Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 04, April-2025

Heart Disease Prediction using Machine Learning

Abhijitha Kandukuri
B.Tech. student,
Dept. of CSD
Institute of Aeronautical
Engineering Hyderabad, India

Harshitha Thumkunta
B.Tech. student,
Dept. of CSD
Institute of Aeronautical
Engineering Hyderabad, India

Kummari Gnanadeep
B.Tech. student,
Dept. of CSD
Institute of Aeronautical
Engineering Hyderabad, India

Abstract— **Heart disease remains a significant global health concern, emphasizing the need for early and accurate prediction to improve patient care and reduce risks. This study presents a machine learning-driven approach that integrates multiple predictive models to assess the likelihood of heart disease. The ensemble method combines Gradient Boosting, Random Forest, AdaBoost, and Support Vector Classifier to enhance prediction reliability and accuracy. Training was conducted on a dataset with features such as age, cholesterol levels, blood pressure, and exercise-induced angina. Among the models, Gradient Boosting achieved the highest accuracy at 74 percent, demonstrating its efficacy in heart disease prediction. To facilitate accessibility, a web-based application was developed using Flask for backend operations and HTML, CSS, and JavaScript for the frontend. This interface enables users to input relevant health data and receive instant predictions, providing a practical and user-friendly solution. The project highlights the potential of machine learning in healthcare by offering a scalable, non-invasive tool for early detection of heart disease, potentially aiding healthcare professionals and individuals in managing cardiovascular risks more effectively.**

KeyWords : **Heart Disease, Machine Learning, Ensemble Approach, Gradient Boosting, Predictive Analytics, Flask Application, Healthcare Technology**

I. INTRODUCTION

Heart disease is one of the leading causes of mortality worldwide, necessitating the development of reliable and efficient tools for early detection and intervention. The increasing prevalence of cardiovascular diseases has prompted the adoption of advanced technologies to address the challenge of timely diagnosis. Traditional methods, while effective to an extent, often require invasive procedures, specialized equipment, or expert analysis, which may not always be accessible to everyone. Machine learning, a rapidly evolving field, offers an opportunity to enhance diagnostic accuracy through data-driven predictions and automated systems. By analyzing complex patterns in medical data, machine learning models can provide valuable insights, aiding in early detection and better management of heart disease. This project focuses on building a machine learning-based ensemble system to predict the likelihood of heart disease. The system integrates multiple algorithms, including Gradient Boosting, Random Forest, Support Vector Classifier, and AdaBoost, to ensure robust and accurate predictions. Ensemble methods are widely recognized for their ability to combine the strengths of individual models, thereby reducing errors and improving overall performance. Using features such as age, blood ressure, cholesterol levels, and exercise-induced angina, the system is trained to recognize patterns associated with heart disease. Among the models evaluated, Gradient Boosting showed superior accuracy, making it the central component of our predictive framework. Additionally, feature engineering techniques were employed to preprocess the data, ensuring high-quality input for model training. The project also included rigorous testing and validation to achieve reliable results across various data subsets.

To enhance accessibility and usability, the predictive model is deployed as a web application. The backend is powered by Python and Flask, which manage the logic and operations of the application. The frontend, built with HTML, CSS, and JavaScript, ensures a seamless and user-friendly experience for end users. The web interface allows individuals to input their health-related data and obtain instant results on their likelihood of having heart disease. This application aims to bridge the gap between advanced medical diagnostics and everyday users, empowering individuals to make informed decisions about their health. One of the key contributions of this project lies in its integration of technology and healthcare. By leveraging machine learning and web technologies, the system offers a scalable, non-invasive solution for heart disease prediction. This is particularly relevant in regions where access to medical facilities and specialists is limited. Additionally, the project emphasizes the importance of preventive healthcare, encouraging individuals to monitor their health and seek timely interventions. The deployment of a user-friendly web application ensures that the system can be used by healthcare providers as well as individuals, broadening its impact.

In conclusion, this project underscores the potential of machine learning in transforming healthcare by providing accurate, accessible, and efficient diagnostic tools. By combining advanced predictive models with an intuitive web interface, it demonstrates a practical application of artificial intelligence in addressing real-world health challenges. The system not only supports early detection of heart disease but also contributes to reducing the burden on healthcare systems by enabling proactive health monitoring. This work serves as a foundation for further advancements in machine learning-based healthcare solutions, paving the way for innovative approaches to tackling critical medical issues.

