International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Advancements in Heart Disease Prediction: A Machine Learning Approach for Early Detection and Risk Assessment

Aniruth Ramanathan V[1], Sriram Yerra[2], Dr. Swetha N. G.[3], Pralipth Gandikota[4]

[1]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India
Email: *aniruthram2003[at]gmail.com*

[2]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India
Email: *yerrasriram.197056[at]gmail.com*

[3]Assistant Professor Senior, Department of Analytics, School of Computer Science and Engineering Vellore Institute of Technology, Vellore, India
Email: *swetha.ng[at]vit.ac.in*

[4]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India
Email: *pralipthgandikota[at]outlook.com*

Abstract: ***This study aims to evaluate the efficacy of various machine learning models in predicting heart disease risk based on clinical data. Heart disease risk prediction is clinically essential, so the authors used machine learning (ML) on clinical data to identify and assess the impact of various features on the classification of patients with and without heart disease. The authors used cross-sectional clinical data in this study. The designed ML approach established the role of various clinical features in the prognosis of heart disease. Among the evaluated attributes, certain features were identified as strong predictors with significant values. The authors employed seven ML classifiers, including Logistic Regression, Decision Tree, Random Forest, k-Nearest Neighbors, Naive Bayes, Support Vector Machine (SVM), and Neural Networks. The performance of each model was assessed based on accuracy metrics, with SVM exhibiting the highest accuracy at 91.51%. The Support Vector Machine (SVM) model exhibited the highest accuracy at 91.51%, demonstrating superior predictive capability among the evaluated models. The findings highlight the potential of advanced computational methodologies in cardiovascular risk assessment and management. This study underscores the potential of machine learning models in enhancing cardiovascular risk assessment and management. The high accuracy of the SVM model demonstrates its value in clinical settings, paving the way for future advancements in personalized medicine and proactive healthcare interventions.***

Keywords: Machine Learning, Heart Disease Prediction, Support Vector Machine, Clinical Data, Cardiovascular Risk Assessment, Predictive Modeling, Personalized Medicine

1. Introduction

Heart disease continues to be a major global cause of death, making it necessary to create precise and effective predictive models to identify people who are at-risk and carry out interventions in a timely manner. The objective of this research work is to provide a prediction framework that may be used for early identification and risk assessment, hence improving proactive healthcare management tactics by utilizing advances in machine learning and data analytics. We are motivated by the pressing need for technological innovations in healthcare to address the increasing burden of heart diseases globally. In the foreseeable future, the application of this predictive model in clinical practice promises to revolutionize cardiovascular healthcare delivery. By harnessing the power of machine learning, healthcare providers will be empowered to proactively identify individuals at heightened risk of heart diseases, enabling early interventions and personalized treatment strategies. This proactive approach not only holds the potential to significantly reduce the incidence and severity of cardiovascular disorders but also to enhance patient outcomes and overall quality of life.

The rest of the paper is organized as follows:

Section I describes the basic introduction to heart disease prediction using machine learning techniques and the motivation behind this research. Section II illustrates the various research works carried out in disease and disorder prediction using various machine learning techniques and Section III explains in detail about the various inferences that can be derived from the literature survey. Section IV provides information on the proposed methodology and section V deals with the results and discussion and sections VI and VII provide the conclusion and references respectively.

2. Literature Review

In [1], R. Indra Kumari, 2T. Poongodi and 3Soumya Ranjan Jena (2020) predicted heart disease using exploratory data analysis by utilizing the algorithm of K- Means clustering. Its advantages, including being unsupervised and computationally efficient, were highlighted. However, challenges were noted, such as the need to select the optimal number of clusters and the deterministic nature of results, which may converge to local rather than global optima.

In [2], Reldean Williams, 2Ali N Hasan and 3ThokozaniShongwe, 4Vikash Rameshar (2021) concluded

the paper titled "Heart Disease Prediction using Machine Learning Techniques," by employing Random Forest Regression with a notable accuracy of 95.08%. High prediction accuracy is achieved through averaging multiple decision trees, and the ability to handle missing data without imputation is utilized as an advantage of the method.

In [3], Vijeta Sharma, 2Shrinkhala Yadav and 3Manjari Gupta (2020) conducted a comprehensive study on predicting heart disease using Random Forest Regression in the paper, achieving an impressive accuracy of 99.7%. The advantages of this method, including its robustness to overfitting and its ability to handle non-linear relationships within complex datasets, are highlighted.

