Grand Asian journal of Computing and Emerging Technologies (GJCET)

# AI powered heart disease cardiovascular prediction using ML Algorithms

*Zain ul Abedeen[1]*, Waqar Azeem Khan[2], Huma Sarfaraz[1], Shahmeen Zahid[1], Amna Bibi[1] Eman Anwar[1]*

Received data:20 June 2025 Revised date: 10 July 2025 Accepted date: 14 July 2025 Published date: 31 August 2025



**Abstract:** This paper presents an AI-powered heart disease prediction system using machine learning algorithms. The system is implemented in an Android application that collects user health data and predicts the presence of heart disease in real-time. The model uses Naive Bayes classification and is connected to the app through a Flask API. The app is equipped with functionalities like user login/signup, chat-bot assistance, dietary recommendations, and a history of past predictions. The goal is to provide a user-friendly tool that enables early risk detection and promotes heart health awareness. Considering the rapid growth in AI-driven healthcare solutions, this project has the potential to make a significant impact in Pakistan**,** addressing a critical need in preventive cardiology. We achieved different accuracy using different models Naïve Bayes is 95%, Decision tree is 84%, CNN is 92% and ANN is 94%.

**Keywords**: Heart Disease, Prediction, Machine Learning, Android Application, Real-time risk.

Issue (1) & volume (1)

Issue Date August 31, 2025

[1]*Department of Computer science, University of Sialkot, Sialkot, Pakistan*

[2]*Department of Computer science, Abbottabad university of science and technology, Pakistan, Email; abbottexpert@gmail.com*

**Corresponding Author*

*zain.abedeen@uskt.edu.pk*

## 1. Introduction

Heart disease remains one of the leading causes of morbidity and mortality worldwide, posing significant challenges to healthcare systems and impacting millions of lives. Early detection and accurate prediction of cardiovascular conditions are critical for effective treatment and management, ultimately improving patient outcomes and reducing healthcare costs. The implementation of this predictive system can thetransform preventive cardiology, enabling timely interventions and promoting better health management strategies for patients at risk of cardiovascular diseases. Cardiovascular diseases (CVDs) are the leading cause of death globally, taking an estimated 17.9 million lives each year. CVDs are a group of disorders of the heart and blood vessels and include coronary heart disease, cerebrovascular disease, rheumatic heart disease and other conditions. More than four out of five CVD deaths are due to heart attacks and strokes, and one third of these deaths occur prematurely in people under 70 years of age (Telsang et al., 2024).

## 2. Problem Statement

Heart disease remains a leading cause of mortality worldwide, making early detection and accurate risk prediction essential for effective management. Existing Heart disease detection systems primarily depends on traditional diagnostic tools such as ECGs, blood tests, and physical examinations, which often require hospital visits and specialized medical professionals (Sun et al., 2023). Our project aims to address this gap by creating a "Heart disease prediction system" that uses models of

machine learning to analyze data of patients and predict the likelihood of heart disease with high accuracy (Baumgartner et al., 2022).

### 2.1 Heart Disease Cardiovascular Prediction System

Main objective of this paper is to develop of a prediction system of heart disease cardiovascular disease. A ML (machine learning) based prediction system that analyzes a patient's medical history, lifestyle, and genetic data to predict their risk of heart disease. The Heart Disease Cardiovascular Prediction System leverages advanced machine learning algorithms to patient data, including medical history, lifestyle factors, and diagnostic test results. By processing this 2 data, the system provides healthcare professionals with accurate predictions of an individual's risk for heart disease. This predictive tool is designed to facilitate early diagnosis and personalized treatment planning, ultimately improving patient outcomes and reducing healthcare costs (Sajja & Kalluri, 2020).