1.1 Research Background

Heart disease continues to be a pressing global health concern, contributing significantly to morbidity and mortality rates worldwide. The growing prevalence of cardiovascular diseases has motivated researchers and healthcare professionals to seek innovative solutions for early diagnosis and prevention. Conventional diagnostic methods, while effective, often involve invasive procedures, high costs, and reliance on specialized medical personnel. These limitations create a barrier to timely detection, particularly in underserved regions. The advent of machine learning and data-driven technologies has opened new avenues for tackling these challenges, providing scalable, cost-effective, and accurate solutions for disease prediction and management.

Machine learning, as a subset of artificial intelligence, has emerged as a transformative tool in healthcare, enabling the analysis of complex datasets to uncover patterns and relationships that may not be apparent through traditional methods. In the context of heart disease prediction, machine learning models can analyze various clinical parameters such as age, cholesterol levels, blood pressure, and electrocardiogram (ECG) readings to predict the likelihood of a cardiac event. By leveraging large datasets and advanced algorithms, these models can learn from historical data, identifying subtle trends and correlations that aid in accurate prediction. Ensemble learning, a method that combines multiple machine learning models, has shown remarkable success in improving prediction accuracy and robustness. The integration of algorithms like Gradient Boosting, Random Forest, Support Vector Classifier (SVC), and AdaBoost enhances the model's ability to generalize and reduce errors. Ensemble techniques capitalize on the strengths of individual models, addressing their weaknesses through collective decision-making. This approach is particularly beneficial in medical applications, where precision and reliability are paramount. In this project, Gradient Boosting was identified as the most effective model, delivering superior performance compared to other algorithms. The availability of large, high-quality medical datasets has further propelled research in this domain. Datasets such as the UCI Heart Disease dataset provide a wealth of information for training and testing predictive models. These datasets often include diverse features and patient records, ensuring that the models are robust across various demographics and medical conditions. Feature engineering and pre- processing steps, such as normalization and handling missing values, play a crucial role in enhancing the predictive power of machine learning models. These preparatory steps ensure that the data fed into the models is accurate, consistent, and free from noise.

In addition to the development of predictive models, the deployment of user-centric applications has become a focal point in healthcare innovation. By integrating machine learning models into web-based applications, researchers aim to make advanced diagnostic tools accessible to non-specialists and patients. These applications bridge the gap between complex technologies and end users, providing instant and interpretable results. For instance, the web application developed in this project allows users to input their medical parameters and receive a real-time prediction of their heart disease risk. Such tools not only empower individuals but also support healthcare professionals by augmenting traditional diagnostic processes. The intersection of machine learning and healthcare is a rapidly evolving field, with numerous opportunities for advancement. This research builds upon existing efforts to create scalable, efficient, and user-friendly solutions for heart disease prediction. By addressing gaps in traditional diagnostic methods and leveraging the power of data, this project contributes to the broader objective of improving patient outcomes and reducing the global burden of heart disease. It also sets the stage for future explorations into integrating machine learning with wearable devices, real-time monitoring systems, and personalized medicine.

1.2 Literature Survey

The prediction and diagnosis of heart disease have been extensively studied in the field of medical informatics. With the rising prevalence of cardiovascular conditions, researchers have turned to computational techniques to develop reliable predictive models. Early studies primarily relied on statistical methods such as logistic regression, which provided insight into the relationship between specific risk factors and heart disease. While these techniques were foundational, they had limitations in handling complex, non-linear relationships and large datasets. This prompted a shift towards machine learning algorithms that could analyze intricate patterns in data with greater accuracy and efficiency.

