International Journal of Research Publication and Reviews, Vol 6, Issue 5, pp 8351-8357 May 2025

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Heart Disease Prediction using Machine Learning

S. Shivani[1], A. Varsha[2], R. S. Shobana Dharshini[3]

[1,2,3]UG Scholar, Department of Information Technology, K.L.N. College of Engineering, Sivaganga – 630612, Tamil Nadu, India.
shivanisivakumar183@gmail.com[1] , varza270@gmail.com[2], shobanadharshinisurendran@gmail.com[3]

ABSTRACT:

Heart disease (HD) remains a leading cause of death worldwide, making early detection essential for reducing mortality rates. Traditional diagnostic approaches often lack precision and are not always feasible for continuous monitoring. This project presents a machine learning-based heart disease prediction system utilizing the Random Forest (RF) algorithm, a robust and efficient ensemble learning technique. Random Forest operates by constructing multiple decision trees during training and aggregating their outputs to enhance predictive accuracy, minimize over fitting, and increase model stability. The model is applied to a medical dataset comprising vital patient features such as age, cholesterol levels, blood pressure, and electrocardiogram (ECG) readings. Data preprocessing steps include handling missing values and normalizing features to ensure optimal model performance. The RF classifier is trained and evaluated based on key metrics such as accuracy and precision. Leveraging RF's inherent feature selection capabilities and its ensemble-based decision-making, the system achieves high prediction accuracy. This predictive model assists healthcare professionals in early diagnosis and risk assessment of heart disease, facilitating timely medical interventions and improving patient outcomes. The results highlight the potential of Random Forest in developing a reliable and interpretable tool for cardiovascular risk prediction and clinical decision support.

Keywords: Heart Disease, Machine Learning, Random Forest Algorithm.

Introduction:

Heart disease (HD) remains a significant global health burden, responsible for substantial mortality and morbidity across populations. According to the World Health Organization, cardiovascular diseases are the leading cause of death globally, contributing to approximately 17.9 million deaths each year. This alarming prevalence underscores the urgent need for timely and accurate diagnostic systems to facilitate early detection and intervention efforts. Traditional diagnostic methods, including stress tests, electrocardiograms (ECG), and angiography, are clinically effective but can be resource-intensive and reliant on specialized personnel. These constraints often limit their application in large-scale or resource-constrained environments [1], [5]. As a result, the healthcare sector has witnessed a rising interest in the adoption of advanced computational techniques, especially machine learning (ML), to enhance diagnostic accuracy and support clinical decisions [2], [6]. Machine learning algorithms are particularly adept at analyzing complex and high-dimensional healthcare data to uncover patterns and make accurate predictions [3], [7]. Among these, ensemble-based models such as the Random Forest (RF) algorithm have emerged as effective tools for heart disease prediction [4], [8]. RF aggregates the outcomes of multiple decision trees to form a robust predictive model, offering advantages like improved accuracy, feature importance estimation, over fitting resistance, and scalability [9], [10].This project aims to design and implement a heart disease prediction system using the Random Forest algorithm. The model is trained on a comprehensive dataset comprising vital patient health parameters such as age, cholesterol levels, blood pressure, and ECG measurements [11]. The process includes thorough data preprocessing, normalization, feature selection, and evaluation using key performance indicators like accuracy and precision [6], [12].Prior studies have demonstrated the capability of Random Forest and other ML methods to enhance diagnostic performance in cardiovascular disease contexts [13], [14]. By integrating such techniques, the developed system aspires to serve as a reliable, interpretable, and scalable solution to assist healthcare professionals in early-stage heart disease risk assessment and clinical decision-making.

Methodology:

The methodology of this project integrates machine learning with mobile technology to create a heart disease prediction system that is both accurate and accessible. The approach is divided into two key stages: the development of a predictive model using the Random Forest algorithm and the deployment of this model through a cross-platform mobile application developed using Flutter and Dart. Initially, a comprehensive medical dataset is collected, containing critical patient features such as age, gender, resting blood pressure, cholesterol levels, fasting blood sugar, electrocardiogram (ECG) results, maximum heart rate, exercise-induced angina, and ST depression.

