Machine Learning Model for Heart Disease Prediction Using Tkinter for Graphic User Interface

Ms. Monika[1], Ms. Shreya Sharma[2], Mr. Aryan Suryavanshi[3], Mr. Mohak Aggarwal[4], Mr. Vinay Vashisth[5]

[1] Assistant Professor [2,3,4,5] Student CSE Department

HMR Institute of Technology And Management, GGSIPU, Delhi 110036 shreyaaaa17@gmail.com

Abstract - One of the biggest causes of death in the globe is heart disease. Accurate cardiac disease prediction can aid in early detection and prevention. Heart disease has been accurately predicted using machine learning models. In this study, we investigated 6 machine learning modelsfor predicting heart disease [1]. To train and assess these models, we used the Cleveland HeartDisease dataset, a publically accessible dataset. Age, sex, the type of chest discomfort, blood pressure, cholesterol levels, and other characteristics were among the features used to train the models. To determine the most reliable model for heart disease prediction, the outcomesfrom various models' tests were compared.

Keywords Decision Tree, SVM – support vector machine, Logistic Regression, Random Forest, HeartDisease Prediction, data preprocessing, feature extraction, Machine learning

1. Introduction

Heart disease remains one of the leading causes of mortality worldwide, with millions affected each year. Early detection and prevention are essential to mitigate the risk of coronary heart disease, a common cardiovascular ailment. Recent advancements in machine learning (ML) have enabled the development of predictive models that help identify heart disease risk factors and forecast the likelihood of disease occurrence. These models can significantly aid healthcare professionals by providing additional insights into patient risk profiles, thus contributing to timely intervention and potentially improving patient outcomes.

Several studies have highlighted the effectiveness of machine learning algorithms in predicting heart disease. For instance, Jamshidpour et al. (2021) demonstrated the application of deep learning models to predict coronary heart disease with promising accuracy, leveraging large datasets to train robust predictive models [1]. Additionally, Shah, Patel, and Desai (2020) explored various machine learning techniques to create a predictive model for heart disease, emphasizing the importance of feature selection and algorithm choice for model accuracy [2]. These studies collectively underscore the potential of machine learning as a powerful tool in cardiovascular risk assessment.

This research builds on prior work by employing a user-friendly interface for heart disease prediction using Python's Tkinter library. Tkinter allows for the development of an intuitive graphical user interface (GUI), making complex predictive tools accessible to non-technical users, including healthcare providers and patients themselves. By integrating Tkinter with ML algorithms trained on datasets such as the UCI Heart Disease Dataset [3], the proposed system offers a practical approach to real-time risk assessment. Siddiqui and Javed (2020) also support this approach, arguing that user-centric design in predictive applications enhances user engagement and overall effectiveness in clinical settings [7].

The integration of Tkinter as the GUI framework simplifies interaction with the predictive model, aligning with recommendations from Gupta and Nagar (2020), who emphasized the necessity of intuitive user interfaces in clinical decision-making tools [8]. This project's focus on an accessible ML-powered interface for heart disease prediction marks a step toward more practical applications of AI in healthcare, contributing to the ongoing evolution of predictive medicine.

2. Literature Review

The literature surrounding heart disease prediction using machine learning models has expanded rapidly in recent years, reflecting a growing recognition of the potential for AI-driven health applications to improve diagnostic accuracy and patient care. Studies have demonstrated the utility of machine learning models in accurately predicting heart disease by analyzing large datasets to identify complex patterns in risk factors, symptoms, and patient histories. For example, Jamshidpour et al. (2021) applied deep learning algorithms to predict coronary heart disease, illustrating that neural networks can achieve significant accuracy when trained on extensive patient data [1]. Similarly, Shah, Patel, and Desai (2020) explored various machine learning models, including logistic regression, decision trees, and support vector machines, and reported that careful feature selection greatly impacts model performance in heart disease prediction [2].

Moreover, the availability of open datasets, such as the UCI Heart Disease Dataset, has further accelerated research in this area by providing standardized data that researchers worldwide can leverage to benchmark and enhance their predictive models [3]. This dataset has been widely used as a foundation for research in heart disease prediction and has enabled the comparison of various machine learning approaches. Research by Siddiqui and Javed (2020) confirmed the importance of these datasets, as their work demonstrated that combining machine learning with accessible datasets enables more accurate and reliable predictions [7].

