Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 11 Issue 44 (September 2026) PP. 975-991




JOURNAL OF INFORMATION SYSTEM AND TECHNOLOGY MANAGEMENT
(JISTM)
www.gaexcellence.com/jistm

# CARDIORISK_ME: A MACHINE LEARNING-BASED PREDICTIVE ANALYTICS FOR HEART DISEASE

Daniyal Rosli[1], Zanariah Idrus[2]*, Ahmad Afif Ahmarofi[3], Nurul Husna Mahadzir[4]

[1]Advanced Micro Devices (AMD), Penang, Malaysia
dbinrosl@amd.com https://orcid.org/0009-0004-9016-8030
[2]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Kedah, Malaysia
zanaidrus@uitm.edu.my https://orcid.org/0009-0007-3544-2051
[3]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Kedah, Malaysia
ahmadafif@uitm.edu.my https://orcid.org/0000-0001-9270-0597
[4]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Kedah, Malaysia
husnamahadzir@uitm.edu.my https://orcid.org/0000-0002-8607-5438
*Corresponding Author

**Article Info:**

**Article history:**
Received date: 26.07.2026
Revised date: 20.08.2026
Accepted date: 24.09.2026
Published date: 30.09.2026

**To cite this document:**

Rosli, D., Idrus, Z., Ahmarofi, A. A., & Mahadzir, N. H. (2026). CardioRisk_Me: A Machine Learning-Based Predictive Analytics for Heart Disease. *Journal of Information System and Technology Management, 11* (44), 975-991.

**Abstract:**

Heart disease remains one of the leading causes of death worldwide. Therefore, timely detection is essential to help improve patients' health. While machine learning and artificial intelligence have advanced significantly, implementation and development challenges remain. This study addresses these issues by providing a machine learning-based heart disease prediction model and presenting the results through an accessible, user-friendly web interface. In addition, this research analyzes and assesses clinical and physiological parameters relevant to heart disease risk prediction and develops a prediction system using Logistic Regression, Decision Tree, and Random Forest classifiers. This work supports efforts to increase awareness and control of heart disease, which continues to affect public health globally. In this study, a cardiac risk prediction system (CardioRisk_Me) was developed using a machine learning approach based on clinical and physiological attributes. The dataset is represented by 304 patient records from open-access repositories. The pre-processed data includes scaled features, encoded variables, cleaned extractions, and a training-testing split. Three machine learning algorithms were tested: Logistic Regression, Decision Tree, and Random Forest, using test-based measures of Accuracy, Precision, Recall, F1-Score, and ROC-AUC. Logistic Regression achieved the highest accuracy (86.27%) and the closest to perfect precision, Recall, F1-score, and ROC-AUC (90.00%, 87.10%, 88.52%, and 89.19%, respectively). The chosen model was incorporated into a web-based prototype built with Next.js and Django, creating a real time heart disease risk prediction system. Overall, the

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 11 Issue 44 (September 2026) PP. 975-991

results illustrate the promise of machine learning in supporting risk assessment and preventive healthcare.

**DOI:** 10.35631/JISTM.1144057 **Keywords:**

Prediction, Machine Learning, Artificial Intelligence, Heart Care, Cardiovascular

© The authors (2026). This is an Open Access article distributed under the terms of the Creative Commons Attribution (CC BY NC) (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact jistm@gaexcllence.com.

## Introduction

Heart disease is a huge health crisis and a big contributor to death rates around the world. It remains a major global health issue, responsible for around 31% of all deaths worldwide (Vaduganathan et al., 2022). Although medicines and medical technologies have improved and treatments are available, the burden of heart disease is still increasing, especially in low-and middle-income countries (Qureshi et al., 2021). Early detection and good management of cardiovascular disease are key to optimizing patient care (Ali et al., 2020). Thus, early detection and identification of at-risk individuals can improve treatment effectiveness and help control healthcare costs (Ahmarofi, Abidin, & Zabidi, 2022).

Traditional risk assessment systems, such as the Framingham Risk Score, have been widely extended to estimate an individual's risk of cardiovascular disease (CVD) (Aliasgharzadeh et al., 2026). Yet many of these tools rely on only a handful of risk factors and do not fully reflect the intricate interplay of physiological, genetic, and lifestyle factors that affect the development of heart disease (Law et al., 2023).

Recent developments in machine learning (ML) and artificial intelligence (AI) have enabled more accurate, personalized prediction models and more comprehensive risk assessments (Udegbe, 2023; Ahmarofi et al., 2022). Several studies have shown the potential of ML algorithms to predict heart disease risk. Using EHR data, Li (2024) developed a ML model that outperforms traditional clinical risk scores in forecasting cardiovascular events.

However, many existing models are not practical for risk assessment, and the lack of easy-to-use prediction systems remains a challenge in heart disease prediction. A system is needed that provides ongoing risk assessment and early warnings for potential cardiac events, including data from wearables and remote monitoring systems (Hughes et al., 2023). It can improve

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 11 Issue 44 (September 2026) PP. 975-991
access and decision-making by embedding the most successful model in a web-based platform that medical professionals can access.

## Literature Review

Heart disease has become a leading global health concern, with coronary heart disease accounting for 16% of deaths worldwide since 2019 (Zhang et al., 2023). Mortality rates are particularly rising in high-income regions, reflecting an urgent need for more effective diagnostic and preventive measures. Therefore, the increase in heart disease and related deaths highlights the need for innovative and accessible solutions to reduce its impact and improve patient outcomes (Husnain et al., 2026).

Coronary heart disease and heart ailments are a leading cause of death for many people today. It is rated the number one cause of death across the globe, with the count tallying up to about 17.9 million deaths per year, accounting for 32 % of all deaths globally (Gaziano, 2022). Major risk factors include hypertension, hypercholesterolemia, smoking, obesity, physical inactivity, and diabetes (Bevan et al., 2023), with age, sex, and genetic diversity being additional influences. Early detection is critical, as symptoms usually appear only in advanced stages, so predictive tools are needed to identify at-risk individuals before symptoms develop (Ullah et al., 2023).

Predictive analytics in healthcare combines historical and real-time data with ML and statistical techniques to anticipate consequences. ML techniques can analyze large sets of health records to predict the likelihood that people will develop diseases. It affects health outcomes, from disease prediction to patient readmissions, to improve decision-making, patient care, cost efficiency, and operational effectiveness. Predictive analytics is being integrated into the control of noncommunicable diseases such as diabetes, cancer, and heart disease by combining medical history, genetic predisposition, and lifestyle factors to assess risk. Despite notable technical challenges, including data privacy, integration into clinical workflows, and model interpretability (Mackay, 2023), predictive analytics is expected to grow in popularity in healthcare.