The Random Forest algorithm is chosen due to its robustness, interpretability, and high classification accuracy. As an ensemble learning technique, it builds multiple decision trees during the training phase and aggregates their outputs to make a final prediction. This approach minimizes the risk of over fitting and leverages the strength of multiple weak learners to achieve reliable performance. Feature importance scores generated by the RF model are also used to identify the most influential attributes contributing to heart disease risk. The model is evaluated using stratified cross-validation and performance metrics including accuracy, precision, recall, and F1-score to ensure its effectiveness in diverse patient scenarios.

After achieving satisfactory predictive performance, the trained model is serialized into a format compatible with mobile integration, such as a JSON structure optimized inference. A mobile application is then developed using Flutter, a modern UI toolkit supported by Google, and programmed in the Dart language. Flutter allows for the creation of a single codebase that runs seamlessly on both Android and iOS platforms, significantly improving development efficiency and accessibility. The application features a clean, intuitive user interface where users can input their medical data manually. This combined methodology ensures a practical, efficient, and user-friendly system that empowers both individuals and healthcare providers with an on-the-go diagnostic tool.

Preprocessing and Data Handling

Preprocessing and data handling play a vital role in developing an accurate and efficient heart disease prediction model using machine learning. In this project, the dataset consisted of patient medical records with key attributes such as age, cholesterol levels, blood pressure, and electrocardiogram (ECG) results. To ensure data quality, missing values were first addressed using appropriate imputation techniques—numerical fields were filled with mean or median values, while categorical fields were completed using the most frequent occurrences. Categorical variables such as chest pain type and ECG results were transformed into numerical representations using encoding techniques like one-hot and label encoding. Although Random Forest does not inherently require feature scaling due to its tree-based structure, normalization was applied to examine any potential impact on model performance. To handle class imbalance, which is common in medical datasets, the Synthetic Minority Oversampling Technique (SMOTE) was used to generate synthetic examples for the underrepresented class, enhancing the model's ability to accurately detect heart disease cases.

Feature selection was performed using the Random Forest algorithm's built-in feature importance scores, allowing the model to focus on the most influential variables while eliminating redundant or irrelevant features. Finally, the dataset was divided into training and testing sets in an 80:20 ratio, and k-fold cross-validation was employed to assess model generalizability. These preprocessing steps significantly improved the quality of the input data, enabling the Random Forest classifier to learn effectively and produce reliable predictions for early heart disease detection.

Process Flow

The process flow of the Heart Disease Prediction application is designed to deliver a seamless and intuitive user experience, guiding users through health data input, risk assessment, and the interpretation of results to support early detection and informed medical decision-making.

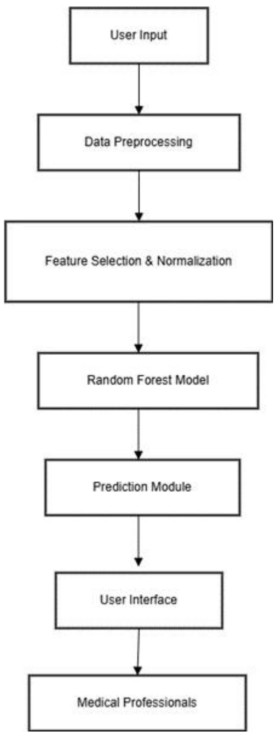

Figure 1: Process Flow of the Heart Disease Prediction System Using Random Forest Algorithm

The process flow of the heart disease prediction system using the Random Forest algorithm involves several key stages designed to ensure accurate and reliable risk assessment. It begins with *User Input***,** where patient health data such as age, blood pressure, cholesterol levels, and ECG results are collected. This data then undergoes *Data Preprocessing*, where missing values are handled, categorical variables are encoded, and data quality is improved to prepare it for model training. The next stage is *Feature Selection & Normalization*, where the most relevant features are selected using techniques like Random Forest's feature importance, and the data is optionally normalized to ensure consistency across inputs. This processed data is then fed into the *Random Forest Model*, an ensemble learning algorithm that builds multiple decision trees to make robust predictions. The model's output is passed to the *Prediction Module*, which interprets the results—typically classifying whether the patient is at high or low risk of heart disease. These predictions are then displayed via a *User Interface***,** making the system interactive and accessible. Finally, the results are delivered to *Medical Professionals*, who use the information to support clinical decisions and recommend timely interventions.