In [4], Harshit Jindal, 2Sarthak Agrawal, 3Rishabh Khera,4Rachna Jain and 5Preeti Nagrath (2021) predicted heart disease using machine learning algorithms with specific focus on K Nearest Neighbors and Logistic Regression, both achieving an accuracy of 88.5%. While simplicity and interpretability are offered by Logistic Regression, it may encounter difficulties with non-linear decision boundaries and capturing complex feature interactions.

In [5], Abhijeet Jagtap, 2Priya Malewadkar, 3Omkar Baswat and 4Harshali Rambade (2019) researched the paper and investigated Heart disease prediction focusing on the Support Vector Machine (SVM) method, with an accuracy of 64.4% achieved. Advantages are offered by SVMs in handling high-dimensional data and versatility in kernel selection, enabling complex relationship modeling.

In [6], Keshav Srivastav and 2Dilip Kumar Choubey (2020) employed K Nearest Neighbours (KNN) for heart disease prediction achieving an accuracy of 87%. KNN's simplicity and lack of a training phase are lauded, making it beginner-friendly and computationally efficient during prediction.

In [7], Md. Mahbubur Rahman, 2Morshedur Rahman Rana and 3Md. Nur-A-Alam (2022) proposed a web-based heart disease prediction system by employing a Decision Tree algorithm with a high accuracy rate of99%. Its advantages are found in its interpretability, which facilitates communication with stakeholders and healthcare professionals, and its ability to handle nonlinear relationships within complex datasets.

In [8], Raparthi Yaswanth and 2Y. Md. Riyazuddin (2021) employed Artificial Neural Networks (ANN) for heart disease prediction in the paper, achieving an accuracy of92.30%. The advantages of ANN include its ability to capture complex non-linear relationships and automatically learn relevant features from raw data.

In [9], L. Chandrika and 2K. Madhavi (2021) proposed hybrid framework, HRFLM, combines Random Forest and linear models to enhance heart disease prediction accuracy (88.4%). By leveraging the strengths of both models, improved performance and robustness against overfitting are offered.

In [10], Nurul Absar, 2Emom Kumar Das and 3M. R. I. Faruque (2022) concluded the efficiency of a machine-learning- supported smart system for heart disease prediction is explored in the study conducted with the K Nearest Neighbors (KNN) algorithm primarily employed, achieving an impressive accuracy of 97.826%. The advantages of KNN, including its simplicity and lack of a training period, are highlighted, making it accessible for beginners.

3. Literature Inferences from Literature Survey

1) Most Used Model: Random Forest Regression and K Nearest Neighbors are the most frequently used models, each appearing multiple times in the table.
2) Major Concerns: Complexity, overfitting, and computational resources are major concerns across various models.
3) Hybrid Approach: A hybrid approach combining random forest with a linear model is proposed to improve prediction accuracy while addressing overfitting.
4) Considerations: Interpretability, simplicity, and the trade- off between accuracy and model complexity are important considerations in selecting the appropriate model for heart disease prediction.

4. Proposed Methodology

This research provides the clarity and ability to predict the probabilistically classified as heart disease. In this study, we have used various machine learning algorithms to perform the predictions and can be applied to many real-time applications specifically in the medical field. The Methodology describes the path we followed in the study. See Fig. 1 for pathway flowchart.

a) Data Collection:

Data collection is the initial phase of approach and the important phase as well. Data can be collected from surveys, public records, from government publications, from magazines and journals, from online platforms like Kaggle, UCI-machine learning repository. This study utilizes the dataset from Kaggle repository, which is the biggest hub of data science resources. There are over 50,000 datasets that cover many domains, without scarifying the quality of data.

b) Data Pre-processing:

The Dataset taken for the study isn't clean, it contains some missing data. Pre-processing is done to prepare the data for prediction modelling. Handling missing values is analogue to an art and it can be done in many ways, by removing the missing value records (suggestable when dataset has a greater number of records) replacing the missing values with mean of entire feature values replacing the missing values with mode of entire feature values. In this study, we have made a hybrid approach of using mode imputation for categorical values and mean imputation for numerical values which resulted better. Then it is followed by shifting columns for feature scaling, encoding with categorical variables.