Predictive models support patient-specific preventive and curative measures by considering factors such as age, cholesterol level, blood pressure, and lifestyle. Logistic Regression, Decision Trees, Random Forests, and Neural Networks are key algorithms that recognize patterns and predict the likelihood of heart disease, supporting proactive management of high-risk patients (Yang et al., 2023). Logistic Regression is widely used in clinical conditions (Pakdil et al., 2026). Logistic Regression and Framingham Risk Score provide a baseline, though they are limited by linear assumptions. The Framingham Risk Score is a risk assessment tool that lets physicians estimate a patient's approximate 10-year risk of developing cardiovascular disease, such as a heart attack or coronary heart disease (CHD) (Liu, 2023; Gray et al., 2023).

Decision Trees, Random Forests, and Neural Neetworks often achieve the highest accuracy because they handle nonlinear relationships (Ahmarofi, 2022); however, they can overfit, are computationally complex, and offer limited interpretability (Karna, 2024; Choudhury et al., 2024). Each algorithm produces different results depending on dataset size, complexity, feature engineering, and the balance between interpretability and performance. Most studies use terms such as accuracy, precision, recall, and F1-score to assess and enhance predictive models

Journal of Information System and Technology Management
EISSN: 0128-1666
Volume 11 Issue 44 (September 2026) PP. 975-991
(Albahri et al., 2023). Basic model structures benefit from cleaning, normalization, and feature selection, which greatly boost performance (Pan et al., 2023). Examples include the Framingham Heart Study, which includes important patient information for building and testing predictive models (Thomas et al., 2024).

One of the primary challenges in managing heart disease is the complexity and cost of multiple tests, such as chest X-rays, angiography, and electrocardiograms (ECGs), which can be time-consuming and financially burdensome (Netala et al., 2024). This complexity can delay risk prediction, increase the risk of more serious health issues, and limit access to early intervention, especially in under-resourced areas. This has made easier, more affordable diagnostic tools a top priority. Current solutions to improve heart disease detection through prediction systems have lack of efficiency. The current models need updating to be readily applicable for widespread use, or they must be accurate enough for risk assessment (Ghaffar et al., 2023).

This study aims to improve early detection and risk assessment in clinical settings. The project matters because it can reduce falls among older adults and improve patient care through Iona-based, tailored preventive interventions that improve individual outcomes (Dogheim et al., 2023). In addition, ML methods such as Logistic Regression, Decision Trees, Random Forests, and Neural Networks can provide insight into risk factors and complex interactions that traditional statistical methods may overlook. By incorporating a web app, clinicians can address a key health challenge, better data visualization, and improve clinical decisions and patient communication.

## Methodology

This research project aimed to create an early-detection predictive system using predictive analytics to forecast heart disease. The methodology began with a broad literature review, data collection, and identification of heart disease-related risk factors, including age, blood pressure, and cholesterol. Numerous data pre-processing steps such as handling missing data, encoding categorical data, normalizing features, and removing outliers were performed. The patterns or relationships found in the data through exploration analysis (EDA) and feature importance analysis were used to identify the best predictors. Using GridSearchCV for hyperparameter tuning, ML models such as Random Forest, Decision Tree, and Logistic Regression were trained and optimized. The system was developed as a web application, using a MySQL database; Django and Flask for the backend; and Next.js for the front end. It was put to the test for stress, scalability, and usability testing with medical experts. Lastly, the app was deployed on Vercel, a cloud app service for building web apps, and documented in depth to ensure reliability and accessibility.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 11 Issue 44 (September 2026) PP. 975-991

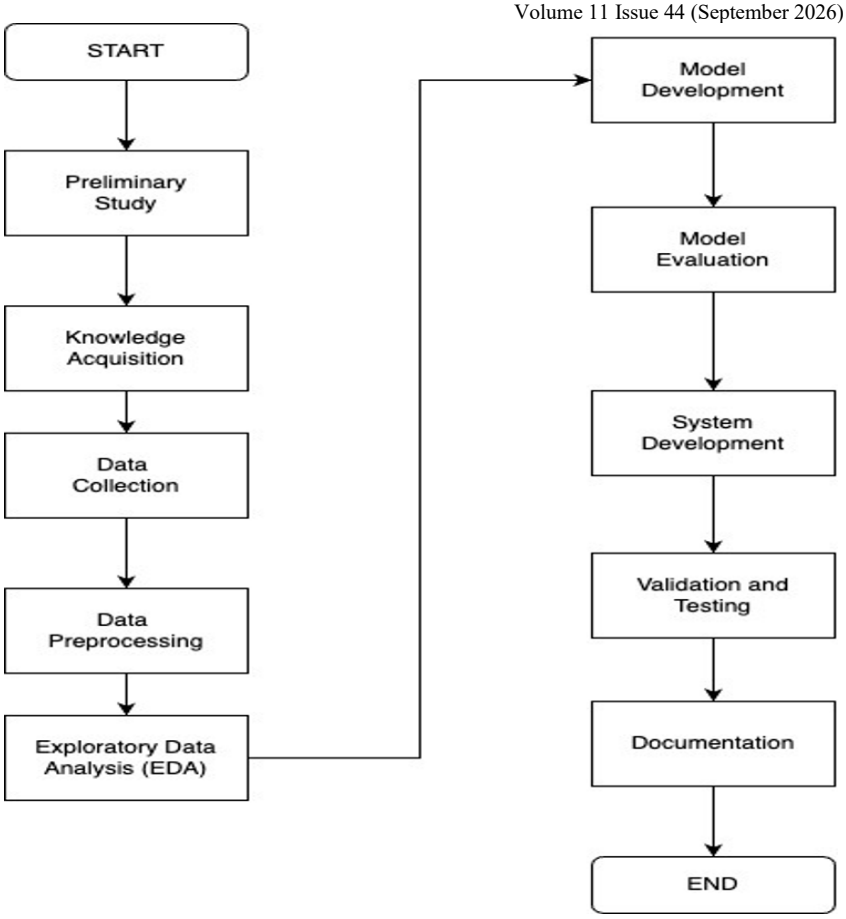



**Figure 1: Research Methodology**

Figure 1 shows the Research Methodology used in this research. The following subsections explain it in detail.