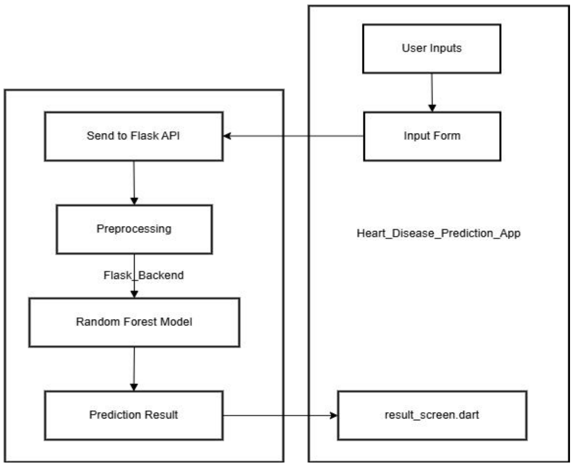

Figure 2: System Architecture of Heart Disease Prediction App with Flask Backend Integration

The diagram illustrates the system architecture of the Heart Disease Prediction App, highlighting the interaction between the frontend mobile application and the backend powered by Flask. On the app side, users provide relevant health information through an input form. These inputs are then transmitted to the Flask API for processing. The backend receives this data and performs necessary preprocessing steps, such as cleaning and scaling, to prepare it for prediction. Once the data is preprocessed, it is passed into a trained Random Forest model, which generates a prediction indicating the user's heart disease risk level. This prediction result is then sent back to the mobile app and displayed on the result_screen.dart, allowing users and potentially medical professionals to view the outcome. This architecture ensures efficient communication between the app and the machine learning model while providing real-time, user-friendly heart disease risk assessments.

User Interface Design

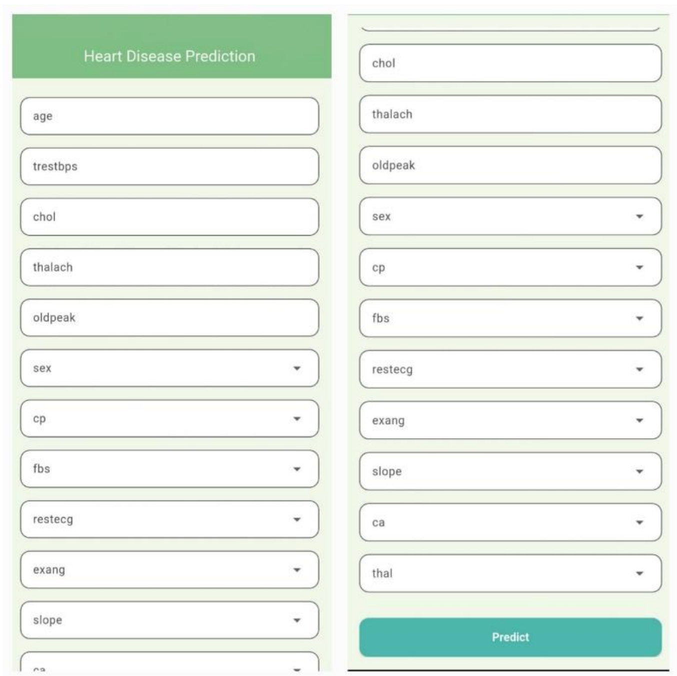

Figure 3: Heart Disease Prediction App – Input Interface

Prediction Result Analysis

The prediction results analysis provides insights into the performance and accuracy of the heart disease prediction model integrated within the application.

