

OPEN ACCESS

This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

¹Department of Electrical and Electronics Engineering Sri Eshwar College of Engineering, Coimbatore, Tamil Nadu, India

²Department of Mechatronics Engineering, Nehru Institute of Engineering and Technology, Coimbatore, Tamil Nadu, India

³Department of Electronics and Communication Engineering, V.S.B. College of Engineering Technical Campus, Coimbatore, Tamil Nadu, India

Correspondence to: Hariharan R, hariharan.r@sece.ac.in

Additional material is published online only. To view please visit the journal online.

Cite this as: Hariharan R, M Maheswaran, T Janani, Pushpalatha N, Akalya A, Deepthisre G and Malarvizhi M. Predictive Modeling for Early Detection of Aortic Disorder. Premier Journal of Science 2025;15:100211

DOI: <https://doi.org/10.70389/PJS.100211>

Peer Review

Received: 14 August 2025

Last revised: 31 October 2025

Accepted: 17 December 2025

Version accepted: 5

Published: 31 January 2026

Predictive Modeling for Early Detection of Aortic Disorder

R Hariharan¹, M Maheswaran², T Janani³, N Pushpalatha¹, A Akalya¹, G Deepthisre¹ and M Malarvizhi¹

ABSTRACT

This study details a novel predictive modeling framework to support the early identification of aortic disorders, by combining real-time physiological monitoring and machine learning techniques. The proposed framework utilizes multimodal clinical data and sensor derived measurements and predictive algorithms to allow precise and timely risk assessment for aortic pathologies. While given data analysis and monitoring of infants is more standard, this new framework bonces its focus and aims to help clinicians understand the cardiovascular health of their patients. Specifically, the goal would be aimed at helping to identify patients exposed to high risk for aortic abnormalities by employing non-invasive means. Finally, the study also reported that a combination of accuracy, sensitivity, specificity AUC scores and confidence intervals were reported to show the framework's robustness and the system's clinical reliability. Experimental validation illustrated the operationalization of synergistically using sensor data alongside advanced machine learning approaches to support early diagnosis of patients and potentially improve patient outcomes.

Keywords: Infant aortic disorder prediction, Sensor-fusion health monitoring, Electromyography-driven diagnostics, Logistic regression cardiac model, Non-invasive neonatal care system

Introduction

Cardiovascular diseases are the number one cause of death globally, and heart disease in particular poses a significant risk to public health. The role of early identification and diagnosis is essential for reducing morbidity and mortality through timely medical treatment. Conventional diagnostic techniques in heart disease often include invasive procedures or expensive imaging techniques, which can result in delays in treatment at the time of greatest need.

In this study, we propose a predictive modeling technique using machine learning to identify the risk of heart disease in patients from clinically relevant patient data. Specifically, the method uses the Cleveland heart disease database, which contains 13 physiological and clinical attributes that can be used by cardiologists, such as age, cholesterol level, type of chest pain and blood pressure. In the supervised learning method, particularly logistic regression, the model develops an ability to classify patients as high-risk and low-risk.

The contribution of this research is that it demonstrates that simple and interpretable machine learning methods can provide accurate and efficient diagnostic support for healthcare professionals. Predications by the model can be used as a non-invasive, low-cost and

reliable method to inform clinicians' decision-making and initiate preventive treatment early on. based monitoring to detect symptoms associated with early warning signs of an aortic abnormality. The intention is to provide reliable, real-time clinical insight to better inform clinicians and allow for a preventative mode of treatment, hence minimizing the burden of aortic related complications later in the patient pathway.

Literature Survey

The use of predictive modeling and smart monitoring systems to improve the early detection and diagnosis of high-stakes conditions (such as aortic disorders in infants) within the healthcare field has added value. Advances in machine learning algorithms and sensor fusion work together to create robust, non-invasive measures for early detection of disorders. A machine learning model has been created to evaluate the different algorithms to predict heart disease. The goal of the study is to provide an opportunity for doctors to recognize those at risk before the individual arrives for their consultation. As part of the study, Random Forest, Support Vector Machines (SVM), and logistic Regression were trained and tested incorporating clinical data. The methodology described provides timely medical intervention, while delivering diagnostic improvement.¹ In this study, cardiac disease prediction using machine learning algorithms from SVM, decision Trees and logistic Regression will be explored. Clinical characteristics, with patient information, were used for data training and testing. Performance was measured using variables in the (recall, accuracy and precision).² In an effort to protect patient data and health and maintain diagnostic quality, a safe A machine-learning system for monitoring adrenal diseases was published.³ A. Gadde and S. Chintala's document, "A Comprehensive Study on Heart Disease Prediction & Risk Stratification using Artificial Intelligence Explainable Technique," expands on (comprehensive report) the implementation of Explainable AI in heart disease risk activity prediction by discard combines machine learning models with SHAP techniques to communicate prediction for the heart disease prediction study.⁴ The supports vector machines (SVM) to identify heart failure in a timely manner is described. The study uses SVM to explore clinical data to diagnose individuals with the risk of heart failure. The study shows how well SVM can deal with complex datasets to support accurate predictions. The strategic plan aims to improve patient outcomes with heart failure therapy and early intervention.⁵ Further, the study uses machine learning to diagnose aortic and mitral stenosis using phonocardiogram signals, addressing an increase in

Ethical approval: N/a

Consent: N/a

Funding: No industry funding

Conflicts of interest: N/a

Author contribution:

R Hariharan, M Maheswaran, T Janani, N Pushpalatha, A Akalya, G Deepthisre and M Malarvizhi – Conceptualization, Writing – original draft, review and editing

Guarantor: R Hariharan

Provenance and peer-review: Unsolicited and externally peer-reviewed

Data availability statement: N/a

diagnostic accuracy and promote early detections, making advances towards non-invasive diagnoses for cardiac disorders.⁶ Heart Disease Prediction Using Logistic Regression, presents the use of logistic regression to predict heart disease and assesses its efficacy using accuracies, confusion matrix type metrics, focusing on early detection and medical interventions.⁷ The system employs machine learning techniques to evaluate physiological data and effectively predict possible strokes prior to their occurrence. The model uses algorithms such as Random Forest and Multilayer Perceptron to achieve high standard deviation, sensitivity and specificity, making it an effective system to provide timely medical treatment. This approach increases the potential for early diagnosis and prevention positively impacting patients.⁸ In all of the outlined studies,^{7–10} research was based on the predictions of heart disease. It looks at a patient's history, analyses the physiological data, and finds patterns in the data that are indicative of a patient's heart condition. The analysis would use different algorithms such as Support Vector Machines, Decision Trees, and Logistic Regression. This study seeks to enhance early detection and hope to guide physicians toward informed solutions. Using these machine learning models, the study will have a positive impact on patients' outcomes and diagnostic reliability in terms of cardiovascular health.⁹ A system that identifies cardiac anomalies early in patients by using machine learning approaches. The system will identify patterns in patient data that correlate to heart problems to allow for an arrangement of timely diagnosis. This strategy aims to enhance diagnostic accuracy and inform clinicians on optimal choices. With specifics strategies and an early intervention, this method intends to improve patient outcomes.¹⁰ The article "Prediction of Heart Disease Using Machine Learning Algorithms" begins by comparing how long various algorithms run to using pre-existing datasets such as the Cleveland heart disease dataset. They explained that logistic regression sits somewhere in the middle in terms of performance and interpretability, highlighting its use in application to what could be real-time decisions in a clinical setting.¹¹ The article "Application of Machine Learning in Early Detection of Heart Disease" suggested a hybrid model to clinical practitioners that included Logistic Regression with ensemble learning in order to improve the accuracy for diagnosis. By using correlation and information gain to maximize the features, the model improves accuracy of index in early-stage detection thereby reducing false positives for treatment in health care.¹² The study "Machine Learning Techniques for Heart Disease Prediction: A Comparative Study" is a literature article that compared several classifiers such as Logistic Regression, SVM, Random Forest, and XGBoost. Their results showed that while the more complex models yield slightly better accuracy, logistic regression was still found to be the model of choice in healthcare as it is much less complex conceptually and therefore easier to integrate into the clinical workflow as there is less black box transparency.¹³ The study "Intelligent

Heart Disease Prediction System using Data Mining Techniques" studied the use of data mining algorithms such as Naïve Bayes, Decision Trees, and Neural Networks to do early stage detection of heart disease. It allowed for the processing of large datasets of healthcare data and searched for hidden patterns to obtain a prediction and was able to do so with relatively high accuracy in the prediction of heart disease. In recent studies, the effects of imaging and biomechanics on the aortic disorder detection have been outlined. Ultrasound screening has been shown to reduce mortality associated with abdominal aortic aneurysms. Also, the use of wall stress by utilizing finite element modeling has been shown to better predict rupture compared to the diameter value. More recently, patient-specific biomechanical simulations have improved individualized risk assessment.

Deep learning have even gone further in cardiovascular imaging. Xu et al. used convolutional neural networks (CNNs) to analyze CT scan images for aortic aneurysm detection with a high accuracy, and Bratt et al. developed deep-learning-based segmentation process to improve assessment of aortic morphology. These studies support the idea that combining imaging, biomechanics and AI would facilitate early identification of aortic disorder, which is another aspect we can extend our study in the future Scientists, (2023).

Block Diagram

A system intended to track and identify changes in a number of physiological parameters. Signal analysis is used to process the data collected by input sensors, such as moisture, heart rate, EMG, and PIR sensors. The data that has been analyzed is next assessed to find any anomalies. To alert caretakers of any anomalies, an alert system is triggered. Potential uses for this technology include research, patient care, and healthcare monitoring, as shown in Figure 1.

A multi-sensor technique is used in the suggested system architecture, which is depicted in the block diagram (Figure 1), to monitor and identify physiological abnormalities. The system starts with input devices that continuously gather physiological data from the baby in real time, such as an ultrasonic sensor, moisture sensor, EMG sensor, and heart rate sensor.¹⁴

The system guarantees instant follow-up with the alarm and caregiver notification module in case an anomaly is detected. All reviewed data, regardless of anomaly detection, has been transmitted to the data storage for future use, research and longitudinal health analysis. The structure offers mostly non-invasive, effective, and real-time health monitoring of aortic and cardiac anomalies to enable early detection and improved care of babies.