### *Preliminary Study*

This preliminary research examined how heart disease is currently predicted, current trends, and the challenges in predicting it. The research work was based on a review of research papers, medical publications, and reports on heart disease prediction methods. The study identified three widely used ML models, key risk factors, and their accuracy. The research compared ML algorithms such as Random Forest, Decision Tree, and Logistic Regression. This evaluation helped define the project's scope. These results also guided a few research questions and empirical hypotheses, which identified gaps in the study literature and contributed to current knowledge.

### *Knowledge Acquisition*

At this stage, we collected information from reliable sources, including scientific journals, symposium papers, and medical literature. The goal is to build a solid basis for selecting features when creating a predictive model. The most important factors influencing heart disease prediction are age, sex, blood pressure, cholesterol, blood sugar level, ECG reading, and heart rate. It's also a matter of "reading up" knowledge that medical experts require in order to understand what it's all about. These inputs helped improve data quality and ensured the predictive model captured the key attributes that play a major role in a person's risk of heart

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 11 Issue 44 (September 2026) PP. 975-991

disease. In this step, it is confirmed that each specified parameter aligns with the actual diagnostic conditions, allowing the parameters to move to the next stage; in other words, the actual test.

### *Data Collection*

This study used a heart disease dataset from the UCI Machine Learning Repository and Kaggle. The dataset contains data for 304 patients, including various health-status attributes related to heart disease and disease type. It includes demographic, clinical, and physiological data: age, sex, chest pain type, resting blood pressure, heart disease risk prediction, exercise-induced angina, maximum HR attained, ST depression, ST slope, and fasting blood sugar > 120 mg/dl.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | age | sex | cp | trestbps | chol | fbs | restecg | thalach | exang | oldpeak | slope | ca | thal | target |
| | 63 | 1 | 3 | 145 | 233 | 1 | 0 | 150 | 0 | 2.3 | 0 | 0 | 1 | 1 |
| | 37 | 1 | 2 | 130 | 250 | 0 | 1 | 187 | 0 | 3.5 | 0 | 0 | 2 | 1 |
| | 41 | 0 | 1 | 130 | 204 | 0 | 0 | 172 | 0 | 1.4 | 2 | 0 | 2 | 1 |
| | 56 | 1 | 1 | 120 | 236 | 0 | 1 | 178 | 0 | 0.8 | 2 | 0 | 2 | 1 |
| | 57 | 0 | 0 | 120 | 354 | 0 | 1 | 163 | 1 | 0.6 | 2 | 0 | 2 | 1 |
| | 57 | 1 | 0 | 140 | 192 | 0 | 1 | 148 | 0 | 0.4 | 1 | 0 | 1 | 1 |
| | 56 | 0 | 1 | 140 | 294 | 0 | 0 | 153 | 0 | 1.3 | 1 | 0 | 2 | 1 |
| | 44 | 1 | 1 | 120 | 263 | 0 | 1 | 173 | 0 | 0 | 2 | 0 | 3 | 1 |
| | 52 | 1 | 2 | 172 | 199 | 1 | 1 | 162 | 0 | 0.5 | 2 | 0 | 3 | 1 |
| | 57 | 1 | 2 | 150 | 168 | 0 | 1 | 174 | 0 | 1.6 | 2 | 0 | 2 | 1 |
| | 54 | 1 | 0 | 140 | 239 | 0 | 1 | 160 | 0 | 1.2 | 2 | 0 | 2 | 1 |
| | 48 | 0 | 2 | 130 | 275 | 0 | 1 | 139 | 0 | 0.2 | 2 | 0 | 2 | 1 |
| | 49 | 1 | 1 | 130 | 266 | 0 | 1 | 171 | 0 | 0.6 | 2 | 0 | 2 | 1 |
| | 64 | 1 | 3 | 110 | 211 | 0 | 0 | 144 | 1 | 1.8 | 1 | 0 | 2 | 1 |
| | 58 | 0 | 3 | 150 | 283 | 1 | 0 | 162 | 0 | 1 | 2 | 0 | 2 | 1 |

**Figure 2: A Segment of Heart Disease Dataset**

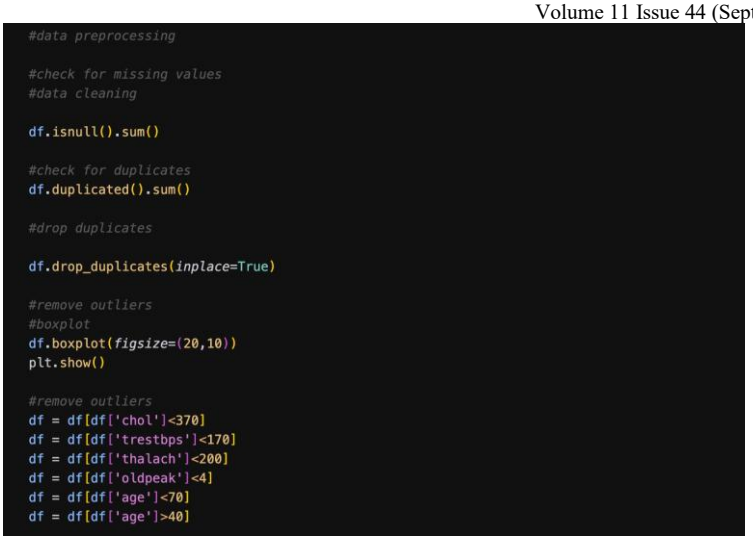
This research used the heart disease data as shown in Figure 2. This dataset is a key tool for developing an efficient heart disease prediction system for timely detection and prevention. Therefore, including key health parameters improved the predictive model's efficacy, accuracy, and reliability.

### *Data Preprocessing*

Data preprocessing enhances data quality and model performance. First, missing values were analyzed and removed to avoid bias in predictions. The second step was to normalize the dataset, followed by statistical and visualization (boxplot) methods to put all of the features and data variables on a comparable scale. Binary encoding was used for categorical variables, and Python's StandardScaler was used to scale the numerical features. Feature selection used correlation analysis to identify the most important predictors of heart disease incidence, improving model efficiency and effectiveness. It removed any unnecessary, duplicate, or out-of-context information.