Table 1: Literature Survey

S.no	Paper Title	Method	Advantages	Disadvantages
1.	Heart Disease Prediction using Exploratory Data Analysis	K- Means Clustering	• Unsupervised Learning Simple and Fast	• Deterministic Result • Number of Clusters Selection in prior
2.	Heart Disease Prediction using Machine Learning Techniques	Random Forest Regression	• High Accuracy • Handles Missing Data	• Computational Complexity • Memory Consumption
3.	Heart Disease Prediction using Machine Learning Techniques	Random Forest Regression	• Robustness to Overfitting • Handles Non-linear Relationships	• Less Interpretable • Bias- Variance Tradeoff
4.	Heart disease prediction using machine learning algorithms.	K Nearest Neighbors and *Logistic Regression	• Simple and Interpretable • Probabilistic Output	• Linear Decision Boundary • Limited Expressiveness
5.	Heart Disease Prediction using Machine Learning	Support Vector Machine	• Effective in high- dimensional spaces • Versatility in kernel selection	• Sensitive to choose Of kernel • Computational complexity
6.	Heart Disease Prediction using Machine Learning and Data Mining	K Nearest Neighbors	• Simple to understand and implement. • No training phase	• Computationally expensive • during prediction • Sensitivity to irrelevant features
7.	A web-based heart disease prediction system using machine learning algorithms.	Decision Tree	• Interpretability • Handle Nonlinear Relationships	• Overfitting • Instability
8.	Heart Prediction using machine Learning techniques	Artificial Neural Networks	• 1.Non-linearity Feature Learning	• Black BoxNature • Training Complexity
9.	A Hybrid Framework for Heart Disease Prediction using Machine Learning Algorithms	hybrid random forest with a linear model	• Improved Prediction • Accuracy, Robustness to Overfitting	• Complexity Interpretability Trade-off
10.	The Efficacy of Machine- Learning Supported Smart System	K Nearest Neighbors	• Simplicity • No Training Period	• Computational Complexity • Sensitivity to Noise and Outliers

c) Feature Engineering:

Feature Engineering is the process of generating new features/ transforming or removing existing features to improve the model score and performance. This is an additional unique phase of the approach which gives improved accuracy over previous versions. In this model, we have used mutual information scores to perform removal of a set of features as part of feature engineering. This involves more mathematics, as different combinations of sets to be tried in each run and find a combination that is mathematically stable. In this study, there was a gain of one percent from the previous version using feature engineering.

d) Model Training (SVM):

The accuracy of a model is predicted under this phase, which will list the base of accuracy score, and places for improvements. We initiate this phase by splitting the dataset into training and test set. Processing data to ease the model training must be done only on training set, the test set must be kept it unmanipulated. Feature Scaling is implemented as one such process to ease the model training, which scales the feature values to one standard scale. It uses Standardization or Normalization as part of feature scaling. Now, each model is trained and tested one by one. For this Study, Decision Tree, Logistic Regression, Support Vector Machine, Naïve Bayes, K-Nearest Neighbors, Random Forest Models are executed for comparison and SVM Model is chosen as it provides better accuracy than others. Feature Engineering is fully dependent on accuracy score given under this phase.

e) Principal Component Analysis:

Principal Component Analysis, commonly known as PCA, is a statistical technique for dimensionality reduction. It transforms the data with new co- ordinates such that dimensions are compressed to a lower ordinality for visualization purposes. It's also proven that it improves model score sometimes. purposes and diagnostical purposes.

f) Metrics and Evaluation Score:

This Phase is the score prediction phase, the output predicted under each model is displayed and feature engineering and other processes are done according to the predicted accuracy to improve the score. For this Study, the metrics used for evaluation are confusion matrix, accuracy score and standard deviation. See Fig. 3. Confusion matrix is metric under machine learning model score performance which provides a comprehensive view of its predictions from the calculations of true positives, true negatives, false positives, false negatives. Most of the time it is a 2x2 table, sometimes it extends as per feature requirements of the dataset. The formula for calculating accuracy and precision is depicted in Fig. 4. Cross Validation Score metric is also used under this study to predict the overall accuracy and standard deviation of the model.

g) Models Used for Comparison:

As mentioned in the Model training phase, for this study, Decision tree, Random Forest, Support Vector Machine, Logistic Regression, Naïve Bayes, K-Nearest Neighbors are used. Each Model has its own way of proceedings, advantages, and disadvantages. Out of the models used in this study, after trying out each combination of set under feature engineering, SVM is suggested either under linear or radial basis function kernel.