## 3. Literature Review

The literature review for the "Heart Disease Cardiovascular Prediction System" explores various research studies, methodologies, and advancements in the prediction of heart disease using machine learning and artificial intelligence. This review covers the key areas of data collection, machine learning algorithms, deep learning approaches, feature selection, and integration with healthcare systems (Balasubramaniam et al., 2022). The study evaluates multiple machine learning algorithms for heart disease prediction (Telsang et al., 2024), comparing their accuracy to determine the most effective model. The tested algorithms include Logistic Regression (LR), Naïve Bayes (NB), K-Nearest Neighbors (KNN), Random Forest (RF), Gradient Boosting Machine (GBM), Support Vector Machine (SVM), and CatBoost (CB). The accuracy results indicate that CatBoost (CB) achieved the highest performance with 94.28%, followed by Gradient Boosting Machine (GBM) at 92.46% and Support Vector Machine (SVM) at 87.15%. While Logistic Regression (80.38%) and Naïve Bayes (78.69%) also showed competitive results, KNN had the lowest accuracy at 67.21%. Despite the claim that SVM is the most accurate model, the findings suggest that CatBoost outperformed all other algorithms (Chu et al., 2021). The paper explores the role of artificial intelligence (AI) in the diagnosis and treatment of cardiovascular diseases (2023), including heart failure, atrial fibrillation, valvular heart disease, hypertrophic cardiomyopathy, and congenital heart disease. AI techniques, such as Convolutional Neural Networks (CNN), Logistic Regression, Support Vector Machines (SVM), and K-Nearest Neighbors (KNN), have been applied to enhance disease detection and management (Wang et al., 2021) Current Status of Nutritional Knowledge Regarding Cardiovascular Disease (CVD) (2023) The review highlights a significant knowledge gap regarding cardiovascular diseases (CVDs) and their risk factors, particularly among individuals with lower socioeconomic status. This limited understanding can contribute to poor prevention and management of CVDs (Faheem & Faheem, 2023)

## 4. General Methodology

The methodology for creating app of prediction system involves several phases:

- Planning
- Development
- Deployment phase
- Conceptual view
- Model training
- Working of app
- Testing
- Results

In the "Heart Disease Cardiovascular Prediction System," various machine learning models are employed to predict the likelihood of heart disease. Each model has its own strengths and weaknesses, making it suitable for different types of data and prediction requirements. Below are detailed descriptions of the machine learning models used. We train different models to check out the high accuracy to predict heart disease prediction system.

**Table 1.** Summary of recent studies on cardiovascular disease prediction, diagnosis, and risk assessment using machine learning, deep learning, and statistical models (2020–2024).

| Year | Paper Title | Model Used | Results |
|---|---|---|---|
| 2024 | Heart Disease Predication using Heart Machine Learning (Telsang et al., 2024) | Methodology to determine which one is best:<br>Logistic Regression<br>Random Forest<br>Navies Bayes<br>Cat Boast<br>KNN<br>SVM<br>Gradient Boosting Machine | Algorithms/Accuracy<br>LR: 80.38%<br>NB: 78.69%<br>KNN: 67.21%<br>RF:73.77%<br>GBM:92.46%<br>SVM:87.15%<br>CB:94.28%<br>Hence Most accurate ML algorithm is SVM |
| 2023 | Artificial intelligence in cardiovascular diseases: diagnostic and therapeutic perspectives (Sun et al., 2023) | CNN model<br>Logistic Regression<br>SVM<br>KNN | AI algorithms have demonstrated the potential to enhance disease stratification, typing, and outcome prediction for CVD patients, surpassing the capabilities of traditional methods with 81 % Accuracy. |
| 2023 | Current Status of Nutritional Knowledge regarding cardiovascular Disease (CVD)(Faheem & Faheem, 2023) | Search engines used to locate with (2010—2023) timeframe were Google& Google Scholar. Included, cross-Sectional observational, case control | The review found that: Many people have limited understanding of CVD and risk factors. The knowledge Gap is particularly because of lower socioeconomic status. |
| 2022 | The Global Burden of Cardiovascular Diseases and Risk (Vaduganathan et al., 2022) | The study employs models to estimate the burden of CVDs attributable to various risk factors globally, regionally, and nationally | High systolic blood pressure remains the leading modifiable risk factor for premature cardiovascular deaths globally, Lead exposure is a major contributor to CVD mortality, linked with hypertension, stroke, coronary artery disease. Reduced kidney function is also a significant contributor to cardiovascular deaths |
| 2022 | Feature Selection and Dwarf Mongoose Optimization Enabled Deep Learning for Heart Disease Detection (Balasubramaniam et al., 2022) | Squeeze Net is employed for heart disease prediction. Dwarf Mongoose Optimization Algorithm (DMOA) is utilized to optimize Squeeze Net. | The DMOA-Squeeze Net method achieved a maximum accuracy of 0.925, sensitivity of 0.926, and specificity of 0.918. |
| 2021 | Impact of established cardiovascular disease on 10-year death after coronary revascularization for complex coronary artery disease (Jindal et al., 2021) | The study utilized the SYNTAX score II 2020 to identify patients who would benefit the most from either CABG or PCI | Patients with established CVD had a significantly higher risk of 10-year all cause death compared to those without CVD. |