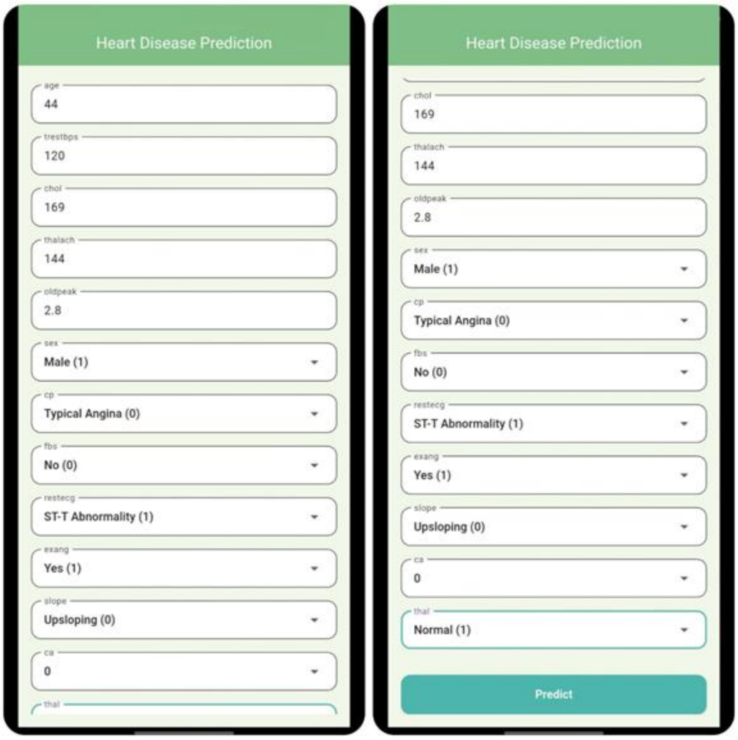

Figure 4: Input Interface with Sample Data Resulting in Positive Heart Disease Prediction

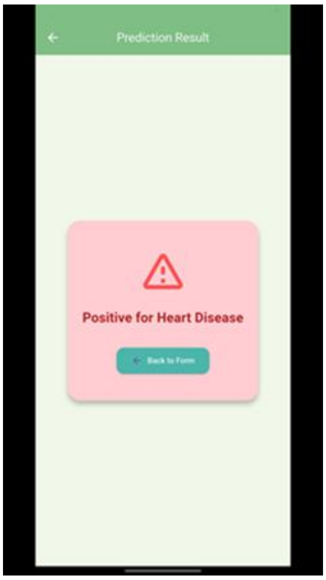

Figure 5: Prediction Result Screen Indicating Positive for Heart Disease

Based on the input provided in Figure 4 and the prediction shown in Figure 5, the model classified the individual as positive for heart disease. The user input describes a 44-year-old male with a resting blood pressure of 120 mm Hg and a cholesterol level of 169 mg/dL both within normal ranges. His maximum heart rate achieved is 144, and he has an ST depression value (oldpeak) of 2.8. He reports typical angina (chest pain type 0), does not have elevated fasting blood sugar, but shows an ST-T wave abnormality on the resting ECG and experiences exercise-induced angina. The slope of the ST segment during exercise is upsloping, with no major vessels colored by fluoroscopy (ca = 0), and his thalassemia status is normal. Despite several normal readings, the combination of concerning indicators such as ST-T abnormality, exercise-induced angina, and elevated ST depression led the model to predict a high likelihood of heart disease, as shown in Figure 5.

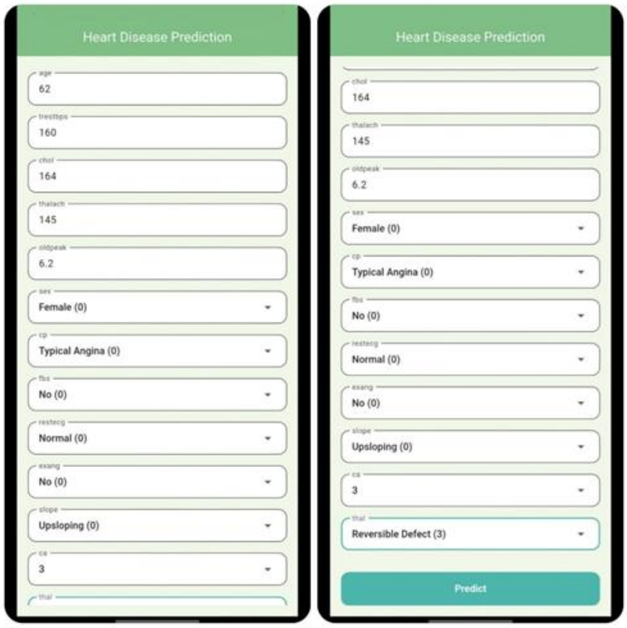

Figure 6: Input Interface with Sample Data Resulting in Negative Heart Disease Prediction