Hardware Implementation

The baby health monitor system is constructed to provide uninterrupted health monitoring of a baby in real time by assimilating four basic sensors into the system's design which includes an EMG sensor, moisture sensor, heart rate sensor, and PIR sensor to

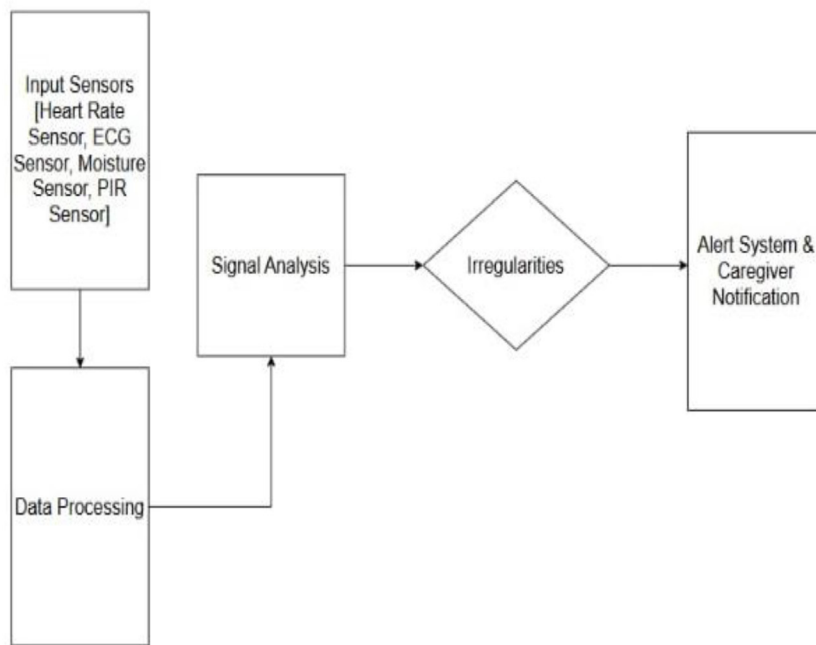


Fig 1 | Block diagram

recognize baby presence under a cover. The hardware's non-invasive design allows caregivers to focus on baby health monitoring to the health monitoring metrics while keeping baby comfortable at the same time.

Heart Rate Sensor

The By continually monitoring the infant's heart rate at any time, the heart rate sensor can help recognize early signs of heart diseases and the system ensures continuous and non-invasive monitoring process, which gives caregivers confidence. Additionally, recognizing the need for early detection through this sensor means that timely treatment can be provided, thereby improving the chances of better health outcomes for the infant.¹⁵

EMG Sensor

It identifies electrical activity of muscles with a focus on signals related to heart function. It provides important information about cardiac health by monitoring electrical impulses and muscle contractions. In addition, the sensor furthers heart rate monitoring by providing a greater understanding of the baby's cardiac muscle activity.

Moisture Sensor

It keeps the infant clean and comfortable by warning caretakers when the diaper needs to be changed. Through the prevention of extended exposure to moisture, the sensor lowers the risk of skin infections and diaper rash. This guarantees prompt attention and improves the baby's general health.

PIR Sensor

Non-invasive Transcutaneous bilirubinometers are non-invasive sensors used to detect infant jaundice. These devices utilize specific wavelengths of light to

measure bilirubin levels within a baby's skin, providing a painless and safe opportunity for early diagnosis. Jaundice is a prevalent disease of neonates in which increased blood bilirubin levels result in a yellowish hue of the skin. Non-invasive imaging sensors enable caregivers to monitor bilirubin levels continuously and intervenes medically in a timely manner to support the infant's health and wellbeing.

Software

The heart disease prediction system was implemented using Python for its flexibility and extensive data science libraries. Google Colaboratory was used for development, allowing for real-time code execution and teamwork. Data preprocessing was done by Pandas, while numerical calculations were provided by NumPy. The machine learning model was constructed, trained, and assessed using Scikit-learn. The sigmoid function is used to convert inputs to probabilities in the main technique, Logistic Regression, which is perfect for binary classification. It used characteristics including age, blood pressure, cholesterol, and the type of chest discomfort to forecast the risk of heart disease.

Python

The planned system was developed using the versatile programming language Python, which has a high level of data science libraries. Google Collaboratory which was cloud-based Jupyter notebook environment that allowed for real-time code execution was used as the development platform. Pandas library was used to work with data manipulation and preprocessing. The NumPy library was used for numerical calculations with the added benefits of array reshaping during predictions. All machine learning operations were done using the Scikit-learn library, which worked best to allow model building, training, and evaluations of models. All of these tools helped to increase development time with high degrees of repeatability by providing a flexible, and interactive environment for developing and executing predictive models.¹⁶

Logistic Regression

A structure a supervised machine learning technique developed specifically for problems of binary classification, where the goal is to classify an input instance into one of two classes. On the contrary to linear regression, which predicts continuous values, logistic regression uses the sigmoid (logistic) function to convert expected values into a probability between 0 and 1. A classification threshold - typically 0.5 - is used to convert probability into a binary output. In this study, logistic regression was used to determine a patient's probability of having heart disease, based on medical attributes, including age, resting blood pressure, cholesterol, and type of chest pain.

Flowchart

The flowchart represents the operational flow of the "Full Pampering of Infant Babies" system. It begins with data collection from sensors such as heart rate,

EMG, moisture, PIR, and temperature sensors. The collected data is sent to the ESP32 microcontroller, where it undergoes filtering and normalization to remove noise and ensure consistency.