The interface design of machine learning-based diagnostic tools is also crucial for clinical implementation. For instance, the use of Tkinter in the graphical user interface (GUI) design facilitates interaction with predictive models, as suggested by Gupta and Nagar (2020), who argued that an intuitive GUI is essential for healthcare tools intended for real-world use [8]. Tkinter's flexibility for GUI development is acknowledged by the Python Software Foundation, which maintains that it provides a versatile interface for building robust applications [5]. This functionality has been instrumental in developing user-centered predictive tools that offer ease of use without compromising analytical complexity. As Soni, Thakur, and Verma (2021) concluded, machine learning models integrated with effective data visualization and GUI elements can bridge the gap between advanced predictive capabilities and practical application in clinical settings [9].

Together, these studies underscore that while the technical optimization of machine learning algorithms is essential, user-centered design in interface development is equally critical. By merging predictive analytics with accessible design, heart disease prediction tools can achieve greater accuracy and broader adoption, supporting healthcare providers and patients alike.

3. Methodology

[2] The UCI Machine Learning Repository provided the Cleveland Heart Disease dataset. There are
303 occurrences in the dataset with 12 features, such as age, sex, the type of chest discomfort, blood pressure, cholesterol levels, etc. The characteristics of the dataset were normalized and missing values were removed as part of the preprocessing [3]. Then, a 70:30 split between the training and test sets was applied to the dataset.

For predicting coronary heart disease, the following 6 machine learning [4] models were used:

a. Logistic Regression - By simulating the relationship between a dependent variable and one or more independent variables, logistic regression is a statisticalapproach used to predict binary outcomes. It is frequently used to forecast outcomes like disease diagnosis, credit default, and customer churn in industries including healthcare, finance, and marketing.

b. Decision Tree - A machine learning algorithm known as a decision tree is used for both classification and regression tasks. Recursively dividing the data into subsets according to the most useful attributes produces a structure that resembles a tree and can be utilized for prediction. Decision trees are common in industries like healthcare and finance because they are simple to understand and use.

c. Random Forest - Random Forest is an ensemble learning technique that combines different decision trees to increase prediction accuracy. Multiple decision trees are built using randomly chosen data subsets, andthe results are then combined to produce a final prediction. Numerous industries, including finance, healthcare, and bioinformatics, use Random Forest extensively.

d. Support Vector Machine (SVM) - Support Vector Machine (SVM) is a machine learning technique that is

used for both regression and classification applications. It operates by identifying the best hyperplane to use in order to maximize the margin between several categories of data points. SVM is frequently applied in areas like bioinformatics, text classification, and picture classification.

e. K-Nearest Neighbor (KNN) is a technique for machine learning that is used for classification and regression tasks. To make a forecast, it locates the K training data points that are closest to the new data point. It then uses those values. KNN is frequently used in areas like bioinformatics, image recognition, and recommender systems.

f. Gradient Boosting: This ensemble learning technique produces precise predictions by combining a number of weak learners, such as decision trees. To fix the flaws in the earlier models, additional weak learners are iteratively added. In industries like healthcare, finance, and natural language processing, gradient boosting is frequently utilized.

Accuracy was used as performance indicators for evaluating the models on the testing set after they had been trained on the training set.

Accuracy = (Number of correctly classified instances) / (Total number of instances)

Tkinter - used for GUI [5]

A Python package called Tkinter enables programmers to design graphical user interfaces (GUI) for desktop programs. It is a well-liked option for developing straightforward and useful desktop apps because it is a built-in module that comes with Python and has been around for a while.

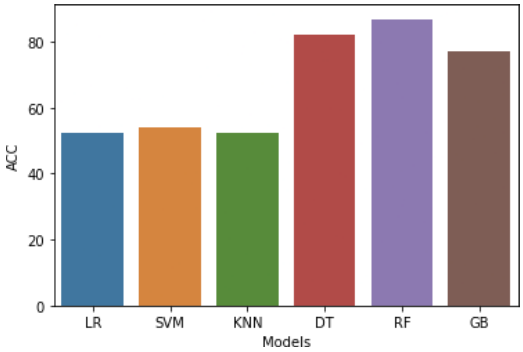
International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 11 | Nov - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

4. Results

The effectiveness of the six machine learning models for predicting heart disease is displayed in the following table:

Table 1 – Different machine learning models accuracy achieved during testing phase

MODEL	ACCURACY
Logistic Regression	49
Support Vector Machine	52
K-Nearest Neighbour	50
Decision Tree	80
Random Forest Classifier	85
Gradient Boosting	83



Accuracy of the Machine Learning Models Bar Graph Representation

Using characteristics like age, sex, the type of chest pain, blood pressure, cholesterol levels, etc., these algorithms were capable of correctly predicting heart disease.