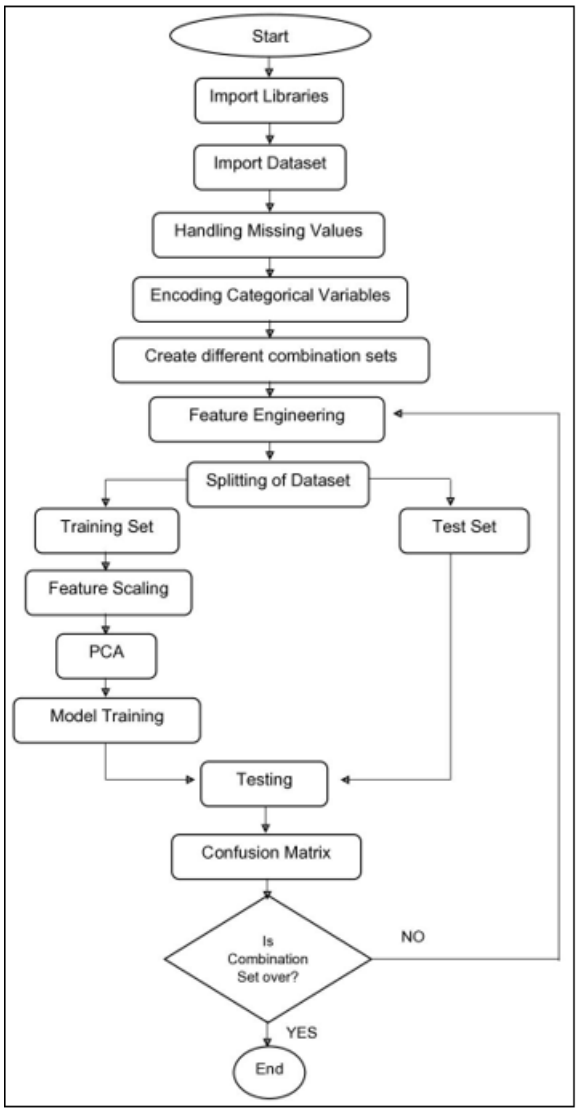

Figure 1: Flow Diagram

- *Logistic Regression*: Logistic regression comes under the category of supervised machine learning model under classification. Classification model predicts an output of either one or zero with respect to the probability of the output. Logistic Regression uses the S-shaped curve or sigmoid function to perform the prediction.
- *Naïve Bayes:* Naïve Bayes comes under the category of supervised machine learning algorithm under classification. Naïve Bayes is basically a model formed out of the Baye's probability theorem. The Bayes theorem is the formula relating the conditional probability.
- *K-Nearest Neighbors*: This model comes under supervised machine learning model under classification. Choosing the number k of neighbors form the initial step of this model. TakeK nearest neighbors of the new point, according to Euclidean distance.
- *Support Vector Machine (SVM):* Support Vector Machine forms two support vectors where the hyperplane lies in the middle providing a maximum margin between the planes and vectors thus separating the points of distinct classes. It has special parameters that distinguish it from other models, which include hyperparameter, kernel type, Degree, Kernel Coefficient (gamma). It can also be used to map in higher dimensions, but it can be computed intensive. The intuition of SVM Model is depicted in Fig 2.
- *Decision Tree:* Decision Tree model comes under supervised machine learning model, which is a non-parametric model. It consists of a root node, which splits again and again to purely classify the different and distinct categories, that subdivides the tree into decision nodes and leaf nodes, that is, the terminal nodes.
- *Random Forest:* Random Forest Model comes under the supervised machine learning model, which is an extended version of decision tree. Decision tree makes the prediction using a single tree, whereas the random forest makes the prediction from many groups of trees, which form a forest in a random manner.

Graphical Representation of Comparison of all mentioned models to analyze which model suits best for the study in terms of accuracy and background and best fits for the features of the dataset.
See Fig 5.