The processed data is analyzed for health and hygiene parameters, including heart health, muscle activity, diaper wetness, and presence detector as shown in this Figure 2. If irregularities are detected, real-time alerts are sent to caregivers. Otherwise, the data is stored for long-term monitoring and trend analysis.¹⁷

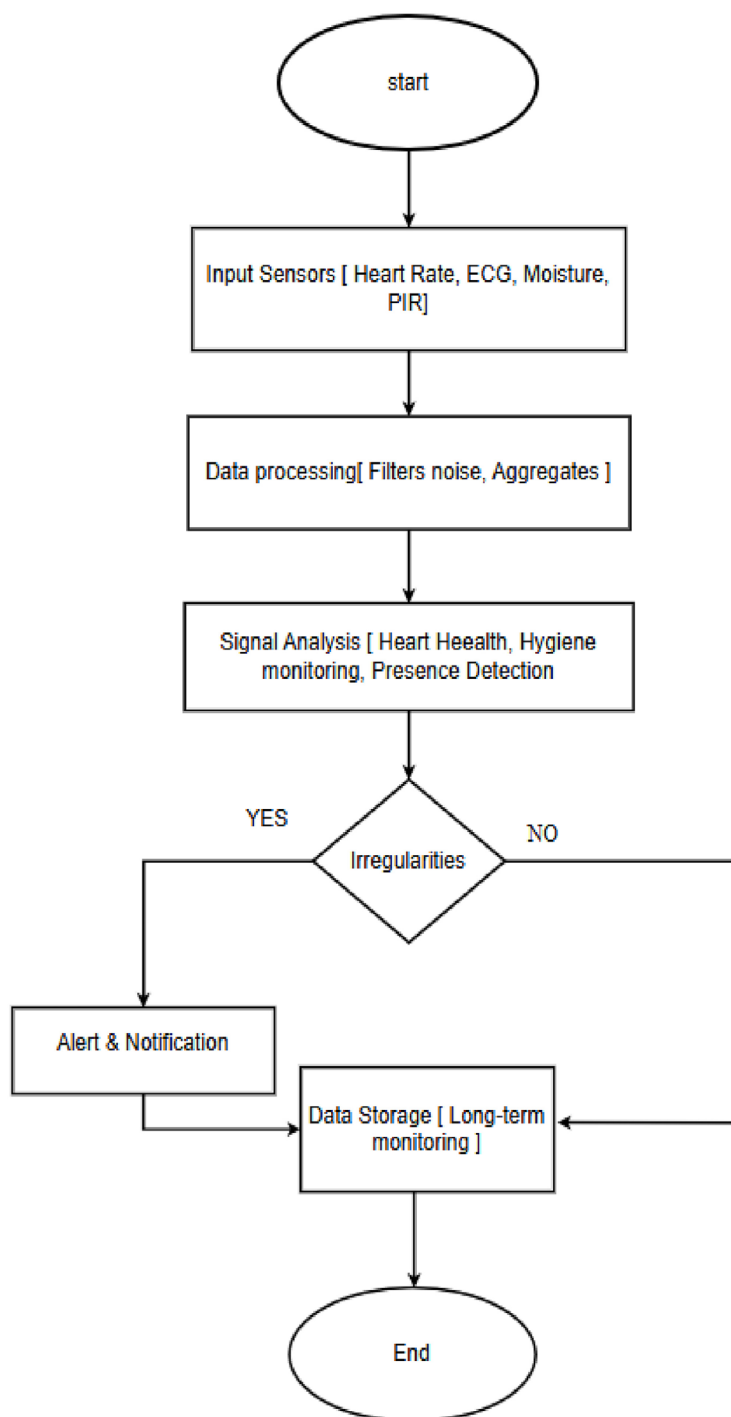


Fig 2 | Flowchart

The system ensures continuous infant care by providing actionable insights and notifications.

Circuit Diagram

An electrical circuit is represented graphically by a circuit diagram. Simple pictures of the circuit's parts are used in pictorial circuit diagrams, whereas standard symbolic representations are used in schematic diagrams to depict the circuit's parts and connections. The schematic diagram's depiction of the connections between circuit components does not always match the physical configurations in the completed product.

The heart rate sensor uses an analog output to send signals to the ESP32, which can be connected to a 3.3V or 5V supply, ground, and analog input pins like GPIO 34. The PIR sensor detects motion by outputting a HIGH signal when motion is detected, which the ESP32 can interpret as a trigger event.

The moisture sensor measures soil water content with two pins for power and one for analog output as shown in the Figure 3. The ESP32 reads moisture levels from the sensor and interprets the data. The temperature sensor measures environmental temperature with three pins: VCC, GND, and data output. The DHT11 or DHT22 sensor measures environmental temperature with three pins: VCC, GND, and data output.

A library is needed to interface with the DHT sensor and read temperature and humidity values. The ECG sensor monitors the electrical activity of the heart with multiple leads: one for the signal, one for reference, and one for ground. The ground lead connects to the GND pin of the ESP32, the reference lead to a digital pin, and the signal lead to an analog input pin. The ECG sensor outputs an analog signal that the ESP32 can read to monitor ECG waveforms.

Proposed System Architecture

The suggested system for heart disease prediction is a machine learning framework that encompasses data acquisition, data preprocessing, model training and evaluation. The architecture aims to deliver accurate predictions while maintaining clinical interpretability. Figure 1 shows the workflow of the system.

1. Data Acquisition

Patient data were acquired from the Cleveland Heart Disease dataset, which has 303 records and 13 clinical attributes (age, sex, type of chest pain, blood pressure, cholesterol, fasting blood sugar, resting ECG results, and exercise mindfully angina). The outcomes were binary, indicating the presence (1) or absence (0) of heart disease.

