revolutionized by this discovery, and patient outcomes could be greatly enhanced. Healthcare practitioners can gain from more accurate risk assessments, individualized treatment plans, and proactive interventions by leveraging the potential of machine learning techniques.
By identifying high-risk individuals early on, it is possible to provide timely preventive actions and lessen the burden of heart disease on patients and healthcare systems.
In this study, we evaluate the most effective machine learning methods for predicting heart disease. We will compare the efficiency of various algorithms, assess their performance using pertinent datasets, and talk about the practical application of these conclusions. We can pave the path for future developments in this crucial area and help to improve healthcare outcomes for those at risk of heart disease by comprehending the benefits and drawbacks of machine learning approaches in the prediction of heart disease.

II. PROBLEM STATEMENT

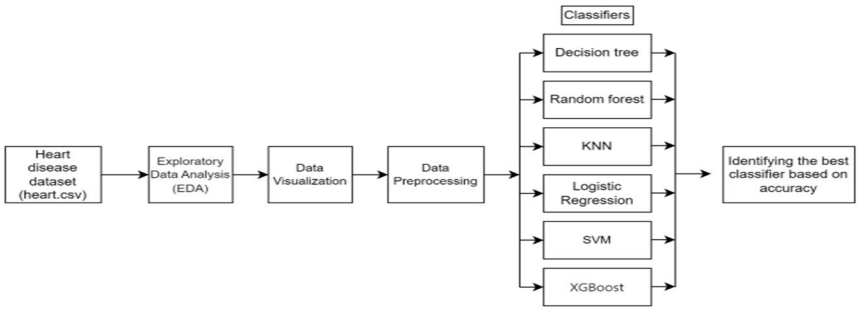
The problem statement of the Heart Disease Prediction is heart arrest is one of the leading causes of death today. Since heart disease is becoming a much bigger problem for people, our initiative focuses on estimating its likelihood using deep learning Techniques.

III. PROPOSED SYSTEM

The proposed approach was created with the intention of classifying individuals with heart disease and healthy individuals. We used the "heart.csv" illness dataset, which has been used in a number of investigations. On both all available and chosen features, the effectiveness of various deep learning predictive models for the diagnosis of heart disease was evaluated. By using hyper parameter tuning and other techniques, we have discovered the potential of a number of additional classifiers and enhanced prediction score.

©IJRASET: All Rights are Reserved

A. Architecture



IV. MATHEMATICAL MODELLING

A. Decision Tree

A decision tree is a visual representation of potential outcomes resulting from a sequence of interconnected choices. It provides individuals or organizations with a framework to compare different actions based on their associated costs, probabilities, and benefits. By following the branches of the tree, decisions can be made by evaluating the possible consequences at each step.

B. Random Forest

Random forest is a widely used machine learning algorithm that combines the predictions of multiple decision trees to generate a single, aggregated result. By creating an ensemble of decision trees and combining their outputs, random forest improves prediction accuracy and reduces over-fitting. Each decision tree in the random forest is trained on a different subset of the data and makes independent predictions, which are then combined to produce the final result.

C. K-Nearest Neighbors (KNN)

The k-nearest neighbors algorithm, often referred to as KNN or KNN, is a non-parametric and supervised learning classifier. It operates based on the principle of proximity, where a data point's classification or prediction is determined by the classes of its k nearest neighbors in the feature space. KNN is a versatile algorithm used for both classification and regression tasks, relying on the similarity between data points to make predictions.

D. Logistic Regression

Logistic regression is a widely adopted machine learning algorithm employed in the supervised learning domain. It is specifically designed for predicting categorical outcomes using a given set of independent variables or features. Unlike linear regression, which predicts continuous values, logistic regression estimates the probability of an event belonging to a particular class. By fitting a logistic function to the data, logistic regression provides a probabilistic interpretation of the relationship between the features and the target variable.

E. Support Vector Machine (SVM)

Support Vector Machine, is a powerful supervised learning algorithm used for classification and regression tasks. In SVM, data points are mapped to a high-dimensional space, where a hyperplane is generated to separate different classes or categories. The goal of SVM is to find the optimal hyperplane that maximizes the margin between the classes, effectively creating a decision boundary. SVM can handle both linearly separable and non-linearly separable data by using kernel functions to transform the data into higher dimensions.