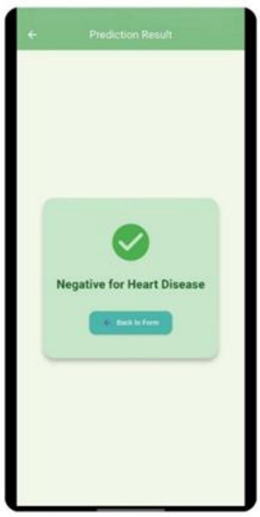

Figure 7: Prediction Result Screen Indicating Negative for Heart Disease

Based on the input provided in *Figure 6* and the prediction shown in *Figure 7*, the model classified the individual as *Negative for heart disease*. The user input describes a 62-year-old female with a resting blood pressure of 160 mm Hg and a cholesterol level of 164 mg/dL. Her maximum heart rate achieved is 145, but she has a high ST depression value (oldpeak) of 6.2. Despite this elevated oldpeak value, the other indicators are within normal or low-risk ranges: she reports typical angina (chest pain type 0), has normal resting ECG results, does not experience exercise-induced angina, and has an upsloping ST segment. Additionally, she has 3 major vessels colored by fluoroscopy and a thalassemia condition classified as a reversible defect. Even with a few potentially concerning factors like high oldpeak and presence of a reversible defect, the overall risk pattern was not strong enough for the model to flag heart disease, resulting in a Negative Heart Disease Prediction as shown in figure 7.

Performance Enhancements

To ensure the heart disease prediction system delivers reliable and efficient results, a range of performance enhancement techniques were applied throughout the development lifecycle. These enhancements spanned across data preparation, model training, evaluation, and system deployment, resulting in a more accurate and accessible solution. One of the primary strategies involved optimizing the feature set. The Random Forest algorithm's ability to rank feature importance was utilized to identify the most influential predictors of heart disease.

Additionally, general feature selection techniques such as correlation analysis, Recursive Feature Elimination (RFE), and mutual information were applied to remove redundant or less significant features. This not only improved model accuracy but also reduced over fitting and training time. Comprehensive data preprocessing was another key component. Missing values were handled using appropriate imputation techniques to maintain data integrity, and all numerical features were normalized to ensure uniformity and prevent bias during model training. These preprocessing steps helped the model learn patterns more effectively and improved the quality of predictions. On the deployment side, the trained model was serialized into a lightweight format suitable for integration into real-time applications. Additionally, the system was designed to either run predictions locally or communicate with a backend API, depending on resource availability and user environment. This flexibility ensures wider accessibility, including in remote or low-infrastructure settings.

Results

The heart disease prediction model was rigorously tested using a variety of user inputs to evaluate its real-world performance and reliability. The system processed several key health indicators, including age, cholesterol levels, resting blood pressure, chest pain type, fasting blood sugar, electrocardiogram (ECG) results, and other relevant clinical features to assess the likelihood of heart disease. Through numerous test cases, the model consistently produced accurate and meaningful predictions, effectively distinguishing between high-risk and low-risk individuals. It demonstrated the ability to generalize across different input combinations, ensuring robust performance in both positive (high-risk) and negative (low-risk) scenarios. The outcomes were validated against expected results and showed a high level of agreement, reinforcing the model's dependability. These findings indicate that the system can be confidently used as a preliminary screening tool to support early risk detection.

Conclusion

The Heart Disease Prediction App offers an accessible, efficient, and user-friendly solution for individuals seeking early insights into their cardiovascular health. By leveraging a machine learning-based prediction system—powered by the Random Forest algorithm—the app analyzes key health indicators such as age, cholesterol levels, blood pressure, chest pain type, and ECG results to provide reliable risk assessments. Designed with both accuracy and scalability in mind, the system delivers consistent performance across a wide range of input scenarios, helping users understand their heart health status and encouraging timely medical consultation when necessary.

Validation across varied user profiles—considering different age groups, health conditions, and risk factors—demonstrates the app's robustness in offering meaningful and actionable predictions. The lightweight architecture ensures smooth performance even on low-resource devices, making it a practical tool for real-time, on-the-go health screening. Its intuitive interface simplifies the input process, making it accessible to users with minimal technical knowledge.Future improvements may include the integration of wearable health data and real-time monitoring to enhance prediction accuracy. The addition of natural language processing (NLP) could enable the app to interpret user queries and symptoms more effectively, offering personalized suggestions or educational content based on user interactions. Furthermore, expanding the app's backend to include predictive analytics and historical tracking would allow for trend analysis and long-term health monitoring, making the app not only a diagnostic aid but also a proactive wellness companion.