2. Data Preprocessing

The preprocessing included handling of the categorical features, normalizing the continuous features, and splitting the dataset. To mitigate class imbalance, we used stratified k-fold cross-validation (ensuring that both positive and negative cases were represented proportionately in each fold).

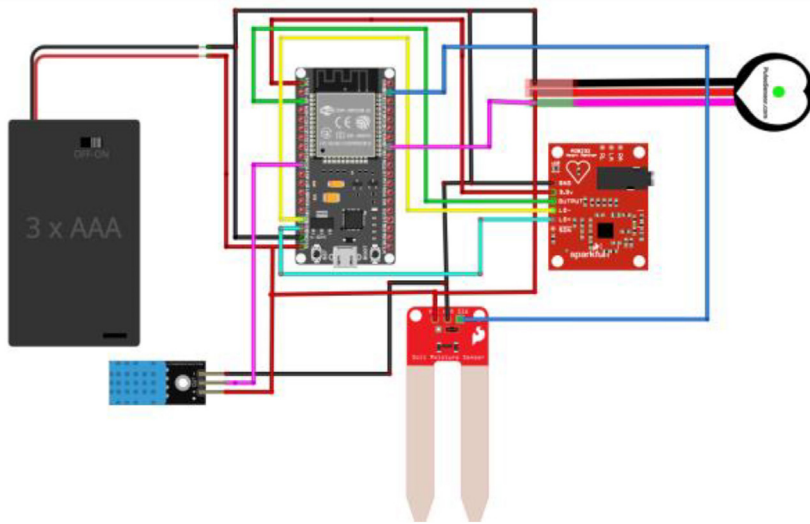


Fig 3 | Circuit diagram

3. Model Training

Logistic Regression was selected as the baseline classifier because of its interpretability and its capacity to classify in binary. The model was trained on 80% of the dataset, with validation completed on 20%. Cross-validation was conducted to evaluate robustness and minimize overfitting.

4. Evaluation Metrics

Model performance was assessed using standard metrics including Accuracy, Precision, Recall, F1-score, Area Under the Receiver Operating Characteristic Curve (AUC), and 95% Confidence Intervals (CI). These measures provide a comprehensive view of the model's predictive ability and clinical relevance.

5. Prediction

Once trained, the system can classify new patient records by analyzing the 13 clinical inputs and predicting the probability of heart disease. The outcome can assist clinicians in early risk identification and decision-making, enabling timely medical intervention.

Proposed Methodology

Dataset Description and Preprocessing

The main goal of the heart disease dataset is to forecast a person's likelihood of developing heart disease by analyzing extensive medical data gathered from a variety of patients. The entire medical profile of a single patient is represented by each data entry in the dataset, which is called an instance. Every instance has these characteristics in addition to a target variable, which is a binary value that denotes the existence or absence of cardiac disease accordingly and can be either `0` or `1`. These characteristics, which can be categorical or numerical, are carefully selected according to their importance in medical diagnosis.

These characteristics, which can be categorical or numerical, are carefully selected according to their

Table 1 | Dataset characteristics and description

Dataset Characteristics	Description
Number of Instances	303
Number of Features	13
Target Variables	Presence of Heart Disease (0 or 1)
Feature Type	All continuous or categorical
Missing Values	None

importance in medical diagnosis. Since there are no missing values in this dataset, it is ideal for training classification models as shown in Table 1. To properly construct and assess machine learning models, the data was pre-processed by dividing it into training and testing subsets.¹⁸

As shown in the Table 1, the dataset includes 13 medically relevant features used by clinicians to assess cardiovascular health, including age, sex, chest pain type, resting blood pressure, cholesterol levels, fasting blood sugar, electrocardiographic results, thalach, exercise- included angina, oldspeak slope.ca (number of major vessels coloured by fluoroscopy, ranging from 0 to 3), and thal (a categorical variable representing thalassemia type: 3 = normal, 6 = fixed defect, 7 = reversible defect). These features provide insights into a patient's heart condition and form the basis for predicting heart disease likelihood using machine learning models.

The dataset was examined for class balance and null values prior to training. There were no missing values discovered. There was a small imbalance in the distribution of classes, with heart disease being present in most cases.

The dataset was split into 80% training and 20% testing sets using stratified sampling to preserve class distribution after the characteristics were removed from the target in order to prepare it for model training. Although it was not required for Logistic Regression, data normalization (such as Z-score normalization) might be used optionally to enhance model convergence.

Problem Formulation

The objective is to predict whether a person has heart disease based on their physiological and clinical data. This is a binary classification problem, where each instance is labelled as:

0: No heart disease (healthy)

1: Presence of heart disease

Let the feature space be represented as $F = \{f_1, f_2, \dots, f_n\}$, where $n = 13$ features. The target is to learn a function $h: F \rightarrow \{0, 1\}$ that accurately classifies patient records. This can be framed as a supervised learning task using Logistic Regression.

Feature and Target Separation

The dataset was divided into input features and the target variable. All columns except the 'target' column were considered as features (denoted as X), and the 'target' column was used as the output label (denoted

as Y). The target variable is binary, where 0 represents a healthy heart and 1 indicates the presence of heart disease.

Train- Test Split

To correctly evaluate model performance, the dataset was divided into training set and test set in an 80:20 split. This involved additionally using stratified sampling during splitting so that both sets of samples had an equal distribution of the target classes to ensure the model received a balanced set of samples during training and testing. In stratified sampling, during splitting, the proportions of the target classes were maintained during splitting of the dataset. This helps prevent data imbalance during the training of the model and assigns an even measurement of performance on the testing set.