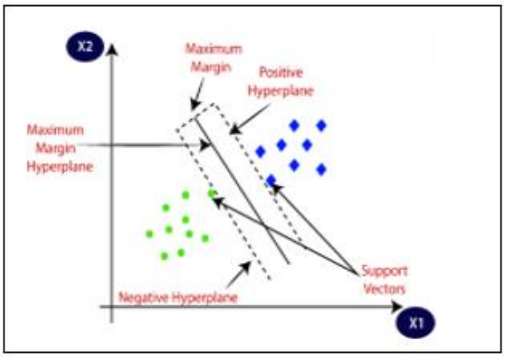

Figure 2: SVM Intuition

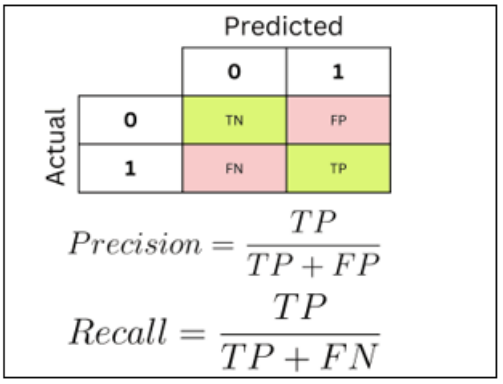

Figure 3: Confusion matrix

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

Figure 4: Accuracy Score and Precision

5. Results and Discussion:

Accuracy is metric to evaluate the predicted output in classification model. Using this metric, the best model that suits the dataset can be obtained. Basically, it is the score of how often the model is correct. It is just a mathematical

formula of correct predictions out of total predictions which range a value between 0 and 1, multiplied with 100 to project the output in terms of percentage. Precision is the metric which tells how often the model is correct in predicting the target class mostly used with categorical output whereas accuracy with numerical output. There's always a trade-off between accuracy and prediction, prioritizing them depends on the problem statement.

a) Environmental Setup:

Environmental setup for this study is depicted as PC configuration in which all the models and approach we rerun successfully.

Device name: Aniruth
Processor: Intel(R) Core (TM) i5-10300H CPU @ 2.50GHz
Installed RAM: 8.00 GB (7.84 GB usable)
Systemtype:64-bit operating system, x64-basedprocessor
Pen and touch: No pen or touch input is available for this display Edition: Windows 11 Home Single Language
Experience: Windows Feature Experience Pack1000.22684.1000.0

Under this setup, the whole study was conducted and was successfully provided with an errorless output. The result for each model is depicted in the following figures and hence it is represented in a comparison chart. See Fig. 5.

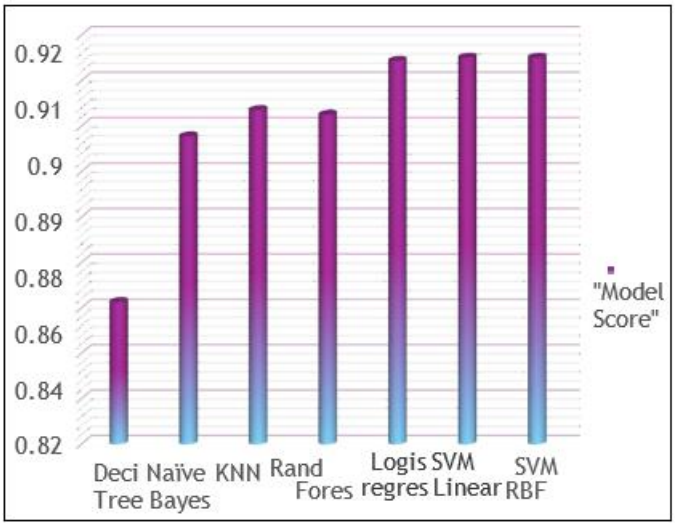

Figure 5: Graphical Representation of model

b) About Data:

The Data for this study consists of 17 independent features and 1 predicting feature which is heart disease. The values under heart disease feature are either 'yes' or 'no' which bring the prediction under the category of classification. The 17 independent features are symptoms and driving factors related to heart disease which includes body mass index, age, sex, Asthma, Kidney Disease, Skin cancer and many more.

c) Model Scores:

- Decision Tree Model gives an accuracy of 86.1% from the approach used in the study.
- Naïve Bayes Model gives an accuracy of 89.7% from the approach used in the study.
- KNN Model gives an accuracy of 90.3% from the approach used in the study.
- Random Forest Model gives an accuracy of 90.2% from the approach used in the study.
- The Logistic Regression Model gives an accuracy of 91.44% from the approach used in the study.
- The SVM Linear Model gives an accuracy of 91.51% from the approach used in the study.
- The SVM Kernel 'rbf' Model gives an accuracy of 91.51% from the approach used in the study.