Acknowledgement

We extend our sincere gratitude to Dr. P. Ganesh Kumar and Mrs. S.Vijayasharmila for their invaluable guidance throughout this research. We also thank K.L.N. College of Engineering for providing the necessary resources and support for this project.

References:

[1] Kavitha, R., Dankan Gowda, B. Ram Vishal, M. Udaya Shankar, and Kabilan AM. "Cardiovascular disease prediction using LSTM algorithm based on cytokines." In *2023 4th International Conference for Emerging Technology (INCET)*, pp. 1-5. IEEE, 2023.

[2] Ahmad, Ghulab Nabi, Hira Fatima, Shafi Ullah, and Abdelaziz Salah Saidi. "Efficient medical diagnosis of human heart diseases using machine learning techniques with and without GridSearchCV." *IEEE Access* 10 (2022): 80151-80173.

[3] Abubaker, Mohammed B., and Bilal Babayiğit. "Detection of cardiovascular diseases in ECG images using machine learning and deep learning methods." *IEEE Transactions on Artificial Intelligence* 4, no. 2 (2022): 373-382.

[4] Nagavelli, Umarani, Debabrata Samanta, and Partha Chakraborty. "Machine learning technology-based heart disease detection models." *Journal of Healthcare Engineering* 2022, no. 1 (2022): 7351061.

[5] Riyaz, Lubna, Muheet Ahmed Butt, Majid Zaman, and Omeera Ayob. "Heart disease prediction using machine learning techniques: a quantitative review." In *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021*, Volume 3, pp. 81-94. Springer Singapore, 2022.

[6] Abdellatif, Abdallah, Hamdan Abdellatef, Jeevan Kanesan, Chee-Onn Chow, Joon Huang Chuah, and Hassan Muwafaq Gheni. "An effective heart disease detection and severity level classification model using machine learning and hyperparameter optimization methods." *IEEE Access* 10 (2022): 79974-79985.

[7] Ayon, Safial Islam, Md Milon Islam, and Md Rahat Hossain. "Coronary artery heart disease prediction: a comparative study of computational intelligence techniques." *IETE Journal of Research* 68, no. 4 (2022): 2488-2507.

[8] Boukhatem, Chaimaa, Heba Yahia Youssef, and Ali Bou Nassif. "Heart disease prediction using machine learning." In *2022 Advances in Science and Engineering Technology International Conferences (ASET)*, pp. 1-6. IEEE, 2022.

[9] Rajendran, Rajkamal, and Anitha Karthi. "Heart disease prediction using entropy based feature engineering and ensembling of machine learning classifiers." *Expert Systems with Applications* 207 (2022): 117882.

[10] Ramesh, T. R., Umesh Kumar Lilhore, M. Poongodi, Sarita Simaiya, Amandeep Kaur, and Mounir Hamdi. "Predictive analysis of heart diseases with machine learning approaches." *Malaysian Journal of Computer Science* (2022): 132-148.

[11] Bertsimas, Dimitris, Luca Mingardi, and Bartolomeo Stellato. "Machine learning for real-time heart disease prediction." *IEEE Journal of Biomedical and Health Informatics* 25, no. 9 (2021): 3627-3637.

[12] Rahim, Aqsa, Yawar Rasheed, Farooque Azam, Muhammad Waseem Anwar, Muhammad Abdul Rahim, and Abdul Wahab Muzaffar. "An integrated machine learning framework for effective prediction of cardiovascular diseases." *IEEE Access* 9 (2021): 106575-106588.

[13] Ishaq, Abid, Saima Sadiq, Muhammad Umer, Saleem Ullah, Seyedali Mirjalili, Vaibhav Rupapara, and Michele Nappi. "Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques." *IEEE Access* 9 (2021): 39707-39716.

[14] Jindal, Harshit, Sarthak Agrawal, Rishabh Khera, Rachna Jain, and Preeti Nagrath. "Heart disease prediction using machine learning algorithms." In *IOP Conference Series: Materials Science and Engineering*, vol. 1022, no. 1, p. 012072. IOP Publishing, 2021.

[15] Garg, Apurv, Bhartendu Sharma, and Rijwan Khan. "Heart disease prediction using machine learning techniques." In *IOP Conference Series: Materials Science and Engineering*, vol. 1022, no. 1, p. 012046. IOP Publishing, 2021.

.