Support Vector Machines (SVMs) and Random Forests are among the earlier machine learning models applied in this domain. SVMs have proven effective in binary classification tasks due to their ability to create hyperplanes that separate data points into distinct classes. Random Forests, on the other hand, operate as ensemble methods that combine multiple decision trees to enhance prediction accuracy. Researchers have reported satisfactory results using these models for heart disease prediction, particularly when optimized with feature selection techniques. However, challenges such as overfitting and limited interpretability have been noted, highlighting the need for more robust solutions. Ensemble learning techniques have significantly advanced the field by addressing some of the challenges faced by single-model approaches. Algorithms like AdaBoost and Gradient Boosting have gained popularity for their ability to improve predictive performance by combining weaker classifiers into a strong ensemble. Studies have demonstrated that these methods outperform traditional machine learning models by minimizing errors and adapting to various data distributions. Gradient Boosting, in particular, has shown exceptional promise in medical applications due to its iterative nature, which emphasizes learning from previous mistakes and refining the model over time. The availability of extensive medical datasets has further enhanced the scope of heart disease prediction. Datasets such as the Cleveland Heart Disease dataset and others from UCI Machine Learning Repository have become benchmarks for evaluating predictive models. These datasets contain diverse clinical parameters, such as blood pressure, cholesterol levels, and ECG results, enabling comprehensive model training and testing. Pre-processing techniques, including data normalization, feature scaling, and outlier removal, are often employed to improve data quality and ensure consistency. Studies have highlighted that effective data preparation is as crucial as model selection in achieving high accuracy.

Recent advancements in web-based applications have also influenced the practical deployment of machine learning models for heart disease prediction. Researchers have emphasized the importance of creating user-friendly interfaces that make advanced algorithms accessible to healthcare professionals and patients alike. Web applications integrated with predictive models allow users to input their medical data and receive real-time diagnostic insights. Such tools have been praised for their potential to reduce diagnostic time, enhance decision-making, and extend healthcare services to remote or underserved areas. The growing body of research underscores the transformative potential of machine learning in healthcare. However, challenges such as ensuring model transparency, addressing bias in medical datasets, and validating models in real-world scenarios remain areas of active exploration. This project builds on the existing literature by leveraging ensemble learning techniques, particularly Gradient Boosting, to create a high-accuracy heart disease prediction model. Coupled with a web-based application, the research aims to bridge the gap between advanced computational methods and practical healthcare needs, contributing to a more accessible and efficient diagnostic process.

II. PROPOSED METHODOLOGY

The proposed methodology focuses on developing a machine learning-based predictive system for heart disease, integrating multiple computational and design approaches to ensure accuracy and usability. The core of the system lies in building and deploying an ensemble model, with Gradient Boosting Classifier selected as the optimal algorithm after extensive experimentation. This methodology involves systematic steps, including data collection, preprocessing, model training, evaluation, and deployment through a web application interface. By combining technical rigor and user-centered design, the project ensures both high-performance metrics and practical utility.

Data pre-processing is the first critical step in the methodology. The dataset used includes various clinical features such as age, blood pressure, cholesterol levels, and heart rate, extracted from publicly available medical datasets. Pre-processing steps include handling missing values, scaling numeric features to a standard range, and encoding categorical variables. These measures are essential for eliminating inconsistencies and ensuring that the models receive high-quality input data. Additionally, feature selection techniques are employed to identify the most relevant predictors, thereby reducing noise and improving model performance.

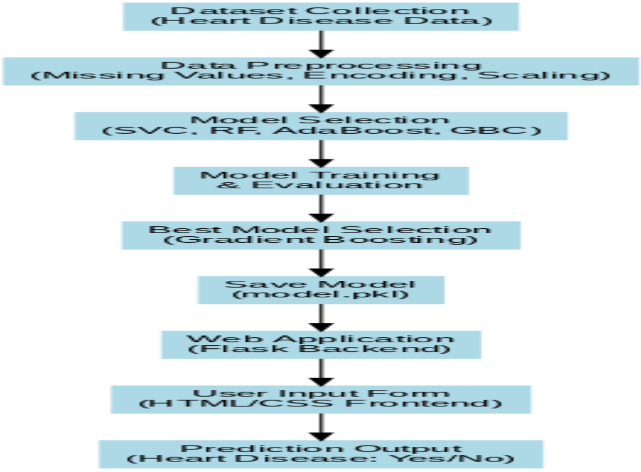


Fig. 1 Machine Learning Workflow for Prediction

The methodology employs four machine learning models—Support Vector Classifier (SVC), Random Forest Classifier, AdaBoost, and Gradient Boosting Classifier. These models are developed using the Python programming language within the Jupyter Notebook environment, leveraging libraries such as scikit-learn. Each model is evaluated using standard performance metrics like precision, recall, F1-score, and accuracy. Among these, the Gradient Boosting Classifier achieves the highest performance due to its iterative learning process, which focuses on reducing residual errors in successive iterations. The model is then serialized using the joblib library and saved as model.pkl for integration into the web application.

