Cureus Journal of Computer Science

Open Access Original Article

Post-COVID-19 Heart Attack Mortality Rate Prediction and Classification in Vaccinated Adults Using Random Forest Algorithm

Received 05/22/2025
Review began 06/10/2025
Review ended 08/12/2025
Published 09/02/2025

DOI:
https://doi.org/10.7759/s44389-025-06920-z

Uttam U. Deshpande [1] , Mahendra M. Dixit [2] , Ramchandra Alias Ameet Chate [3] , Sowmyashree H. Srinivasaiah [4] , Aishwarya Jamakandi [1]

1. Department of Electronics and Communication Engineering, KLS Gogte Institute of Technology, Belagavi, IND 2. Department of Electronics and Communication Engineering, KLS Vishwanathrao Deshpande Institute of Technology, Haliyal, IND 3. Department of MBA, KLS Gogte Institute of Technology, Belagavi, IND 4. Department of Computer Applications, Bangalore Institute of Technology, Bengaluru, IND

Corresponding author: Uttam U. Deshpande, uttamudeshpande@gmail.com

Abstract

The COVID-19 pandemic has had a significant effect on lifestyle, health, and the economy globally, with about 12.10 million people suffering from serious health issues. Emerging clinical evidence suggests that people recovering from COVID-19 are more vulnerable to cardiovascular problems, including heart attacks, and case trends in India are varied with noticeable spikes. Comprehensive data on prognostic indicators and long-term outcomes, particularly in vaccinated populations, are still lacking, nevertheless. By creating a prediction model based on machine learning to evaluate heart disease risk and categorization in the post-COVID era, this study seeks to close that gap. Key risk variables and underlying trends are identified by the model using Random Forest Regressor and Classifier algorithms to examine patient data. Given the pandemic's effects on lifestyle choices, heart health, and access to healthcare, the study focuses on the time after the initial waves of COVID-19. The suggested method identified at-risk patients and classified different types of cardiac disease with an accuracy rate of almost 92%, providing a useful tool for targeted healthcare and early intervention in the post-pandemic environment.

Categories: Algorithm Analysis, Forensic Analysis, Machine Learning (ML)
Keywords: covid-19, post-pandemic, mortality rate, heart disease classification, covid vaccination, random forest regressor and classifier

Introduction

According to the World Health Organization, cardiovascular disease continues to be the leading cause of mortality worldwide, contributing to almost one-third of all fatalities annually [1]. Early detection and precise prognosis of cardiac problems remain difficult despite tremendous progress in medical diagnostics and treatment. Heart illness includes a broad spectrum of ailments that impact the structure and function of the heart, such as heart failure, arrhythmia, valve problems, cardiomyopathies, and coronary artery disease. A complicated interaction of lifestyle variables, underlying comorbidities, and genetic predispositions influences these disorders. Although they frequently call for specialized equipment and skilled interpretation, traditional diagnostic techniques like electrocardiograms (ECG), stress tests, and echocardiograms aid in identifying structural and functional issues.

The emergence of machine learning in recent years has created new opportunities for cardiac disease prediction and early diagnosis [2]. Complex datasets can be analyzed by machine learning models to find connections and patterns that traditional approaches can miss. These technologies have the potential to minimize the need for intrusive testing while increasing accuracy and lowering diagnostic costs [3]. Notably, machine learning algorithms such as Random Forest (RF), Support Vector Machine (SVM), and Neural Networks have demonstrated encouraging outcomes in automating tasks related to the identification of cardiac disease, which enhances clinical effectiveness and decision-making [4]. Management of cardiovascular health has become much more challenging because of the COVID-19 epidemic [5]. Myocarditis, pericarditis, and thromboembolic events are among the cardiac problems that people recovering from COVID-19 may be more susceptible to. Furthermore, lifestyle changes brought on by lockdowns, postponed medical attention, and increased stress have made pre-existing cardiac disorders worse and caused new cardiovascular problems. Given the uncommon but noteworthy side effects, like vaccine-associated myocarditis, that have occasionally been documented, vaccinated populations also need to be closely monitored.