F. XGBoost

XGBoost is a highly optimized distributed gradient boosting library designed to be efficient, flexible, and portable. It is based on the gradient boosting framework and implements machine learning algorithms. XGBoost utilizes an ensemble of weak predictive models, such as decision trees, to iteratively improve the overall prediction performance. It employs a combination of boosting and regularization techniques to handle complex relationships and deliver accurate predictions. XGBoost is widely used in various machine learning competitions and has gained popularity for its efficiency and effectiveness.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

V. SIMULATION AND RESULTS

The simulation of the proposed project would involve training and evaluating the models using the dataset of heart disease. The proceedings inherent in the emulation may be encapsulated thusly:

Model Evaluation: After training, the performance of the models can be evaluated in the Figure 1. Metrics such as accuracy, precision, recall, and F1-score could be calculated to measure the model's classification performance. This evaluation step helps in fine-tuning the models and adjusting hyper parameters if needed.

Decision Tree

Accuracy Score: 87.68%

Classification Report: Precision Score: 91.19% Recall Score: 86.70% F1 score: 88.89%

Random Forest

Accuracy Score: 93.28%

Classification Report: Precision Score: 93.66% Recall Score: 94.58% F1 score: 94.12%

KNN

Accuracy Score: 87.39%

Classification Report: Precision Score: 87.98% Recall Score: 90.15% F1 score: 89.05%

Logistic Regression

Accuracy Score: 87.39%

Classification Report: Precision Score: 88.35% Recall Score: 89.66% F1 score: 89.00%

SVM

Accuracy Score: 89.36%

Classification Report: Precision Score: 89.10% Recall Score: 92.61% F1 score: 90.82%

XG Boost

Accuracy Score: 92.16%

Classification Report: Precision Score: 92.68% Recall Score: 93.60% F1 score: 93.14%

Figure 1

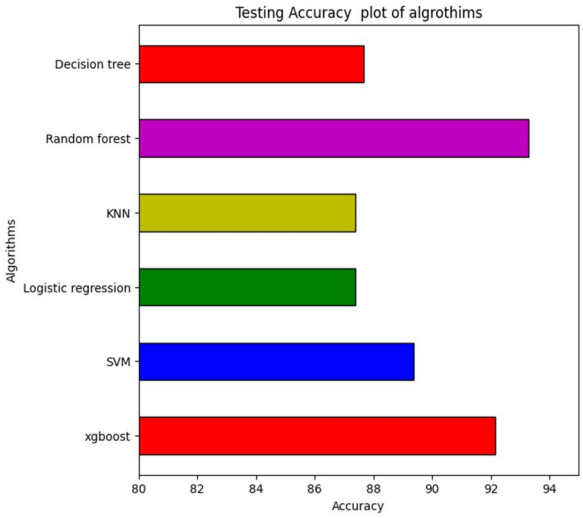


Figure 2

VI. CONCLUSION

In conclusion, the project successfully developed a heart disease prediction using deep learning. The models were trained on a dataset of heart disease, with proper preprocessing techniques applied to enhance the quality of the data. We have been performed experiments to identify the performance of various algorithms. Certain feature selection algorithms are used because features selection increased classifier performance in terms of accuracy and reduce processing time of classifier. The trained model demonstrated high accuracy and performance. The project's results showcase the potential analysis of heart prediction.