The resulting data was then split into 80% training and 20% testing so the model could train on one portion and test on data it had not seen, providing a fair performance assessment. Below is a segment of the Python code used for data cleansing operations.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 11 Issue 44 (September 2026) PP. 975-991



```
#data preprocessing

#check for missing values
#data cleaning

df.isnull().sum()

#check for duplicates
df.duplicated().sum()

#drop duplicates

df.drop_duplicates(inplace=True)

#remove outliers
#boxplot
df.boxplot(figsize=(20,10))
plt.show()

#remove outliers
df = df[df['chol']<370]
df = df[df['trestbps']<170]
df = df[df['thalach']<200]
df = df[df['oldpeak']<4]
df = df[df['age']<70]
df = df[df['age']>40]
```

**Figure 3: A Segment of Python Code for Data Cleaning**

Figure 3 shows the relation of data pre-processing, and its primary goal is to maintain data integrity by eliminating these pairs. Each entry is duplicated to make them unique, and the extreme values are isolated and corrected to improve data quality. Hence, the dataset is now noiseless and unmixed for ML operations. Age, cholesterol, and blood pressure are continuous variables in the dataset. Categorical variables such as sex, chest pain type, and thalassemia were encoded using either one-hot or label encoding to convert them into numerical vectors that the model can interpret. It spots outliers, distinguishes potential outliers, and identifies them using graphical methods, which were used to treat them.

### *Exploratory Data Analysis (EDA)*

Exploratory Data Analysis (EDA) describes the data's features and identifies relationships between the variables involved. It shows how the information can predict heart disease. The EDA process begins by analyzing the statistical characteristics of features such as blood pressure, cholesterol, age, and other selected medical measures. Correlations are then examined to identify relationships between features and the target and to inform feature selection.

Visualization techniques such as histograms, box plots, and heatmaps can identify patterns, outliers, and missing values. These ideas help to improve data quality to build an effective prediction model. Factor distribution, outlier detection, and the correlation of features with the target variable were analyzed through statistical summaries and visualizations like histograms, heatmaps, boxplots, and scatterplots.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 11 Issue 44 (September 2026) PP. 975-991

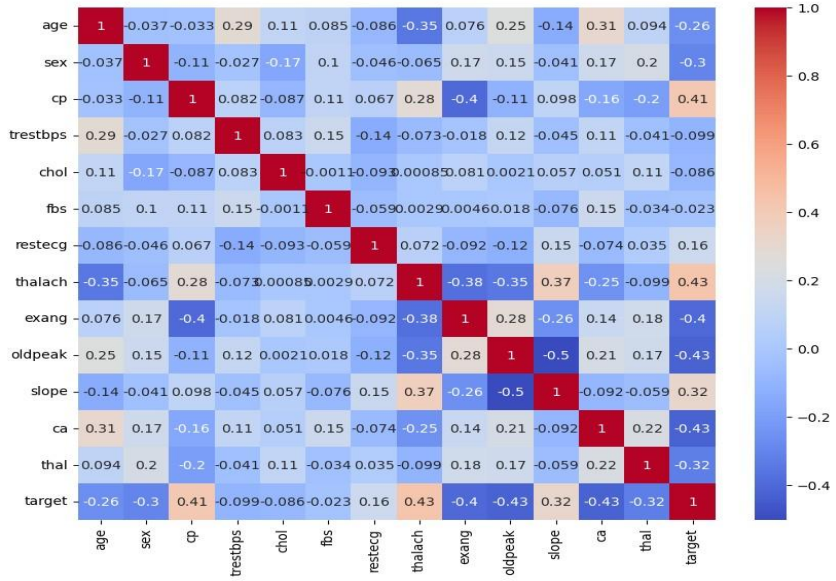



**Figure 4: Heatmap of Features**

The correlation heatmap in Figure 4 shows positive or negative correlations between the features in the heart disease dataset. Correlation coefficients range from -1 to +1 and reflect the strength and direction of the relationship between variables; a positive coefficient indicates a strong positive correlation, while a negative coefficient indicates an inverse relationship. A correlation closer to 0 means that there is little to no relationship. In the heatmap, the values of 'cp' (chest pain type) and 'thalach' (maximum heart rate achieved) have a very similar positive association with the target variable (heart disease or not), whereas 'old peak' and 'ca' (number of major vessels coloured) have highly negative associations with 'heart disease' due to their tendency to be low. This indicates the significance of these features in predicting heart disease.

### *Model Development*

Model development involved formulating and training ML models to accurately predict heart disease outcomes on the pre-processed dataset. The following three ML algorithms were selected for heart disease prediction in this study. Logistic Regression, Decision Trees, and Random Forest were selected to analyze important patient health features. These algorithms use different learning methods and provide a more comprehensive evaluation of heart disease risk.

Hyperparameter tuning was employed to tune each of the algorithms for optimum performance. This encompassed parameters that impact the learning process, such as the Logistic Regression learning rate, Decision Tree max depth, and Random Forest number of trees. This may help increase accuracy, precision, F1 score, and recall. GridSearchCV and RandomizedSearchCV were used for the Hyperparameter search.

This tuning improved performance by strengthening the models and setting the correct hyperparameters for each model. Thus, all models have been developed from pre-processed data. The training set trains the models, while the test set evaluates them. Cross-validation was used to combat overfitting and improve generalization to unseen data.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 11 Issue 44 (September 2026) PP. 975-991

### *Model Evaluation*

This study emphasizes model evaluation in terms of prediction success, reliability, and the predictive model's robustness. The results were evaluated using five metrics to assess model performance: accuracy, precision, recall, F1-score, and Receiver Operating Characteristic (ROC) Area Under the Curve (AUC). The terms accuracy and correctness are used to measure correctness; the number of false positives is reduced by precision, the number of true positives by Recall is increased by the F1-score balances the number of false positives and the number of true positives; the ROC-AUC measures the distinction between the classes; the higher the greater the value. The “best” performing model in terms of these metrics was selected as the one to use for the final prediction.

### *System Development*

During the development phase, the system's user interface was designed through the use of Figma. By combining Next.js for its user-friendliness and speed with Django, a powerful Python web framework for managing data, the application delivers a slick experience and efficient back-end functionality. The platform uses minimal data to analyze heart disease risk. The platform then passes the data to the embedded, pre-trained ML model for processing. The model is optimized for high accuracy and performance to provide accurate risk assessment. This system is designed to help make better decisions and to achieve high-quality heart disease risk assessment.

### *Validation and Testing*

This phase confirms the system's accuracy, reliability, and performance through validation and testing. The ML models for predicting heart disease are validated for accuracy, precision, recall, and F1-score. Testing begins with cross-validation and includes analysis of performance metrics to assess the predictive models' effectiveness. System Testing tests both backend and frontend components, how they handle data, and how they interact with the Database code, to confirm that data is handled and processed correctly and that user interactions occur as expected.