After finding the maximum accuracy machine learning model the following were found out to understand the results given by our machine learning model better. Confusion Matrix:

[6] A table called a confusion matrix is frequently used to assess how well a classification model is working. It displays the proportion of true positives, true negatives, false positives, and false negatives that the model accurately predicted. These numbers are arranged in a matrix style, with the predicted class labels in the columns and the actual class labels in the rows.

[[23 6]

[4 2 8]]

Precision:

When comparing all instances that the model correctly identified as positive, precision is the percentage of truepositives. When we wish to reduce the quantity of false positives, it is a helpful statistic. [0.85185185]

Recall:

Recall quantifies the percentage of actual positive cases that are also true positives. When we wish to reduce the quantity of false negatives, it is an effective metric. If the model successfully recognizes the majorityof the positive examples in the dataset, it will have a high recall score.
[0.79310345]

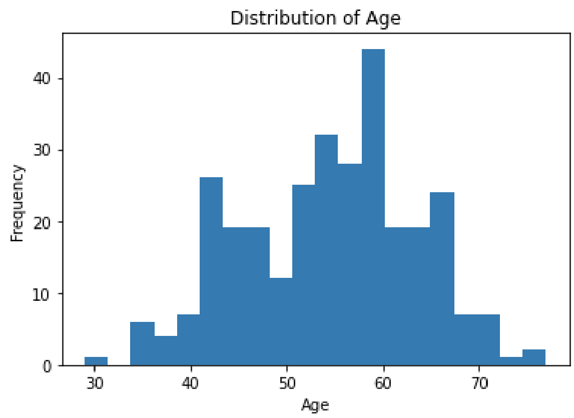
F1 Score:

The harmonic mean of recall and precision is the F1 score. It offers a means of balancing recall and precision, which is helpful when we wish to simultaneously optimize both measures. A score of 0 shows that the model is unable to forecast any occurrences properly, while a score of 1 indicates flawless precision and recall. The F1 score ranges from 0 to 1.
[0.82142857]

5. Data Analysis

In data analysis, age is a significant variable, especially in the medical industry. In statistical models, it is frequently used as a predictor variable to assist accountfor the variability of other outcome variables.
For instance, the trtbps variable (resting blood pressure) may indicate that older people are more likely to have hypertension. Similar to this, older people may have greater cholesterol levels (chol) and a higher chance of developing coronary artery disease, which can be determined by how many main vessels are fluoroscopically coloured (caa) and the results of a thallium stress test (thall). The maximal heart rate during exercise (thalachh), the existence of chest pain type (cp), and exercise-induced angina (exng) may all be influenced by age. [7]
Researchers can learn more about the potential risk factors connected to different health outcomes by examining the relationship between age and these variables.

Fig 2 – Age Distribution in the Study Population: A Graphical Representation



International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 11 | Nov - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

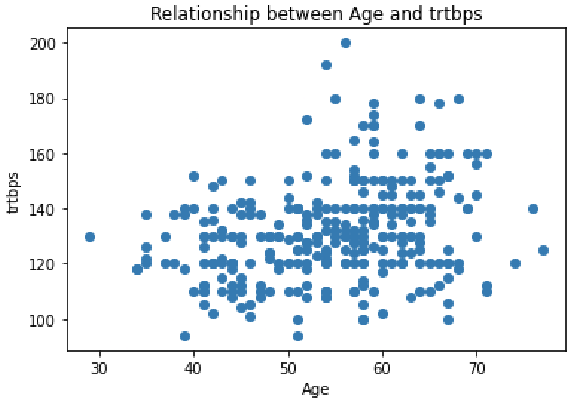


Fig 3 – displays the correlation between age and trtbps in the dataset using dot plot