Model Training

A logistic regression model was selected and trained using the training dataset. Logistic regression was chosen due to its simplicity, interpretability, and effectiveness for binary classification tasks. The model was trained using Scikit-learn's Logistic Regression class without extensive hyperparameter tuning, as the goal was to establish a reliable baseline.

Model Evaluation

The trained model was evaluated using accuracy as the primary metric. Predictions were made on both the training and test datasets, and the accuracy scores were calculated using Scikit-learn's accuracy score function. This step helped assess the model's performance and generalization ability on unseen data.

Dataset Overview

In this study, the Cleveland Heart Disease dataset was used which can be found in the public domain. It consists of 303 records, each with 13 clinical features including: age, cholesterol, blood pressure, chest pain type, fasting blood sugar, resting ECG results, and exercise-induced angina. Each record also contains a binary target variable outlining the presence (1) or absence (0) of heart disease.

Dataset Preparation

The dataset was assessed for missing values and class imbalance. There were no null values, borderline imbalance between diseased and healthy cases was rectified by leveraging stratified sampling. Continuous (ratio & interval) variables were rescaled using Z-score normalization so that the learning algorithm would converge more effectively. The dataset was then split into training and testing subsets using an 80:20 split.

Model Development

Logistic Regression, due its simplicity and interpretability, was selected as the baseline classifier. It has also been widely used in clinical decision support. To robustly evaluate the models, stratified 5-fold cross-validation was conducted where the dataset was

Table 2 | Feature comparison chart

Parameter	Existing Method	Proposed Method
Size	It is large and difficult to move	Compact and portable, easy to use in this proposed method
Cost	High initial and maintenance costs	Cost-effective solution with low maintenance
Usage	Limited real-time monitoring capabilities	Real-time, continuous, and accurate monitoring
Efficiency	Requires skilled professionals to operate	User-friendly interface with minimal training needed
Advantages	Basic vital monitoring with fewer features	Enhanced monitoring, data analytics, and alerts for quick intervention

divided into five folds, then using four folds for training and one for validation in each iteration. This mitigated variance in performance estimation and overfitting.¹⁹

Comparison

Table 2 shows the comparison of existing and proposed methodology with respect to various parameters like size, cost, usage, efficiency and advantage.

Result Analysis

The proposed system is analyzed based on experimental setup and simulation analysis using Google Co lab.

Experimental Setup

Using a publicly accessible dataset obtained from Kaggle, extensive tests were carried out to assess the effectiveness of logistic regression for heart disease prediction. The dataset includes a binary target variable that indicates if cardiac disease is present (1) or not (0), in addition to 13 clinical characteristics. Python was used for the implementation, and key libraries for data processing, model training, and evaluation included Pandas, NumPy, and SciKit Learn.

Algorithm: Logistic Regression

Input:

Heart disease dataset with n features and target variable

Test size split ratio (e.g., 0.2)

Random state seed (for reproducibility)

Output:

Predicted class label: 1-Diseased, 0-Healthy

Begin:

1. Load Dataset

Read CSV file into a DataFrame D.

2. Preprocess Data

- Check for null values.

- Perform statistical analysis on D.

- Separate features X and target Y.

3. Split Dataset

Use "train_test_split" to divide X and Y into

- Training set: (X_{train}, Y_{train})

- Testing set: (X_{test}, Y_{test})

4. Initialize Logistic Regression Model

Model—Logistic Regression

5. Train Model

Fit the model on $X_{\text{train}}, Y_{\text{train}}$

6. Evaluate Model

- Predict on training data: $Y_{\text{train}} - \text{model}$.Predict(X_{train})- Predict on test data: $Y_{\text{test}} - \text{model}$. Predict(X_{test})

- Compute accuracies using "accuracy_score"

7. Build Predictive System

- Input new patient features as array

- Reshape x

- Predict outcome using: $Y - \text{model}$.Predict(x_{reshaped})

8. Return

Predicted output and accuracy scores on train and test data.