Moreover, User Acceptance Testing allows users to provide feedback, which can help enhance the tool based on their experience and expectations. Therefore, the above validation and testing were done systematically to evaluate the system's capability and the robustness, usability, and effectiveness of heart disease prediction.

## Results and Discussion

This section describes the predictions of the Logistic Regression, Random Forest, and Decision Tree models. It provides a detailed account of the project results and paves the way for practical use. The results include model performance, feature importance, and practical actions from the study. The process includes extensive data processing, EDA, model training, model validation, and deployment as a user-friendly web application for healthcare professionals to provide care for the illness. The model comparison results are shown in Table 1.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 11 Issue 44 (September 2026) PP. 975-991

**Table 1: Model Comparison Results**

| Model | Accuracy | Precision | Recall | F1-Score | ROC-AUC |
|---|---|---|---|---|---|
| **Logistic Regression** | 86.27% | 90.00% | 87.10% | 88.52% | 89.19% |
| **Decision Tree** | 66.67% | 75.00% | 67.74% | 71.19% | 66.37% |
| **Random Forest** | 78.43% | 83.33% | 80.65% | 81.97% | 84.60% |

As shown in Table 1, Logistic Regression achieved the highest accuracy (86.27%), precision (90.00%), recall (87.10%), F1 score (88.52%), and ROC-AUC (89.19%) among the models. Other models, Random Forest and Decision Tree, had moderate accuracy of 78.43% and 66.27%, respectively. Among the models, Logistic Regression was the best performer based on the evaluation metrics. This model is widely used because it is precise and has strong recall, particularly in healthcare, where reducing diagnostic mistakes is important. Random Forest also performs well, but Logistic Regression is more explainable and efficient. To make the results easily customizable, a set of 304 records was used to train, evaluate, and compare different ML models. The results were easy to customize because the number of records (304) was used for training, evaluation, and comparison of various ML models.

The high precision (90.00%) indicates that the model is not likely to make many "false positives" or wrong predictions, and the high recall (87.10%) shows that the model is likely to be accurate in capturing the actual number of heart disease cases. These qualities matter in health care products, as misrisk prediction can delay treatment or lead to inaction.

Random Forest is well known for handling non-linear relationships and avoiding overfitting with smaller datasets, but for this reason, it is still less effective than Logistic Regression. The dataset (304 cases) is relatively small and cannot fully take advantage of complex interactions of features with ensemble methods. The least accurate model was Decision Tree, with an accuracy of only 66.67%. A decrease in the number of tests: The model was too precise on the training data and didn't generalize well to unseen samples, likely due to overfitting.

Based on the experimental results, Logistic Regression gave better accuracy in heart disease prediction than both Random Forest and Decision Tree. Logistic Regression is a relatively simple classification algorithm, but it performed well and achieved the best results across all classification evaluation measures. This means that a linear decision boundary is a good model of the relationship between the clinical features that are selected and the risk of heart disease.

Therefore, Logistic Regression was selected as the final trained model because it showed high prediction accuracy, was relatively simple to compute, and was easy to interpret. In healthcare, interpretability is essential because it helps clinicians understand the link between risk factors and predictions, enabling them to deliver effective treatment. That is, Logistic Regression provides a trade-off between performance and interpretability in a heart disease predictor model.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 11 Issue 44 (September 2026) PP. 975-991

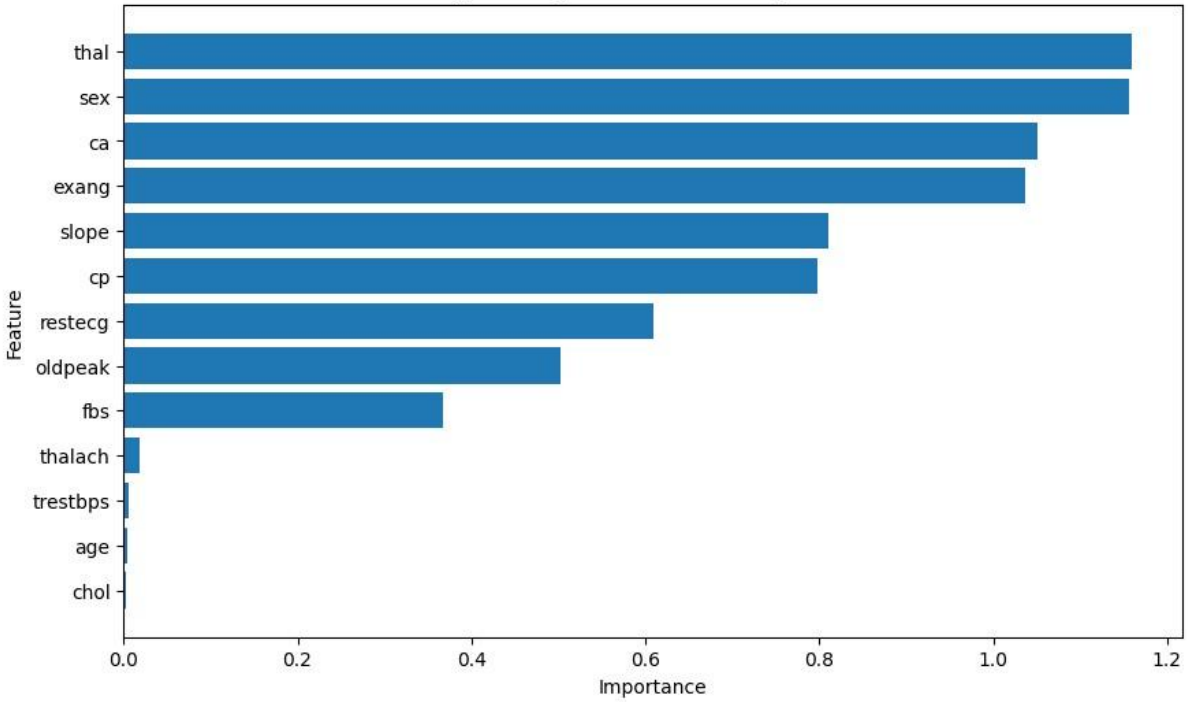



**Figure 5: Logistic Regression Feature Importance**

The feature importance ranking for the Logistic Regression model predicting heart disease is shown in Figure 5. The importance of a feature in the prediction of the target is indicated by the length of the bar: the longer the bar, the more important the feature. Thalassemia (thal) is the top one in the model, and the major vessel has coloured (ca) at ranks 2 and 3. This means that the model prefers to use the features in this order to predict. These features show a very strong relationship with heart disease, indicating they strongly influence model predictions.