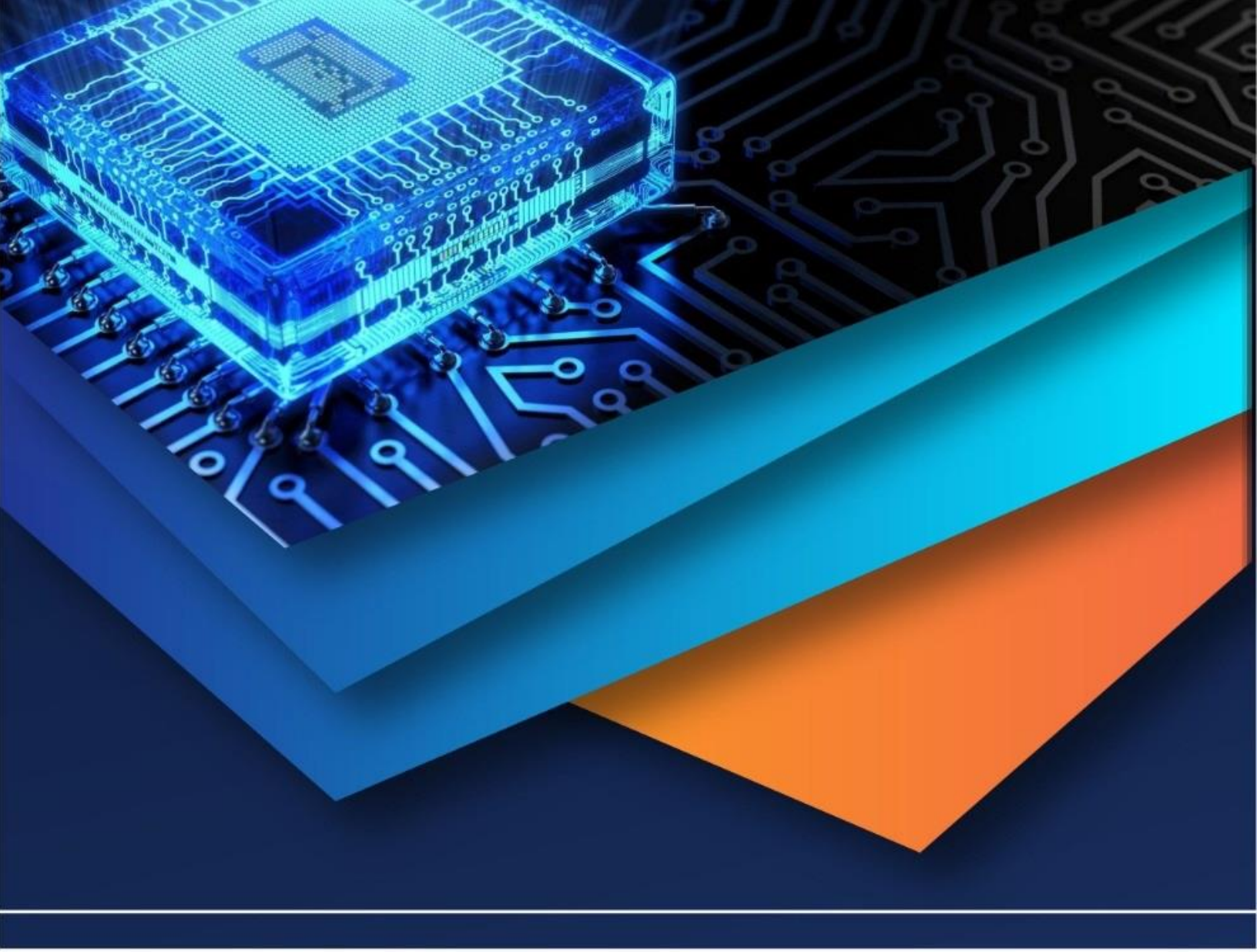
VII. ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to Malla Reddy University for the support and resources provided during the course of this research.

REFERENCES

[1] Patil, Shanta Kumar B., and Y. S. Kumaraswamy. "Extraction of significant patterns from heart disease warehouses for heart attack prediction." IJCSNS 9.2 (2009):228235.ResearchPapers(ChronicKidneyDiseaseprediction):IEEE-Xplore.
[2] Soni, Jyoti, et al. "Predictive data mining for medical diagnosis: An overview of heart disease prediction." International Journal of Computer Applications 17.8 (2011): 43-48.
[3] Chitra, R., and V. Seenivasagam. "Review of heart disease prediction system using data mining and hybrid intelligent techniques." ICTACT journal on soft computing 3.04 (2013): 605-609.
[4] M. Durairaj and N. Ramasamy, "A Comparison of the Perceptive Approaches for Preprocessing the Data Set for Predicting Fertility Success Rate," International Journal of Control theory and Applications, vol. 9, pp. 256-260, 2016.
[5] L. A. Allen, et al., "Decision Making in Advanced Heart Failure:A Scientific Statement From the American Heart Association," American.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)