The accuracy scores and confusion matrix of the models are given in Fig. 6.

6. Conclusion

Our study evaluated seven machine learning models to predict heart diseases through experiments and rigorous analysis we conducted which demonstrated the effectiveness of machine learning methods in accurately predicting heart disease and provide valuable insights into the potential of predictive analytics in clinical practice. Notably, the Support Vector Machine (SVM) model demonstrated exceptional accuracy, achieving rates of up to 91.51%. These results underscore the significant potential of advanced computational methods in healthcare, promising more informed decision-making and improved patient outcomes through predictive analytics. Based on the positive results of our predictive analysis, we strongly advocate for integrating these predictive models into routine clinical practice to facilitate early detection and self-management of cardiovascular diseases. Application of real- life screening tools could allow healthcare providers to identify high-risk individuals, interrupt interventions, and ultimately reduce health risks. Embracing such innovative approaches paves the way for a proactive healthcare paradigm, where prevention and personalized care become paramount in safeguarding public health.

Future studies can explore the integration of additional machine learning models and ensemble techniques to further enhance predictive accuracy. Investigating deep learning models and incorporating more comprehensive datasets could provide deeper insights and improve the robustness of predictive analytics in clinical practice.

Making the Confusion Matrix of KNN Model

```
from sklearn.metrics import confusion_matrix, accuracy_score
cm = confusion_matrix(Y_test,Y_pred)
print(cm)
accuracy_score(Y_test, Y_pred)
```

```
[[95011  1944]
 [ 8269   724]]
0.9036036546230226
```

Making the Confusion Matrix of Random Forest Model

```
from sklearn.metrics import confusion_matrix, accuracy_score
cm = confusion_matrix(Y_test,Y_pred)
print(cm)
accuracy_score(Y_test, Y_pred)
```

```
[[94966  1989]
 [ 8333   660]]
0.9025748480386605
```

Making the Confusion Matrix of Logistic Regression Model

```
from sklearn.metrics import confusion_matrix, accuracy_score
cm = confusion_matrix(Y_test,Y_pred)
print(cm)
accuracy_score(Y_test, Y_pred)
```

```
[[96817   138]
 [ 8929    64]]
0.9144202816476007
```

Making the Confusion Matrix of Naive Bayes Model

```
from sklearn.metrics import confusion_matrix, accuracy_score
cm = confusion_matrix(Y_test,Y_pred)
print(cm)
accuracy_score(Y_test, Y_pred)
```

```
[[94028  2927]
 [ 7899  1094]]
0.8978177974100502
```

Making the Confusion Matrix of SVM Linear Model

```
[38] from sklearn.metrics import confusion_matrix, accuracy_score
cm = confusion_matrix(Y_test,Y_pred)
print(cm)
accuracy_score(Y_test, Y_pred)
```

```
[[87575     0]
 [ 8239     0]]
0.91401047863569
```

Making the Confusion Matrix of SVM Kernel Model

```
from sklearn.metrics import confusion_matrix, accuracy_score
cm = confusion_matrix(Y_test,Y_pred)
print(cm)
accuracy_score(Y_test, Y_pred)
```

```
[[96955     0]
 [ 8993     0]]
0.915118737493865
```

Making the Confusion Matrix of Decision Tree Model

```
from sklearn.metrics import confusion_matrix, accuracy_score
cm = confusion_matrix(Y_test,Y_pred)
print(cm)
accuracy_score(Y_test, Y_pred)
```

```
[[89622  7333]
 [ 7355  1638]]
0.8613659531090724
```

Figure 6: Confusion Matrix and Accuracy Score of all comparison model

References

[1] R. Indra Kumari, T. Poongodi, and S. R. Jena, "Heart Disease Prediction Using Exploratory Data Analysis," *Procedia Computer Science*, vol. 173, pp. 130-139, 2020. DOI: https://doi.org/10.1016/j.procs.2020.06.017.

[2] R. Williams, T. Shongwe, A. N. Hasan, and V. Rameshar, "Heart Disease Prediction Using Machine Learning Techniques," in *2021 International Conference on Data Analytics for Business and Industry (ICDABI)*, IEEE, pp. 118- 123, 2021. DOI: 10.1109/SKIMA59232.2023.10387299.