Furthermore, these three types (exercise-induced angina [exang], ST-segment slope [slope], and chest pain type [cp]) contribute substantially, as they are effective and relevant for heart disease risk prediction. Conversely, resting ECG (restecg) and oldpeak (ST depression on exercise) are less important but still have considerable clinical value when assessing the presence of heart abnormalities.

Less significant features, such as fasting blood sugar (fbs), maximum heart rate achieved (thalach), and resting blood pressure (trestbps), were also much less associated with heart disease risk. Other factors, including age and blood cholesterol (chol), are not very important in this model, although they are important in heart conditions.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 11 Issue 44 (September 2026) PP. 975-991

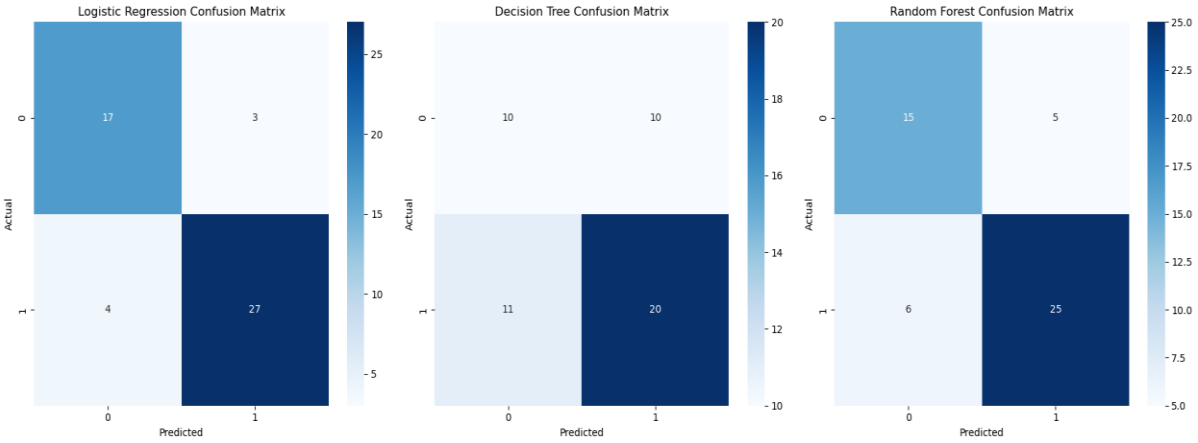



**Figure 6: Confusion Matrix for Logistic Regression, Decision Tree and Random Forest**

Figure 6 shows the confusion matrices produced by the Logistic Regression, Decision Tree, and Random Forest models. The confusion matrices assess the performance and classifications of three ML models: Logistic Regression, Decision Tree, and Random Forest. Logistic Regression performed best with this training and the fewest errors, correctly identifying most cases. Random Forest was the second best, with reasonably good accuracy but a bit more incorrectly classified case. The Decision Tree was the least reliable and also misclassified some of the heart disease-negative cases. Overall, Logistic Regression and Random Forest perform better at predicting heart disease, and improving the Decision Tree can also be beneficial.

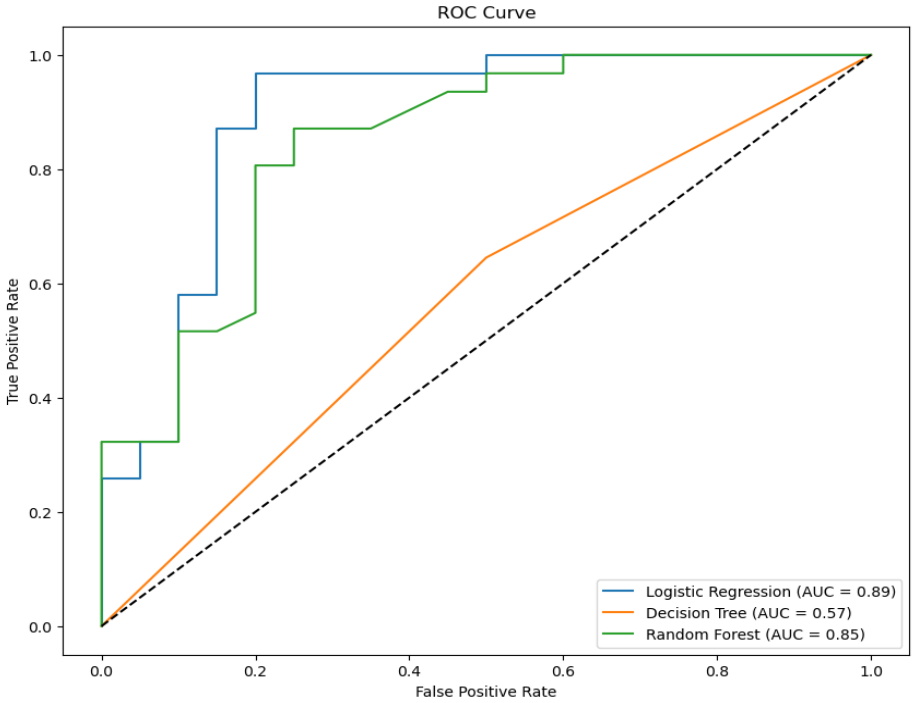



**Figure 7: ROC-AUC Evaluation Metrics**

Figure 7 shows the ROC curves for Logistic Regression, Decision Tree, and Random Forest, which compare true positive and false positive rates. Overall effectiveness is measured by the Area Under the Curve (AUC); a higher score indicates better performance. Logistic Regression gives the best results (AUC = 0.89), followed by Random Forest (AUC = 0.85), while Decision Tree gives the worst results (AUC = 0.57). As shown, Logistic Regression and Random Forest

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 11 Issue 44 (September 2026) PP. 975-991

predict heart disease better, while Decision Trees predict positive and negative cases slightly better, respectively. The system successfully met the objective of predicting heart disease risk with good accuracy.

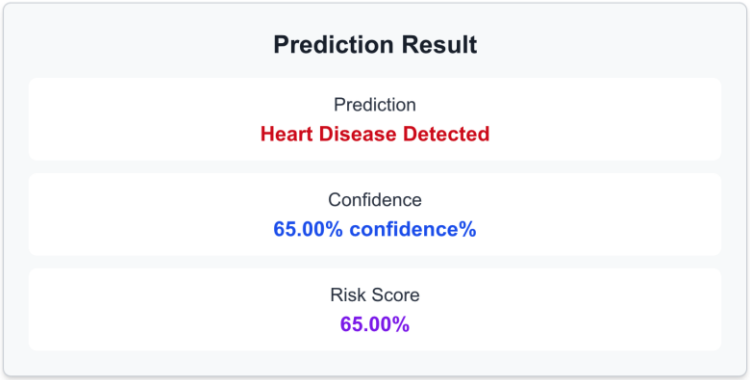



**Figure 8: Risk Prediction Result User Interface**

Figure 8 illustrates the solution, which provides patients with a heart health check based on entering certain health parameters. This simple product allows healthcare professionals to get immediate feedback, making it a more accessible tool for risk assessment and effective disease management.