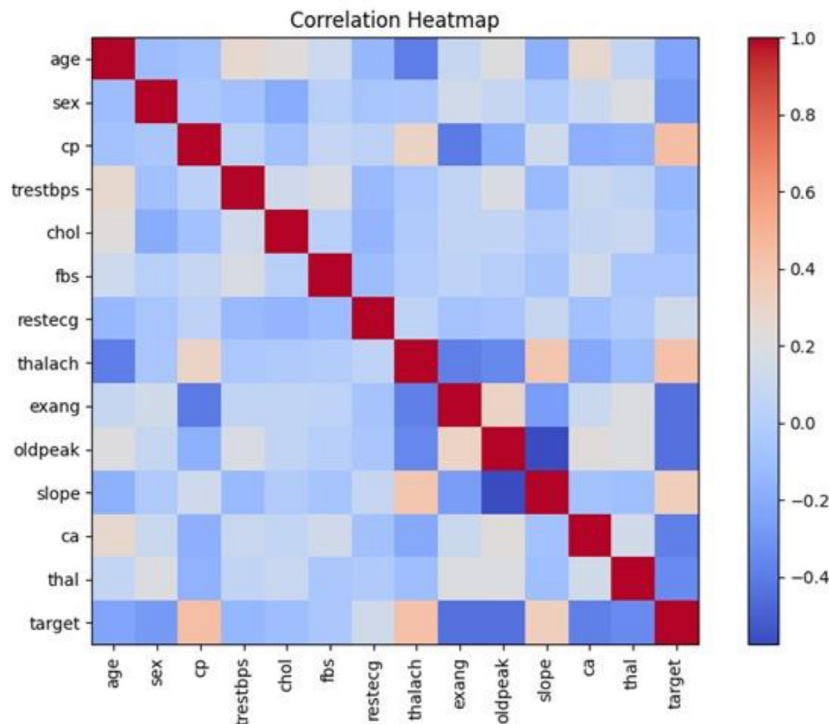


Fig 4 | Correlation analysis

Table 3 | Sample prediction results

Age	Sex (M-1, F=0)	CP	Cholesterol	Heart Rate	Prediction	Heart Disease
52	1	0	212	168	0	No
58	0	2	250	155	1	Yes
45	1	1	230	160	0	No
34	0	0	150	160	0	No
29	0	1	220	150	0	No

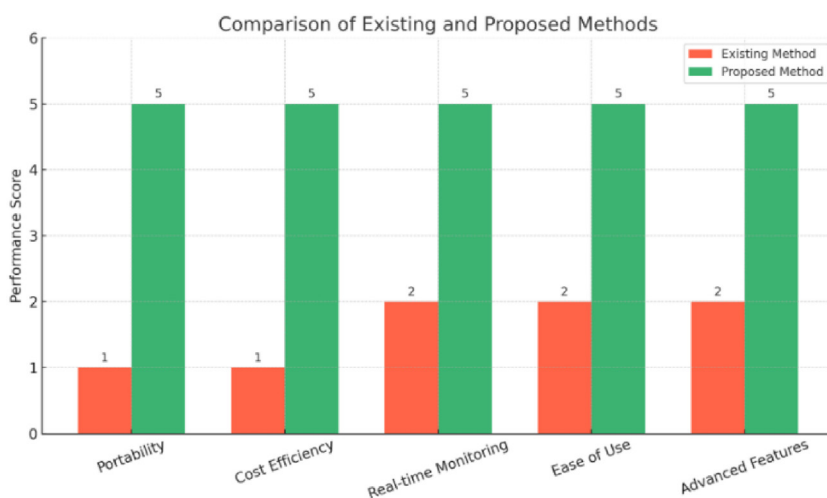


Fig 5 | Comparison chart

Performance Evaluation

The accuracy score, which determines the percentage of properly predicted instances, was used to evaluate the model's performance once it had been trained. Both the test set (to assess generalization) and the training set (to check for underfitting) had predictions produced. These measures aid in assessing the model's performance and reliability for making predictions in the future.²⁰

Real-time Prediction Simulation

A set of 13 values representing a new patient's health data as structured into a NumPy array, reshaped into a manner suited for prediction, and passed into the trained model. The system printed whether the user is likely to get heart disease based on the output (either 0 or 1).

Correlation Heatmap Analysis

The correlation heatmap shown in Figure 4 visually represents how each feature in the heart disease dataset is related to the others. Darker red indicates a strong positive correlation, while darker blue indicates a strong negative correlation. For instance, chest pain type (cp) and maximum heart rate achieved (thalach) are positively correlated with the presence of heart disease, suggesting they are important indicators. In contrast, exercise-induced angina (exang), ST depression (oldpeak), and the number of major vessels colored (ca) show a negative correlation with heart disease. This visualization helps in selecting the most relevant features for building an effective machine learning model by highlighting which variables are most strongly associated with the target outcome.

Sample Prediction Results

Table 3 shows the sample prediction data of the proposed system.

Real-time test cases used to evaluate the model's prediction accuracy. In total, five real-time test cases were evaluated. These cases were manually selected to represent a range of patient profiles varying in age, cholesterol levels, chest pain types, and heart rate values as shown in Table 3. Each entry represents a patient's health parameters passed into the trained system. The predicted outcomes were compared with actual diagnoses.

Comparison of Heart Disease

This graph shows example patient predictions for the identification of heart disease using logistic regression. Important health metrics were employed as input

Table 4 | Accuracy (%) of different classification models

Model	Accuracy (%)
Proposed Model	92
SVM	88
Random Forest	90
Logistic Reg	85

characteristics, including heart rate, cholesterol, age, and sex. Either 0 (no disease) or 1 (disease) is the model's output, and the results are displayed alongside the real data. It illustrates how well the model uses clinical data to determine the risk of heart disease as shown in Figure 5. In healthcare systems, these forecasts facilitate prompt medical action and early diagnosis (Table 4).

Conclusion

The study demonstrates that aortic pathology risk can be accurately and reliably predicted using a modeling strategy incorporating logistic regression. Evaluation metrics provided accuracy, recall, F1 score and AUC performance indicators for the predictions made, demonstrating a reliable model, rather than one producing random outcomes.

The value of the model rises above other predictive models in that they were not only interpretable for clinical decision making, but also provided evidence for predicted probabilities and refined clinical practice. Future work will focus on larger datasets, and deep and advanced learning techniques. Future work will also focus on healthcare systems providing discrete datasets on aortic pathologies, in order to demonstrate to clinicians, a greater generalizability to their practice.