| | | | |
|---|---|---|---|
| 2021 | Heart disease prediction using machine learning algorithms (Chu et al., 2021) | The models employed in the research include Logistic Regression, KNN, and Random Forest Classifiers to predict the likelihood of heart disease in patients | The models achieved an average accuracy of 87.5%, with KNN showing the highest accuracy of 88.52% among the algorithms used. |
| 2021 | Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning (Wang et al., 2021) | Random forest, Logistic Regression, SVM Decision Tree, XGBoost Deep Learning (Sequential model approach 128 neurons, ActivationFunction ReLu) | Machine learning models achieve accuracy rates, with SVM and Random Forest performing well among classifiers tested. Deep learning approaches, Particularly sequential models, outperform traditional algorithms, achieving an impressive accuracy rate of 94.2%. |
| 2021 | Roles of Anxiety and Depression in Predicting Cardiovascular Disease Among Patients with Type 2Diabetes Mellitus: A Machine-Learning Approach. (Nashif et al., 2021) | The study utilized statistical tests, including the Mann Whitney U-test, t-test, and Chi-Square tests, for data analysis. Mean Impact Value (MIV) was employed to identify significant predictors in the DNN model. | Results showed that the developed ensemble model for CVD risk achieved an AUC score of 0.91, accuracy of 87.50%, sensitivity of 88.06%, and specificity of 87.23%. |
| 2020 | A Deep Learning Method for Prediction of Cardiovascular Disease Using Convolutional Neural Network (Sajja & Kalluri, 2020) | Data Preprocessing Model Architecture: Training and Testing: Comparison | The proposed CNN model achieved the following results: Training Accuracy: 95.04% Testing Accuracy: 94.78% |

**4.1 Methodology of ML Model**

This section describes the methodology proposed in this project. This project starts with input data from patients, followed by the implementation of different models like Naïve Bayes, Random Forest, KNN, SVM for the best prediction of the heart disease cardiovascular of humans. Then the final result will be displayed. The overall flow of our project is shown below in form of block diagram.

- Data Collection A series of photos are loaded into a model that is being constructed. This stage is helpful for finetuning the model according to the kind of symbol.
- Training and Testing Phase. In this step, the model receives a set of inputs and specific output is anticipated. The model's accuracy is determined based on the output.

- **4.2 Proposed Methodology**
- The proposed methodology for the heart disease prediction system involves the integration of an AI model within an Android application to allow real-time predictions. The process begins with data collection, where the user inputs health-related information such as age, systolic and diastolic blood pressure, cholesterol levels, glucose levels, smoking status, physical activity, and gender. This input data is sent from the Android application to a Python Flask-based REST API.