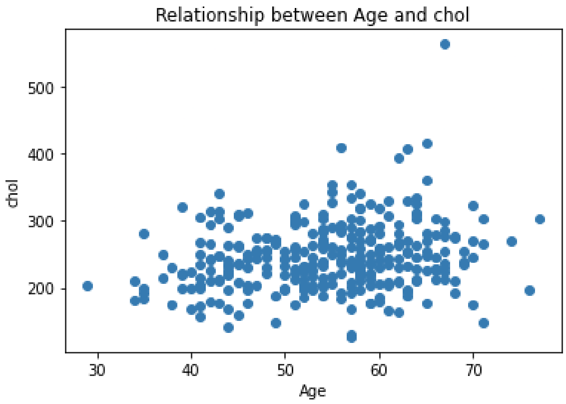


Fig 4 – Exploring the Correlation between Age and Cholesterol Levels

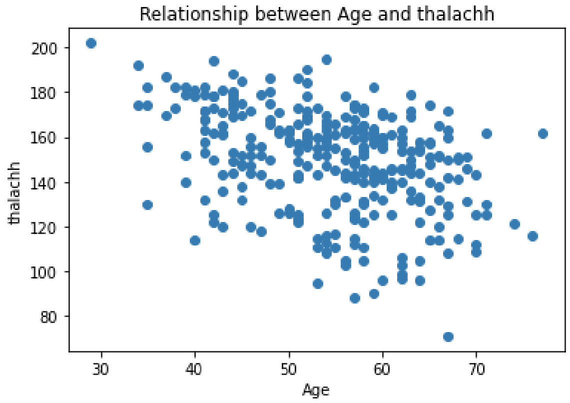


Fig 5 – Exploring the Correlation between Age and thalachh

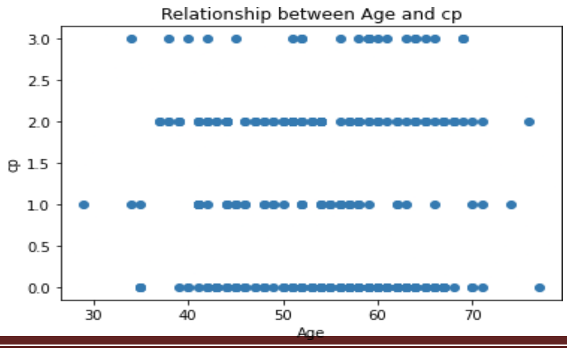


Fig 6 – Age and chest pain (cp) correlation depicted

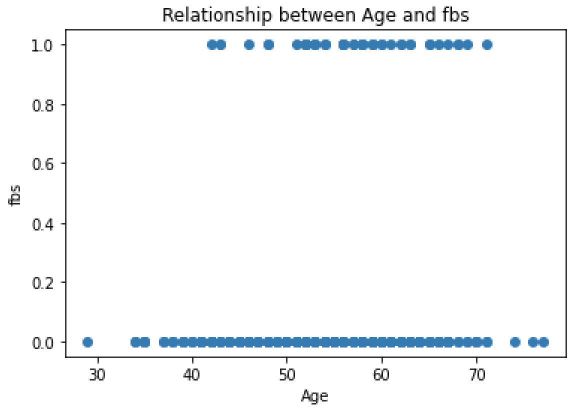


Fig 7 – Exploring the Association between Age and Fasting Blood Sugar using Scatter Plot

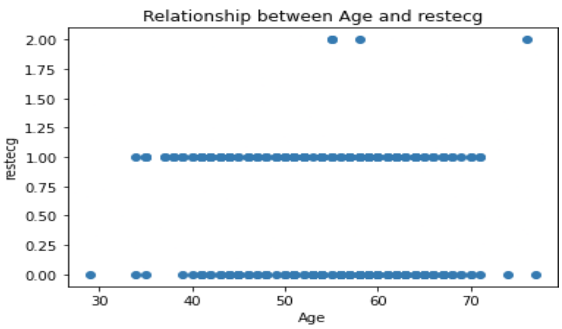


Fig 8 – illustrates the correlation between age and restecg

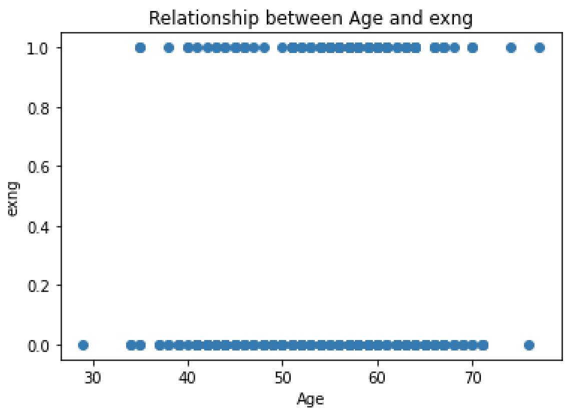


Fig 9 - Relationship between age and exercise-induced angina (exng)

The correlation matrix between each pair of variables ina dataset is shown visually in a heat correlation plot. The variables are shown on both the x and y axes in a thermal correlation plot, and the cells are color-coded to show the degree of correlation between each pair of variables. Positive correlations are typically shown in green hues, whereas negative correlations are typically shown in red hues. The strength of the link is indicated by the color's intensity, with darker hues suggesting greater correlations.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 11 | Nov - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