The deployment phase involves designing a user-friendly web application. Built using Flask for the backend and HTML/CSS for the frontend, the application allows users to input relevant medical parameters and receive real-time predictions. The backend processes user inputs by loading the pre-trained Gradient Boosting Classifier from the model.pkl file, running the prediction logic, and displaying the results in an intuitive format. The application is designed with accessibility in mind, ensuring it caters to users with varying levels of technical expertise. Security measures, such as input validation, are incorporated to maintain data integrity.

The proposed methodology culminates in creating a robust, accessible diagnostic tool for heart disease prediction. By combining advanced machine learning techniques with thoughtful system design, the project not only demonstrates high predictive accuracy but also addresses the practical challenges of deploying machine learning models in real-world healthcare scenarios. The integration of an ensemble learning approach with a web-based interface ensures the system is both reliable and user-friendly, providing valuable insights to patients and healthcare providers alike.

III. EXPERIMENTS AND RESULTS

The experiments conducted in this project aimed to evaluate the performance of multiple machine learning models for predicting heart disease. Four algorithms—Support Vector Classifier (SVC), Random Forest Classifier, AdaBoost Classifier, and Gradient Boosting Classifier—were implemented and tested on a dataset comprising multiple attributes related to patient health. These attributes included age, sex, chest pain type, resting blood pressure, cholesterol levels, fasting blood sugar, and other critical health indicators. The models were compared using classification metrics like precision, recall, and F1-score, along with a confusion matrix to assess prediction accuracy.

The confusion matrices for the four models — Support Vector Classifier (SVC), Random Forest, AdaBoost, and Gradient Boosting — provide valuable insights into their predictive performance in heart disease classification. Each matrix illustrates the distribution of true positives, true negatives, false positives, and false negatives. Among the models, Gradient Boosting achieved the best performance with 18 true positives and only 6 false negatives, reflecting higher sensitivity. AdaBoost also performed well, correctly classifying 16 positive cases with relatively few misclassifications. Random Forest demonstrated moderate performance with 17 true positives but slightly higher false positives. SVC had the weakest performance with a higher number of false negatives, indicating challenges in correctly identifying positive cases. Overall, the analysis of these confusion matrices helped identify Gradient Boosting as the most reliable model for heart disease prediction due to its superior balance between sensitivity and specificity.

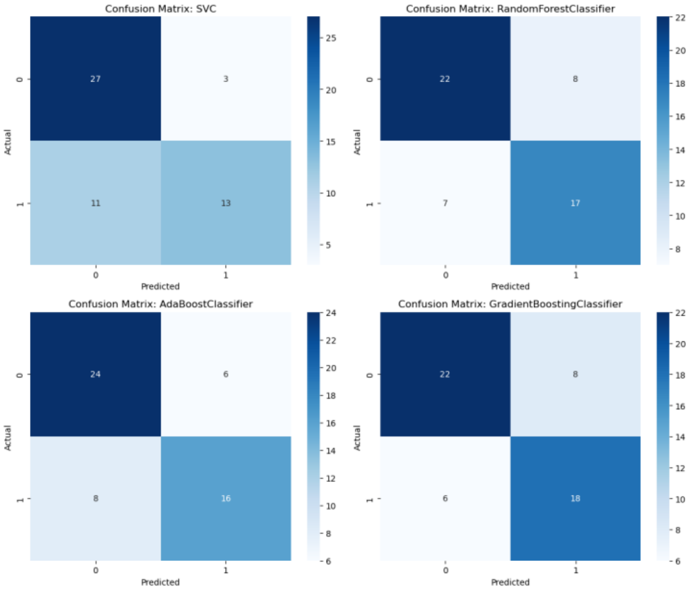


Fig. 2 Confusion matrices of four classifiers.

The bar chart illustrates the accuracy comparison of four machine learning models — SVC, Random Forest Classifier, AdaBoost Classifier, and Gradient Boosting Classifier — applied to heart disease prediction. Among these models, SVC, AdaBoost, and Gradient Boosting achieved the highest accuracy of 74%, indicating consistent performance in distinguishing between individuals with and without heart disease. The Random Forest Classifier demonstrated a slightly lower accuracy of 72%, reflecting marginally reduced predictive capability. These findings emphasize the importance of ensemble learning techniques for improving prediction accuracy in medical diagnosis.