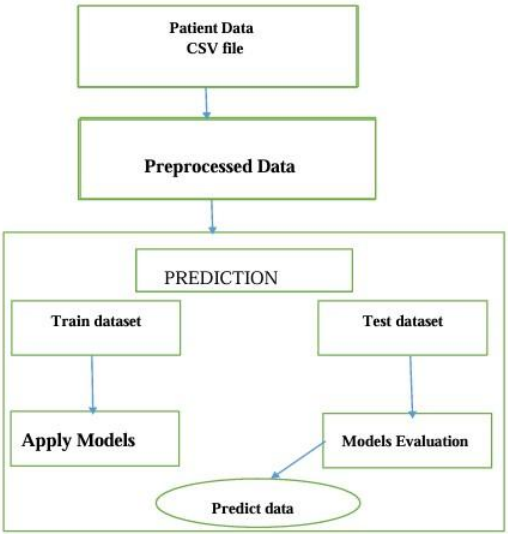


**Fig. 1.** Block diagram of model evaluation

Once the model generates a prediction indicating the presence or absence of heart disease the API sends the result back to the Android application. The result is displayed on the prediction screen and simultaneously stored in Firebase Realtime Database to enable users to view their past reports. Security and usability are ensured using Firebase Authentication for login and signup processes. Other supporting features include a chatbot (for basic heart health queries), a diet chart (suggesting heart-friendly food options), and interactive educational components such as quizzes and tutorials. This methodology ensures a smooth integration of machine learning and mobile development, providing an accessible tool for preliminary heart disease risk assessment.

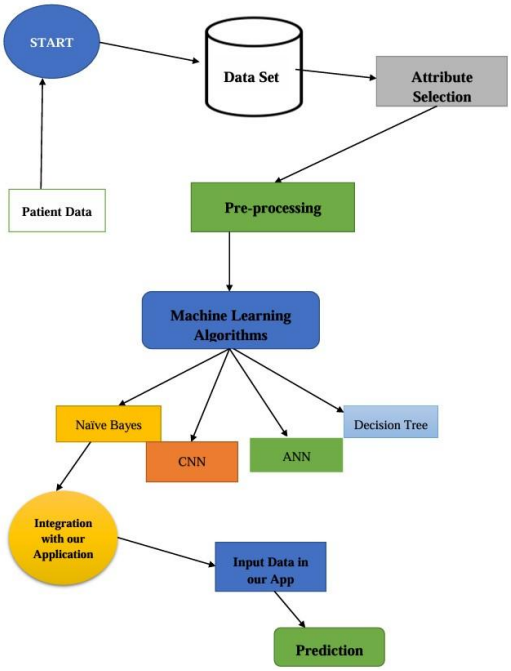


**Fig. 3.** Flowchart of proposed Methodology

## 5. Dataset Preparation

There are several datasets available on Heart disease. We used the heart disease cardiovascular dataset. This was published 5 years ago at the International Conference on Machine Learning. This is an open source dataset and was shared publicly for the Kaggle competition. This Data was based on 70,000 patients and 11 features.

**Features:**

- Age
- gender
- weight
- ap-hi
- ap-lo
- cholesterol
- glucose
- smoke
- alcohol
- active
- cardio

### 5.1 Data Preprocessing & Data Splitting

- The processed Reading the dataset from CSV.
- Removing null or irrelevant values.
- Encoding categorical variables like gender, smoking.
- Feature scaling using MinMax-Scaler.

The train test split function from the selection package divides the dataset into training (85%) and testing (15%) subgroups. This ensures that, after being trained on most of the data, the model is evaluated on a distinct sample of data to determine how well it performed.

## 6. Model Evaluation and Deployment

The model with the highest validation accuracy is loaded. On the testing dataset, predictions are produced in order to assess performance is shown in fig. 4. The evaluation stage verifies the model's excellent capacity for generalizing to fresh data.