## Conclusion

With recent advances in predictive analytics and ML, intelligent systems can analyze medical information and forecast disease risk. The researchers demonstrated machine learning's power to solve real-world problems that impact human lives, such as heart disease, by creating an efficient early-stage heart disease prediction system. Logistic Regression was the most accurate algorithm; compared with Decision Trees and Random Forest, it achieved higher accuracy, precision, recall, F1-Score, and ROC-AUC. In this case, Logistic Regression achieved the highest accuracy, with Decision Tree at 86.27% and Random Forest at 85.25%. For this reason, this study aimed to design and test ML models to predict cardiac disease. Next, to build an interactive web application interface, this study designed and developed a patient data input screen that lets clinicians easily enter patient data, predict multiple events, and make data-informed decisions using Next.js, Django, and Flask. Additional features such as ‘type of chest pain' (ACP, PC, OCP), the number of significant vessels coloured, and ‘ST depression' were associated with an enhanced predictive value and clinical insight.

The selection was successful, and the model was implemented in the CardioRisk_Me web application, demonstrating the feasibility of early disease detection using ML. The study underscores the power of predictive analytics to improve clinical outcomes, enable customized care, and reduce acute responses from the healthcare team. CardioRisk_Me represents a step toward improved clinical results in modern medicine and increased technology integration. Future research will provide more clinical data, develop additional ML algorithms, and conduct more experiments.

| | |
|---|---|
| **Acknowledgements:** | The authors sincerely thank Universiti Teknologi MARA for providing the necessary resources and support throughout this research. We also thank colleagues and peers for valuable insights and constructive feedback that greatly improved this paper. |
| **Funding Statement:** | This research received no external financial support |
| **Conflict of Interest Statement:** | The authors declare that there is no conflict of interest regarding the publication of this paper. All authors have contributed to this work and approved the final version of the manuscript for submission to the International Journal of Information System and Technology Management (JISTM). |
| **Ethics Statement:** | This study did not involve any human participants, animals, or sensitive data requiring ethical approval. The authors confirm that the research was conducted in accordance with accepted academic integrity and ethical publishing standards. |
| **Author Contribution Statement:** | All authors: Daniyal Rosli, Zanariah Idrus, Ahmad Afif Ahmarofi and Nurul Husna Mahadzir contributed significantly to the development of this manuscript. All authors read and approved the final version of the manuscript prior to submission. |
| **Declaration on the Use of Artificial Intelligence (AI)** | "The author(s) declare that no generative artificial intelligence (AI) or AI-assisted technologies were used in the preparation, writing, analysis, or revision of this manuscript." |

## References

Ahmarofi, A. A. (2022). An improved momentum rate in artificial neural networks for estimating product cycle time at semi-automatic production. In *Human-centered technology for a better tomorrow: Proceedings of HUMENS 2021* (pp. 193–202). Springer Singapore. https://doi.org/10.1007/978-981-16-4115-2_15

Ahmarofi, A. A., Abidin, N. Z., & Zabidi, N. Z. (2022). Prediction of intensive care cases for COVID-19 pandemic in Malaysia: An artificial neural networks approach. *ASM Science Journal*, *16*, 1–10. https://doi.org/10.32802/asmscj.2021.747

Ahmarofi, A. A., Gunny, A. A. N., Jamil, J. M., & Amlus, N. (2022). The optimization of the halophilic cellulase production: A 3-2-1 multilayer perceptron artificial neural network approach. In *Symposium on Intelligent Manufacturing and Mechatronics* (pp. 40–46). Springer Nature Singapore. https://doi.org/10.1007/978-981-16-8954-3_5

Albahri, A. S., Duhaim, A. M., Fadhel, M. A., Alnoor, A., Baqer, N. S., Alzubaidi, L., Hamid, R. A., Alamoodi, A. H., Albahri, O. S., Zaidan, B. B., Peh, S. C., Gu, X., Al-qaysi, Z. T., Garfan, S., Mohammed, K. I., Al-Bayati, B., Salih, M. M., Zaidan, A. A., Al-Obaidi, J. R., ... Deveci, M. (2023). A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. *Information Fusion*, *96*, 156–191. https://doi.org/10.1016/j.inffus.2023.03.008

Ali, F., El-Sappagh, S., Islam, S. R., Kwak, D., Ali, A., Imran, M., & Kwak, K. S. (2020). A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion. *Information Fusion*, *63*, 208–222. https://doi.org/10.1016/j.inffus.2020.06.008

Aliasgharzadeh, S., Yagin, N. L., Namazi, N., Haji, S., Alamdari, N. M., Pourmoradian, S., & Ghorbanverdi, Z. (2026). Depression severity and cardiovascular risk in patients with type 2 diabetes mellitus: Insights from the Framingham Risk Score. *Scientific Reports*, 1–8. https://doi.org/10.1038/s41598-026-43077-8

Bevan, G., Pandey, A., Griggs, S., Dalton, J. E., Zidar, D., Patel, S., Wright, J. T., Jr., Rajagopalan, S., & Al-Kindi, S. (2023). Neighborhood-level social vulnerability and prevalence of cardiovascular risk factors and coronary heart disease. *Current Problems in Cardiology*, *48*(8), Article 101182. https://doi.org/10.1016/j.cpcardiol.2022.101182

Choudhury, A., Mondal, A., & Sarkar, S. (2024). Searches for the BSM scenarios at the LHC using decision tree-based machine learning algorithms: A comparative study and review of random forest, AdaBoost, XGBoost and LightGBM frameworks. *The European Physical Journal Special Topics*, 1–39. https://doi.org/10.1140/epjs/s11734-024-01308-x

Dogheim, G. M., & Hussain, A. (2023). Patient care through AI-driven remote monitoring: Analyzing the role of predictive models and intelligent alerts in preventive medicine. *Journal of Contemporary Healthcare Analytics*, *7*(1), 94–110. https://api.semanticscholar.org/CorpusID:276021100