This study investigates how machine learning can be used to assess cardiovascular risk in persons who have received the COVID-19 vaccine. The method achieved substantial generalization and excellent predictive accuracy by merging known cardiac datasets with mortality and immunization information. By including vaccination-linked mortality markers, the work adds to the body of existing literature and

How to cite this article
Deshpande U U, Dixit M M, Chate R, et al. (September 02, 2025) Post-COVID-19 Heart Attack Mortality Rate Prediction and Classification in Vaccinated Adults Using Random Forest Algorithm. Cureus J Comput Sci 2 : es44389-025-06920-z. DOI https://doi.org/10.7759/s44389-025-06920-z

provides a new data dimension for improving clinical decision support in the post-pandemic context.

Literature survey

The goal of utilizing machine learning algorithms is to create an environment with a wealth of data that can enhance the disease prediction's attributes. The literature presents various machine learning algorithms that are developed to predict cardiovascular diseases. Predicting post-pandemic cardiovascular diseases at a reasonable rate is a significant difficulty in many healthcare institutions. It is critical to diagnose cardiovascular diseases at a lower medical cost, and less time is significant. The COVID-19 virus and the COVID-19 vaccination have an impact on the cardiovascular system [6]. Occasionally, COVID-19 vaccination has caused myocarditis or pericarditis conditions, and patients have experienced vague symptoms resulting from heart problems. Myocardial damage, pericarditis, coagulopathy, arrhythmias, heart failure, myocardial infarction, and a chronic post-acute risk of poor cardiovascular outcomes are among the cardiovascular consequences associated with COVID-19 [7]. By comparing patient data from June and December 2020 with that from 2019, the reports have claimed that the COVID-19 pandemic has altered acute decompensated heart failure [8]. Data from 4,806 patients at 22 Indian facilities were analyzed, with a focus on admission trends, causes, outcomes, and adherence to guidelines-directed medical therapy. The findings show that acute decompensated heart failure admissions decreased by 20% during the pandemic (from 2,675 in 2019 to 2,131 in 2020). The involvement of machine learning is changing the way that evidence is created in clinical research [9] in the future. In this article, the potential of machine learning to transform clinical research is thoroughly examined. It highlights the significant advancements that machine learning may bring to several research opportunities, including preclinical drug development and discovery, clinical trial participant management, data collection, and data analysis.

Zakariaee et al. [10] utilized an extensive dataset comprising chest CT severity ratings to evaluate various machine learning techniques for forecasting COVID-19 mortality. The study evaluated 55 characteristics in six categories from 6,854 suspected cases retrospectively and used the Chi-square independence test to identify significant prognostic markers. Eight machine learning models were trained and evaluated using metrics including Area Under the Curve (AUC), sensitivity, specificity, accuracy, and precision: naive Bayes (NB), SVM, K-Nearest Neighbors (K-NN), linear regression, RF, XGBoost, J48, and multi-layer perceptron. The greatest outcomes of the RF approach were 97.2% accuracy, 100% sensitivity, 94.8% precision, 94.5% specificity, 97.3% F1-score, and 99.9% AUC. According to the findings, machine learning models may successfully stratify COVID-19 patients based on risk, which can help with resource allocation and early detection of high-risk patients to increase survival. Ali et al. [11] developed a machine learning technique for analyzing the risk factors associated with patients' heart failure forecasting. In this study, risk factors were analyzed using machine learning techniques to predict the survival of patients with heart failure. The researchers employed five supervised machine learning techniques: Gradient Boosting, RF, Decision Tree Regressor, Decision Tree (DT), and XGBoost. The algorithms' results were evaluated; RF produced the best accuracy of 97.78% based on accuracy, precision, recall, F-measure, and log loss.