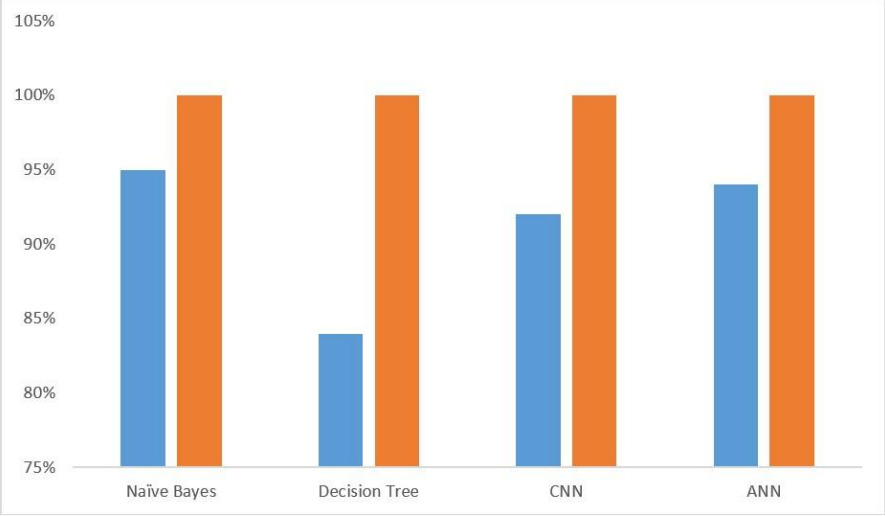


**Fig. 4.** Algorithm Results

### 6.1 Naïve Bayes

The working of naïve bayes is a classification tasks. Here's a step-by-step explanation of how its work.

- Data preparation
- Class probability

We have a dataset of cardiovascular disease and have 11 features but we use features of dataset that is cardio, have find accuracy 95%. We already defined dataset previously. Here is a clear view of our model training of Naïve bayes and have accuracy 95%. We take some new features and add in our dataset for better accuracy. And after that we have final accuracy of naïve bayes that we

defined. With original dataset we have 90% accuracy almost but after some changing we have accuracy 95%. It is easy to implemented and fast algorithm. Also it is good for binary classification tasks. But it is not suitable for complex dataset. Naïve bayes can be biased if the class probabilities are not estimated accurately.

**6.2 Decision Tree**

A decision tree is a type of machine learning algorithm that uses tree-like model to classify data or make predictions. Here's a step-by-step explanation of how we use decision tree model and what's the accuracy we have of our dataset.

- Splitting
- Decision nodes
- Predictions
- Classification
- Select the best features to split the data. Our best features are Cardio and ID.
- Assign a prediction class.
- Use decision tree to make predictions on features.
- We have accuracy on decision tree is 84%.
- Now see of the accuracy pic.

**6.3 CNN**

CNN are typically used for image and signal processing tasks, but they can also be applied on numerical dataset. Here's an explanation of our train model on numerical dataset:

- Reshaping the data
- Normalize the data
- Pooling layers
- Fully connected layers for predictions
- Input layer are (11*1), pooling layers are max pooling, stride=2
- Fully connected layers (128*1)
- Output layer (5*1)

But our dataset doesn't contain more features to produce the better accuracy. We have only 11 features which is not enough for CNN. So accuracy rate is only 92%.

Here is the view of CNN;

- This is how we change numerical data into 1D, 2D, pooling layers.
- Accuracy we have in this dataset.

**6.4 ANN**

ANN is a computational model inspired by the structured and function of human brain. Here's a step-by-step explanation of how we train this model:

- Input layers
- Activation functions
- Hidden layers and output

**7. Application Development Functions**

Heart disease prediction app is an android based application that allows users to use for prediction of heart disease that is user have disease or not. Basically this activity leads the main role in this project. The activity performs the features of dataset. Activity performs to predicts the disease of heart that user have or not.

**Chat-bot activity**

Heart disease prediction app contains chat-bot functions.

- A chatbot that answer about heart disease.it is user input activity
- It uses NLP algorithm to analyze the users input, entities and context.
- It searches its knowledge base or database to find relevant information.
- It delivers its response to the users, through text.