Gaziano, T. A. (2022). Cardiovascular diseases worldwide. In *A public health approach to cardiovascular disease prevention and management* (pp. 8–18). CRC Press. https://doi.org/10.1201/b23266

Ghaffar Nia, N., Kaplanoglu, E., & Nasab, A. (2023). Evaluation of artificial intelligence techniques in disease risk prediction and prediction. *Discover Artificial Intelligence*, *3*(1), Article 5. https://doi.org/10.1007/s44163-023-00049-5

Gray, M. P., Berman, Y., Bottà, G., Grieve, S. M., Ho, A., Hu, J., Nelson, M. R., & Rogers, C. (2023). Incorporating a polygenic risk score-triaged coronary calcium score into cardiovascular disease examinations to identify subclinical coronary artery disease

(ESCALATE): Protocol for a prospective, nonrandomized implementation trial. *American Heart Journal*, *264*, 163–173. https://doi.org/10.1016/j.ahj.2023.06.012

Hughes, A., Shandhi, M. M. H., Master, H., Dunn, J., & Brittain, E. (2023). Wearable devices in cardiovascular medicine. *Circulation Research*, *132*(5), 652–670. https://doi.org/10.1161/CIRCRESAHA.122.3223

Husnain, A., Saeed, A., Hussain, A., Ahmad, A., & Gondal, M. N. (2024). Harnessing AI for early detection of cardiovascular diseases: Insights from predictive models using patient data. *International Journal for Multidisciplinary Research*, *6*(5), 1–12. https://doi.org/10.36948/ijfmr.2024.v06i05.27878

Karna, V. V. R., Karna, V. R., Janamala, V., Devana, V. K. R., Ch, V. R. S., & Tummala, A. B. (2024). A comprehensive review on heart disease risk prediction using machine learning and deep learning algorithms. *Archives of Computational Methods in Engineering*, 1–33. https://doi.org/10.1007/s11831-024-10194-4

Law, J. H., Sultan, N., & Finer, S. (2023). Advancing the communication of genetic risk for cardiometabolic diseases: A critical interpretive synthesis. *BMC Medicine*, *21*, Article 432. https://doi.org/10.1186/s12916-023-03150-9

Li, C., Liu, X., Shen, P., Sun, Y., Zhou, T., Chen, W., Liu, Z., Wang, Z., & Gao, P. (2024). Improving cardiovascular risk prediction through machine learning modeling of irregularly repeated electronic health records. *European Heart Journal – Digital Health*, *5*(1), 30–40. https://doi.org/10.1093/ehjdh/ztad058

Liu, W., Laranjo, L., Klimis, H., Chiang, J., Yue, J., Marschner, S., Thiagalingam, A., & Chow, C. K. (2023). Machine-learning versus traditional approaches for atherosclerotic cardiovascular risk prognostication in primary prevention cohorts: A systematic review and meta-analysis. *European Heart Journal – Quality of Care and Clinical Outcomes*, *9*(4), 310–322. https://doi.org/10.1093/ehjqcco/qcad017

MacKay, C., Klement, W., Vanberkel, P., Lamond, N., Urquhart, R., & Rigby, M. (2023). A framework for implementing machine learning in healthcare based on the concepts of preconditions and postconditions. *Healthcare Analytics*, *3*, Article 100155. https://doi.org/10.1016/j.health.2023.100155

Netala, V. R., Teertam, S. K., Li, H., & Zhang, Z. (2024). A comprehensive review of cardiovascular disease management: Cardiac biomarkers, imaging modalities, pharmacotherapy, surgical interventions, and herbal remedies. *Cells*, *13*(17), Article 1471. https://doi.org/10.3390/cells13171471

Pakdil, F., Ebrahimpour, M., Yih, Y., Zhang, Z., & Viravan, B. (2026). Trends in 30-day readmission prediction: A systematic review and method–condition mapping of analytical approaches. *Computers & Industrial Engineering*, *220*, Article 112230. https://doi.org/10.1016/j.cie.2026.112230

Pan, B., Hirota, K., Jia, Z., & Dai, Y. (2023). A review of multimodal emotion recognition from datasets, preprocessing, features, and fusion methods. *Neurocomputing*, *558*, Article 126866. https://doi.org/10.1016/j.neucom.2023.126866

Qureshi, N. Q., Mufarrih, S. H., Bloomfield, G. S., Tariq, W., Almas, A., Mokdad, A. H., Bartlett, J., Nisar, I., Siddiqi, S., Bhutta, Z., Mark, D., Douglas, P. S., & Samad, Z. (2021). Disparities in cardiovascular research output and disease outcomes among high-, middle- and low-income countries: An analysis of global cardiovascular publications over the last decade (2008–2017). *Global Heart*, *16*(1), Article 4. https://doi.org/10.5334/gh.815

Thomas, A., Ryan, C. P., Caspi, A., Liu, Z., Moffitt, T. E., Sugden, K., Williams, B. S., Belsky, D. W., & Gu, Y. (2024). Diet, pace of biological aging, and risk of dementia in the

Framingham Heart Study. *Annals of Neurology*, *95*(6), 1069–1079. https://doi.org/10.1002/ana.26900

Udegbe, F. C., Nwankwo, E. I., Igwama, G. T., & Olaboye, J. A. (2023). Real-time data integration in diagnostic devices for predictive modeling of infectious disease outbreaks. *Computer Science & IT Research Journal*, *4*(3), 1–18. https://doi.org/10.51594/csitrj.v4i3.1502

Ullah, M., Hamayun, S., Wahab, A., Khan, S. U., Rehman, M. U., Haq, Z. U., Rahman, G., & Naeem, M. (2023). Smart technologies are used as smart tools in the management of cardiovascular disease and their future perspective. *Current Problems in Cardiology*, *48*(11), Article 101922. https://doi.org/10.1016/j.cpcardiol.2023.101922

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

Yang, H., Luo, Y. M., Ma, C. Y., Zhang, T. Y., Zhou, T., Ren, X. L., Zhang, J., & Lin, H. (2023). A gender-specific risk assessment of coronary heart disease based on physical examination data. *NPJ Digital Medicine*, *6*(1), Article 136. https://doi.org/10.1038/s41746-023-00887-8

Zhang, L., Tong, Z., Han, R., Guo, R., Zang, S., Zhang, X., Li, X., & Yang, Y. (2023). Global, regional, and national burdens of ischemic heart disease attributable to smoking from 1990 to 2019. *Journal of the American Heart Association*, *12*(3), Article e028193. https://doi.org/10.1161/JAHA.122.028193