References

- Rabbi MA, Rijon RH, Akhi SS, Hossain A, Jeba SM. A detailed analysis of machine learning algorithm performance in heart disease prediction. In: 2025 4th International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST); Dhaka, Bangladesh. 2025. p. 259–63. <https://doi.org/10.1109/ICREST63960.2025.10914417>
- Balvir P, Katore K, Gourshettiwar P. Prediction of heart disease diagnosis using machine learning algorithms. In: 2025 International Conference on Electronics and Renewable Systems (ICEARS); Tuticorin, India. 2025. p. 2016–21. <https://doi.org/10.1109/ICEARS64219.2025.10941062>
- Poonam, Mallapur SV. Secure healthcare monitoring: algorithms and techniques for heart disease detection and data privacy. In: 2025 International Conference on Intelligent Systems and Computational Networks (ICISCN); Bidar, India. 2025. p. 1–5. <https://doi.org/10.1109/ICISCN64258.2025.10934556>
- Gadde A, Chintala S. Comprehensive study on heart disease prediction and risk stratification using explainable artificial intelligence technique. In: 2025 International Conference on Intelligent Systems and Computational Networks (ICISCN); Bidar, India. 2025. p. 1–5. <https://doi.org/10.1109/ICISCN64258.2025.10934680>
- Prajapati YN, Patel UK, Srivastava M, Yadav J, Srivastava MK, Gupta BK. Protecting hearts with support vector machine analysis for the early detection of heart failure. In: 2025 2nd International Conference on Computational Intelligence, Communication Technology and Networking (CICTN); Ghaziabad, India. 2025. p. 117–9. <https://doi.org/10.1109/CICTN64563.2025.10932449>
- Behera S, Misra IS, Siddiqui KN. Phonocardiogram signal based better prediction of aortic and mitral stenosis heart diseases using machine learning. In: 2025 International Conference on Ambient Intelligence in Health Care (ICAHC); Raipur, India. 2025. p. 1–6. <https://doi.org/10.1109/ICAHC64101.2025.10957655>
- Saha D, et al. Heart disease prediction using logistic regression. In: 2025 International Conference on Computer, Electrical & Communication Engineering (ICCECE); Kolkata, India. 2025. p. 1–6. <https://doi.org/10.1109/ICCECE61355.2025.10940073>
- M M, V KD, Gunapriya D, N P, Karthik SS, A S. Machine learning-based pre-stroke detection system. In: 2024 International Conference on Science Technology Engineering and Management (ICSTEM); Coimbatore, India. 2024. p. 1–5. <https://doi.org/10.1109/ICSTEM61137.2024.10560875>
- Panda AK, Pati C, Pradhan S, Pradhan A, Rath NK. Prediction of heart disease using ML algorithms. In: 2025 First International Conference on Advances in Computer Science, Electrical, Electronics, and Communication Technologies (CE2CT); Bhimtal, India. 2025. p. 930–4. <https://doi.org/10.1109/CE2CT64011.2025.10939603>
- Sudhakar G, Perumalla S, Saturi R, Vulapula SR, Raju Y, Saturi S. Early detection and diagnosis of cardiac disorders using machine learning techniques. In: 2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI); Goathgaun, Nepal. 2025. p. 1312–6. <https://doi.org/10.1109/ICMCSI64620.2025.10883620>
- N P, D K S, S O S, K T, Arulvadiju J, M G. Detection of mishap and myocardial infarction. In: 2024 International Conference on Recent Innovation in Smart and Sustainable Technology (ICRISST); Bengaluru, India. 2024. p. 1–5. <https://doi.org/10.1109/ICRISST59181.2024.10921953>
- Babu WR, et al. Space-based GPS solar tracking system. In: Advances in Engineering Research. 2024. p. 371–82. https://doi.org/10.2991/978-94-6463-529-4_3
- Dhananjayan KS, M M, Kayalvizhi P, Lakshmanan P, N P, Senthoriya OS. Development of a telematic control unit for capturing vital vehicle data without using company fitted telematic ports. In: 2024 International Conference on Science Technology Engineering and Management (ICSTEM); Coimbatore, India. 2024. p. 1–5. <https://doi.org/10.1109/ICSTEM61137.2024.10561194>
- Prasanth A, D L, Dhanaraj RK, Balusamy B, P C S, editors. Cognitive computing for Internet of Medical Things. 1st ed. Chapman and Hall/CRC; 2022. <https://doi.org/10.1201/9781003256243>
- Reddy S, Shetty KK, G S A. Machine learning models for advancing heart disease prediction and diagnosis. In: 2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT); Bengaluru, India. 2025. p. 2045–51. <https://doi.org/10.1109/IDCIoT64235.2025.10915036>
- Zucker EJ, et al. Abdominal aortic aneurysm screening: concepts and controversies. J Clin Imaging Sci. 2018;8:29. https://doi.org/10.4103/jcis.JCIS_40_18
- Benson RA, et al. Ultrasound screening for abdominal aortic aneurysm. J Clin Ultrasound. 2018;46(6):358–64. <https://doi.org/10.1002/jcu.22845>
- Leach JR, et al. Abdominal aortic aneurysm measurement at CT/MRI: a comparison of techniques and implications for clinical practice. J Vasc Interv Radiol. 2021;32(7):1045–52. <https://doi.org/10.1016/j.jvir.2021.03.008>
- Leach JR, et al. Impact of implicit abdominal aortic aneurysm screening in a large healthcare system. J Am Heart Assoc. 2022;11(3):e024571. <https://doi.org/10.1161/JAHA.121.024571>
- Lu JT, et al. DeepAAA: clinically applicable and generalizable detection of abdominal aortic aneurysm using deep learning. Nat Commun. 2019;10:4716. <https://doi.org/10.1038/s41467-019-12715-9>