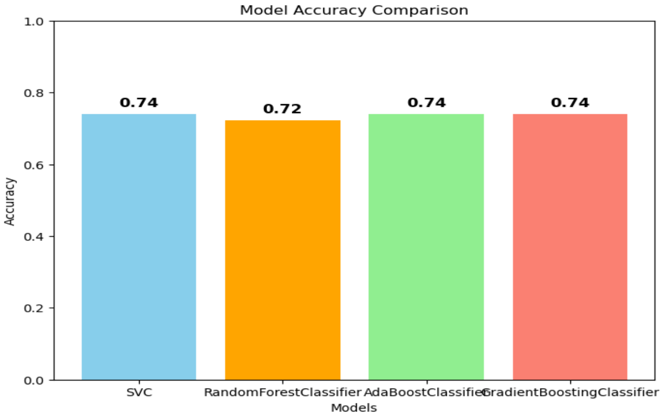
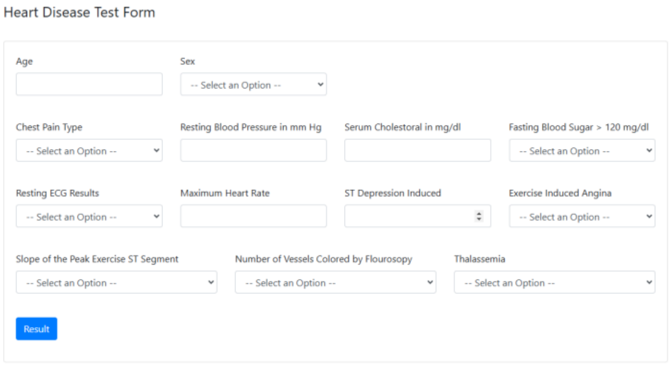


Fig.3 Accuracy comparison of classifiers.

The project features an intuitive and user-friendly graphical user interface (GUI) developed using HTML, CSS, and Flask, designed to ensure easy accessibility and seamless interaction for users from both medical and non-technical backgrounds. This interactive interface serves as a bridge between the user and the machine learning model, allowing for smooth input of crucial medical parameters such as age, gender, chest pain type, resting blood pressure, cholesterol levels, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, and other key indicators relevant to heart disease diagnosis.

The primary goal of the GUI is to simplify the data entry process, ensuring that users can conveniently provide the necessary inputs without requiring in-depth knowledge of the underlying technology. Once the user submits the data, it is sent to the backend where a trained Gradient Boosting Classifier model analyzes the information. This robust and accurate machine learning algorithm processes the inputs to predict the probability of the user having heart disease. The system is designed to deliver fast and reliable predictions, thereby supporting early diagnosis and encouraging timely medical consultation. Overall, the integration of a well-designed frontend with a powerful machine learning backend makes the application both effective and accessible, aiming to contribute positively to healthcare support and preventive diagnostics.



Heart Disease Test Form

Age

Sex: -- Select an Option --

Chest Pain Type: -- Select an Option --

Resting Blood Pressure in mm Hg

Serum Cholestoral in mg/dl

Fasting Blood Sugar > 120 mg/dl: -- Select an Option --

Resting ECG Results: -- Select an Option --

Maximum Heart Rate

ST Depression Induced

Exercise Induced Angina: -- Select an Option --

Slope of the Peak Exercise ST Segment: -- Select an Option --

Number of Vessels Colored by Flourosopy: -- Select an Option --

Thalassemia: -- Select an Option --

Fig. 4 Heart disease prediction input form.

The integration of the GUI bridges the gap between complex machine learning algorithms and practical healthcare applications. Its intuitive design improves usability, making it suitable for real-world scenarios where ease of use is essential. The system demonstrates how machine learning can be applied to assist in early heart disease detection, potentially aiding medical professionals in decision-making. The accompanying GUI image showcases the structured layout of the input form, highlighting the project’s focus on user experience and practical implementation.