[3] V. Sharma, S. Yadav, and M. Gupta, "Heart Disease Prediction Using Machine Learning Techniques," in *2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, IEEE, pp. 177-181, 2020. DOI: https://doi.org/10.1007/s42979-020- 00365-y.

[4] H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, "Heart Disease Prediction Using Machine Learning Algorithms," *IOP Conference Series: Materials Science and Engineering*, vol. 1022, no. 1, p. 012072, 2021. DOI: 10.1088/1757-899X/1022/1/012072.

[5] Jagtap, P. Malewadkar, O. Baswat, and H. Rambade, "Heart Disease Prediction Using Machine Learning," *International Journal of Research in Engineering, Science and Management*, vol. 2, no. 2, pp. 352-355, 2019.

[6] K. Srivastava and D. K. Choubey, "Heart Disease Prediction Using Machine Learning and Data Mining," *International Journal of Recent Technology and Engineering*, vol. 9, no. 1, pp. 212-219, 2020. DOI: http://dx.doi.org/10.35940/ijrte.F9199.059120.

[7] M. M. Rahman, "A Web-Based Heart Disease Prediction System Using Machine Learning Algorithms," *Network Biology*, vol. 12, no. 2, p. 64, 2022. DOI: https://doaj.org/article/5b5476fb239b4e5eb4a1b560d34fbb5c.

[8] R. Yaswanth and Y. M. Riyazuddin, "Heart Disease Prediction Using Machine Learning Techniques," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 5, pp. 1456-1460, 2020. DOI: http://dx.doi.org/10.35940/ijitee.E2862.039520.

[9] L. Chandrika and K. Madhavi, "A Hybrid Framework for Heart Disease Prediction Using Machine Learning Algorithms," *E3S Web of Conferences*, vol. 309, p. 01043, 2021. DOI: https://doi.org/10.1051/e3sconf/202130901043.

[10] N. Absar, E. K. Das, S. N. Shoma, M. U. Khandaker, M. H. Miraz, M. R. I. Faruque, and R. K. Pathan, "The Efficacy of Machine-Learning-Supported Smart System for Heart Disease Prediction," *Healthcare*, vol. 10, no. 6, p. 1137, 2022. DOI: https://doi.org/10.3390/healthcare10061137.

Author Profile

Aniruth Ramanathan V is currently pursuing a Bachelor of Technology (B.Tech) degree, expected to graduate in 2025. He began his undergraduate studies in 2021, focusing on Data Science and Machine Learning (ML). In addition to his academic pursuits, Aniruth worked on various hands-on data science and machine learning projects, where he applied his knowledge to real-world data and developed predictive models.

Sriram Yerra is currently pursuing a Bachelor of Technology (B.Tech) degree, expected to graduate in 2025. He began his undergraduate studies in 2021, focusing on Artificial Intelligence (AI) and Machine Learning (ML). In addition to his academic pursuits, Sriram completed an internship at Sansid from November 2023 to January 2024, where he gained valuable hands-on experience in AI and automation projects. His research interests lie in the application of AI and ML to solve real-world challenges, and he is eager to continue expanding his expertise in these cutting-edge fields.

Dr. Swetha N.G. is an Assistant Professor Senior at the Department of Analytics, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India. With a Ph.D. in Computer Science and Engineering from Anna University, Chennai, Dr. Swetha brings a strong academic foundation to her role. She has obtained her M.E (CSE) from PSG College of Technology and B.E (CSE) from Sri Krishna College of Engineering and Technology. Her career is marked by a dedicated commitment to teaching and research, with over seven years of experience mentoring students. Being a prolific

researcher, she has authored numerous papers and a book chapter in esteemed journals and conferences. Her research interests span a diverse range of topics including Web Service Composition, Parallel Computing, Multi Criteria Decision Making, Cryptography, and Information Security.

Pralipth Gandikota is currently pursuing a Bachelor of Technology (B.Tech) degree with an expected graduation in 2025. He began his undergraduate studies in 2021, focusing on Data Science, Web Development, Cloud Computing, and DevOps. Throughout his academic journey, he has undertaken multiple hands-on projects across these fields, where he applied his knowledge to develop real-world applications, predictive models, and cloud-based solutions.