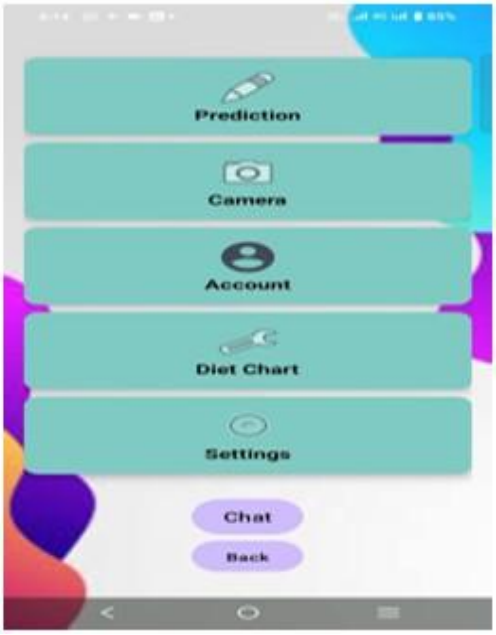


**Fig. 4.** Menu activity of mobile application


**Fig. 5.** Input data activity of application

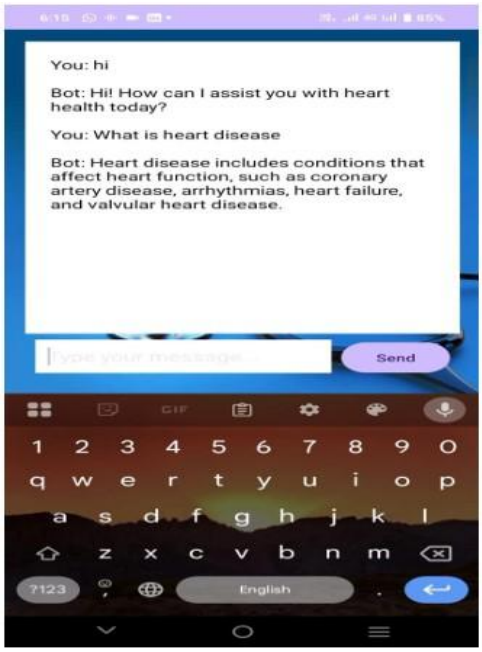


**Fig. 6.** Chabot activity of application

## 8. Conclusion and Future recommendations

This system "AI powered heart disease cardiovascular prediction using ML Algorithms" fills a crucial gap in the healthcare sector, especially in countries like Pakistan, where cardiovascular disease (CVD) rates are high and early detection resources are limited**.** The integration of AI and machine learning in heart disease prediction enhances early diagnosis and preventive healthcare, ultimately reducing the burden on the healthcare system. Considering the rapid growth in AI-driven healthcare solutions, this project has the potential to make a significant impact in Pakistan**,** addressing a critical need in preventive cardiology. The Android application of our project has different and unique features, such as Age, cholesterol, gender, weight, height etc. helps patients to filled data and identify the type of heart disease. The technological advancement facilitates early and accurate diagnosis, which is critical for effective management of disease. The application of our project is text-to-text function and intuitive interface ensures that the app is accessible. While the "Heart Disease Cardiovascular Prediction System" application has made significant contributions, there are several avenues for future work to enhance its functionality and impact:

**Enhancement of user Engagement Features**

**Gamification Features:** Introduce reward-based challenges for users to maintain a healthy lifestyle, such as step count goals, heart-friendly diet tracking, and fitness milestones.

**Personalized Health Insights:** Provide customized daily health tips, exercise plans, and diet recommendations based on the user's risk level.

AI-Powered Preventive Healthcare Systems

- Research AI-based solutions for preventive cardiology, including lifestyle coaching and automated risk assessment.
- Develop digital twin technology to simulate patient heart conditions for better treatment planning.

**Supplementary Materials:** Not Applicable.

**Author Contributions:** All authors equally contribute.

**Funding:** There is no funding for this project

**Institutional Review Board Statement:** Not Applicable.

**Informed Consent Statement:** Not Applicable.

**Data Availability Statement:** Data will be available on request

**Acknowledgments:** The authors gratefully acknowledge the support and facilities provided by the Department of Computer Science, University of Sialkot Pakistan, which made this research work possible.