IV. CONCLUSION

In conclusion, this project successfully demonstrates the application of machine learning techniques in predicting heart disease, offering a reliable and efficient approach to assist in early diagnosis. By leveraging an ensemble method that combines models such as Support Vector Classifier (SVC), Random Forest, AdaBoost, and Gradient Boosting, the system improves prediction accuracy and robustness. The performance evaluation through confusion matrices and accuracy scores indicates that Gradient Boosting outperforms other models with better classification metrics. This emphasizes the importance of integrating multiple models to mitigate individual model limitations and enhance overall performance. The system's ability to handle critical medical parameters and deliver accurate predictions makes it a promising tool for healthcare applications.

Furthermore, the inclusion of a user-friendly graphical user interface (GUI) enhances accessibility for both medical professionals and individuals with minimal technical expertise. The interactive interface simplifies data entry, ensuring that the prediction process is straightforward and time-efficient. This project not only highlights the potential of machine learning in healthcare but also underscores the significance of combining technological innovation with user-centric design. Future enhancements could explore larger datasets, advanced ensemble techniques, or real-time data integration to further improve prediction accuracy and practicality. Overall, this work contributes to the growing field of predictive healthcare by offering a practical solution for heart disease risk assessment.

REFERENCES

[1] World Health Organization. World Health Statistics 2021. World Health Organization; Geneva, Switzerland: 2021.

[2] Iswisi A.F.A., Karan O., Rahebi J. Diagnosis of Multiple Sclerosis Disease in Brain Magnetic Resonance Imaging Based on the Harris Hawks Optimization Algorithm. BioMed Res. Int. 2021;

[3] Al-Safi H., Munilla J., Rahebi J. Harris Hawks Optimization (HHO) Algorithm based on Artificial Neural Network for Heart Disease Diagnosis; Proceedings of the 2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNWC); Tumkur, India. 3–4 December 2021;

[4] Ternacle J., Côté N., Krapf L., Nguyen A., Clavel M.-A., Pibarot P. Chronic kidney disease and the pathophysiology of valvular heart disease. Can. J. Cardiol.2019;

[5] House A.A., Wanner C., Sarnak M.J., Piña I.L., McIntyre C.W., Komenda P., Kasiske B.L., Deswal A., DeFilippi C.R., Cleland J.G.F. Heart failure in chronic kidney disease: Conclusions from a Kidney Disease: Improving Global Outcomes (KDIGO) Controversies Conference. Kidney Int. 2019

[6] 6.Nguyen T., Wang Z.A. Cardiovascular screening and early detection of heart disease in adults with chronic kidney disease. J. Nurse Pract. 2019

[7] Liu R., Ren C., Fu M., Chu Z., Guo J. Platelet Detection Based on Improved YOLO_v3. Cyborg Bionic Syst. 2022;2022:9780569. doi: 10.34133/2022/9780569.

[8] Mohamed A.A.A., Hançerlioğullari A., Rahebi J., Ray M.K., Roy S. Colon Disease Diagnosis with Convolutional Neural Network and Grasshopper Optimization Algorithm. Diagnostics. 2023

[9] Rashid T. Make Your Own Neural Network. CreateSpace Independent Publishing Platform; Scotts Valley, CA, USA: 2016.

[10] Anderson J., Rainie L., Luchsinger A. Artificial intelligence and the future of humans. Pew Res. Cent. 2018;10:12

[11] Dubey A.K., Choudhary K., Sharma R. Predicting Heart Disease Based on Influential Features with Machine Learning. Intell. Autom. Soft Comput. 2021

[12] Karthick K., Aruna S.K., Samikannu R., Kuppusamy R., Teekaraman Y., Thelkar A.R. Implementation of a heart disease risk prediction model using machine learning. Comput. Math. Methods Med. 2022;

[13] Veisi H., Ghaedsharaf H.R., Ebrahimi M. Improving the Performance of Machine Learning Algorithms for Heart Disease Diagnosis by Optimizing Data and Features. Soft Comput. J. 2021;8:70–85.

[14] Sarra R.R., Dinar A.M., Mohammed M.A., Abdulkareem K.H. Enhanced heart disease prediction based on machine learning and $\chi 2$ statistical optimal feature selection model. Designs. 2022;

[15] Singh A., Kumar R. Heart disease prediction using machine learning algorithms; Proceedings of the 2020 International Conference on Electrical and Electronics Engineering (ICE3); Gorakhpur, India. 14–15 February 2020; pp. 452–457