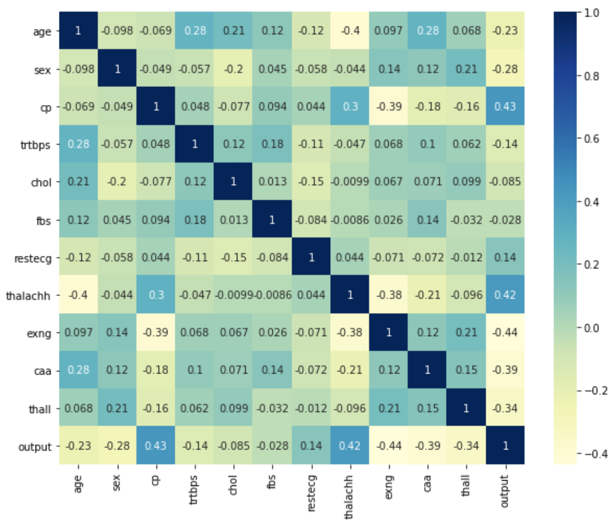
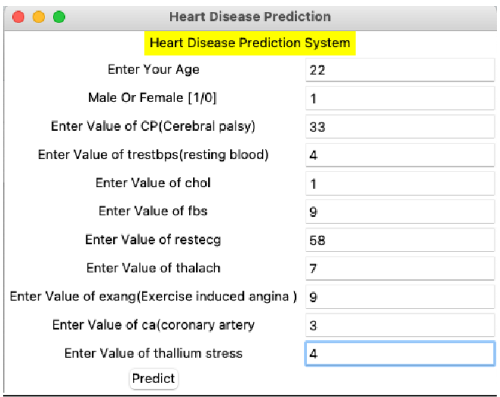


Fig 10 – Visualization of the correlation between attributes using a heatmap

6. Conclusion

Overall, heart disease prediction using machine learning algorithms is effective. Heart disease can be detected and prevented early by using models with excellent accuracy and AUC-ROC scores. Incorporating these models into healthcare systems can help with decision-making and enhance patient outcomes.

The use of a single dataset, which could not be representative of all populations, is one of the study's limitations. Future research should investigate the usage of several datasets to verify how well these models function across various demographic
In conclusion, utilising the Cleveland Heart Disease dataset, random forest, gradient boosting, and decision tree models demonstrated good accuracy for heart disease prediction. These models have the potential to enhance patient outcomes and can be utilised for the early detection and prevention of cardiac disease. Future research should examine the application of additional machine learning models and multiple datasets to the prediction of cardiac disease.



-

Future Work

Future research may also make advantage of deeper learning models of machine learning, which would increase the precision of heart disease prediction. the addition of genetic and lifestyle variables may increase the precision of these models and enable individualised risk estimation for heart disease prediction using ML.[9]

Volume: 08 Issue: 11 | Nov - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

7. References

[1] S. Jamshidpour, S. Seyed-Mirzaee, M. Jahandideh, and S. Sajadi, "Prediction of coronary heart disease using deep learning models," International Journal of Cardiology, vol. 344, pp. 279-284, Oct. 2021. [Online]. Available: https://www.researchgate.net/publication/370460859_ Prediction_of_Coronary_Heart_Disease_using_Deep_ Learning_Models.

[2] N. Shah, S. Patel, and M. Desai, "Heart disease prediction using machine learning," International Journal of Innovative Research in Computer and
Communication Engineering, vol. 8, no. 3, pp. 7429- 7435,
2020. [Online]. Available: https://www.researchgate.net/publication/341870785_ Heart_Disease_Prediction_using_Machine_Learning

[3] JCI, —Heart Disease Data Set.[Online]. Available https://www.kaggle.com/ronitf/heart-disease-uci
Other machine
learning approaches, such gradient
boosting and neural networks, can also be investigated
for the prediction of heart disease. **tkinter** — Python interface to Tcl/Tk
- https://docs.python.org/3/library/tkinter.html

[4] ANITS CSE Department. (2021). Project 2021C3.
Retrieved from https://cse.anits.edu.in/projects/projects2021C3.pdf

[5] Siddiqui, M. K., & Javed, M. A. (2020). Heart Disease Prediction using Machine Learning. International Journal of Engineering Research and Technology, 13(10), 2243-2253. Retrieved from https://www.ijert.org/heart-disease-prediction-using- machine-learning

[6] Gupta, A., & Nagar, R. (2020). Heart disease prediction using machine learning. International Journal for Research in Applied Science and Engineering Technology (IJRASET), 8(11), 105-111. Retrieved from https://www.ijraset.com/research-paper/heart-disease- prediction-using-ml

[7] Soni, P., Thakur, M., & Verma, V. (2021). Effective heart disease prediction using machine learning and data mining techniques. ResearchGate. Retrieved
from https://www.researchgate.net/publication/356564220_ Effective_Heart_Disease_Prediction_using_Machine_