Conflicts of Interest: The authors declare no conflicts of interest.

**References**

Balasubramaniam, S., Satheesh Kumar, K., Kavitha, V., Prasanth, A., & Sivakumar, T. A. (2022). Feature Selection and Dwarf Mongoose Optimization Enabled Deep Learning for Heart Disease Detection. *Computational Intelligence and Neuroscience*, *2022*, 1–11. https://doi.org/10.1155/2022/2819378

Baumgartner, H., Iung, B., Messika-Zeitoun, D., & Otto, C. M. (2022). The year in cardiovascular medicine 2021: valvular heart disease. *Cardiologia Hungarica*, *52*(1), 3–13. https://doi.org/10.15836/ccar2022.44

Chu, H., Chen, L., Yang, X., Qiu, X., Qiao, Z., Song, X., Zhao, E., Zhou, J., Zhang, W., Mehmood, A., Pan, H., & Yang, Y. (2021). Roles of Anxiety and Depression in Predicting Cardiovascular Disease Among Patients With Type 2 Diabetes Mellitus: A Machine Learning Approach. *Frontiers in Psychology*, *12*(April), 1–8. https://doi.org/10.3389/fpsyg.2021.645418

Faheem, A., & Faheem, U. (2023). Current Status of Nutritional Knowledge Regarding Cardiovascular Diseases (CVD). *Article in Biomedical Journal of Scientific & Technical Research*, *December*. https://doi.org/10.26717/BJSTR.2023.53.008473

Jindal, H., Agrawal, S., Khera, R., Jain, R., & Nagrath, P. (2021). Heart disease prediction using machine learning algorithms. *IOP Conference Series: Materials Science and Engineering*, *1022*(1). https://doi.org/10.1088/1757-899X/1022/1/012072

Nashif, S., Raihan, M. R., Islam, M. R., & Imam, M. H. (2018). Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System. *World Journal of Engineering and Technology*, *06*(04), 854–873. https://doi.org/10.4236/wjet.2018.64057

Sajja, T. K., & Kalluri, H. K. (2020). A deep learning method for prediction of cardiovascular disease using convolutional neural network. *Revue d'Intelligence Artificielle*, *34*(5), 601–606. https://doi.org/10.18280/ria.340510

Sun, X., Yin, Y., Yang, Q., & Huo, T. (2023). Artificial intelligence in cardiovascular diseases: diagnostic and therapeutic perspectives. *European Journal of Medical Research*, *28*(1), 1–11. https://doi.org/10.1186/s40001-023-01065-y

Telsang, S., Pokale, T., Posugade, A., Patwal, P., Shirode, P., & Popade, O. (2024). Heart Disease Prediction using Machine Learning. *Proceedings of NKCon 2024 - 3rd Edition of IEEE NKSS's Flagship International Conference: Digital Transformation: Unleashing the Power of Information*, *02*(March), 485–490. https://doi.org/10.1109/NKCon62728.2024.10774628

Vaduganathan, M., Mensah, G. A., Turco, J. V., Fuster, V., & Roth, G. A. (2022). The Global Burden of Cardiovascular Diseases and Risk: A Compass for Future Health. *Journal of the American College of Cardiology*, *80*(25), 2361–2371. https://doi.org/10.1016/j.jacc.2022.11.005

Wang, R., Garg, S., Gao, C., Kawashima, H., Ono, M., Hara, H., van Geuns, R. J., Morice, M. C., Davierwala, P. M., Kappetein, A. P., Holmes, D. R., Wijns, W., Tao, L., Onuma, Y., & Serruys, P. W. (2021). Impact of established

cardiovascular disease on 10-year death after coronary revascularization for complex coronary artery disease. *Clinical Research in Cardiology, 110*(10), 1680–1691. https://doi.org/10.1007/s00392-021-01922-y