Potential impacts of the COVID-19 virus on the heart and circulatory system were discovered by the study [12]. The 2019 COVID-19 has been associated with arrhythmias, cardiomyopathy, and heart attacks. To create a machine learning-based heart disease prediction system for the Indian populace, Maini et al. [13] conducted a pilot study in South India. According to the report, the main cause of death in India is the failure to recognize cardiovascular disease in its early stages. Using machine learning techniques, India may be able to develop an early warning system for cardiovascular diseases that is both economical and efficient. The prediction system was built using the Python programming language and four of the most advanced machine learning algorithms: KNN, NB, Logistic Regression (LR), AdaBoost, and RF. Machine learning provided a precise calculation of the risk of cardiovascular illness. The top-performing (RF) prediction system identified 470 out of 501 medical data points correctly, yielding a diagnosis accuracy of 53.8%. The measured values were 94.6% for specificity and 92.8% for sensitivity. The predictive system achieved a positive predictive value of 94% and a negative predictive value of 93.6%. Karthick et al. [14] developed a machine learning model for heart disease risk prediction using the Gaussian NB, SVM, LR, LightGBM, XGBoost, and RF algorithms.

The top features from the Cleveland heart disease dataset were chosen for this study by the authors using the Chi-square statistical test. The RF Classifier model achieved the highest classification accuracy rate of 88.5% following feature selection [14]. Singh and Kumar [15] examined the prediction of cardiac disease using machine learning methods. This study made use of the UCI repository's heart disease dataset. They compared the performance of several machine learning algorithms, such as LR, KNN, SVM, NB, DT, and RF, in predicting cardiac disease. According to the study, SVM and NB reported 88.52% accuracy, whereas RF reported the highest accuracy of 91.80%. Khdair and Dasari [16] explored different machine learning algorithms for coronary artery disease prediction. Nine features and a dataset of 462 medical cases from the South African heart disease dataset were used in the study. A total of 160 records have coronary heart disease, and 302 have good health. The k-means algorithm and synthetic minority oversampling method were employed in this work to address the issue of unbalanced data. From clinical data, four distinct machine learning techniques, Long Short-Term Memory, SVM, KNN, and Artificial Neural Network, can be used to accurately predict coronary artery disease events. According to the findings, SVM performed the best by producing 78.1% accuracy. The effectiveness of LR, KNN, SVM, NB, DT, RF, and XGBoost machine learning models for heart disease prediction was compared by Kumar Sahoo et al. [17]. The models were trained using the Cleveland heart disease dataset from the UCI machine learning library. With a classification accuracy of 90.16%, the RF algorithm outperformed the other tested machine learning algorithms [17].

Our research aims to improve health outcomes by providing advanced tools for early detection, classification, and prevention of heart disease in the post-pandemic population, addressing the complex challenges posed by the pandemic on cardiovascular health. The main aim of the proposed work is to establish and apply frameworks for machine learning that can precisely detect and categorize cardiac conditions and forecast death rates in the post-pandemic adult vaccination era. We examine how the coronavirus outbreak has affected the incidence and mortality rate of cardiac disease. We utilize an RF Regressor to analyze the effect of the COVID-19 vaccine on patients' hearts. A DT is split using the RF algorithm, which then gathers characteristics at random to choose features. Because of this, the model is less susceptible to certain traits and is better equipped to represent the complex correlations seen in the data without being unduly impacted by a single variable. RF uses the majority voting method to determine the final prediction. An advantage of RF is its ensemble-based decision-making ability. It produces precise predictions and is incredibly efficient at managing huge datasets with complex decision requirements.

Materials And Methods

Proposed system

It is now vital to comprehend how immunization, infection history, and heart health interact as the world recovers from the COVID-19 epidemic. To identify and categorize cardiac disease in vaccinated people with high accuracy and clinical relevance, our suggested method integrates real-time predictions, machine learning algorithms, and public health data. Vaccinated adults' post-COVID-19 heart attack mortality rates are predicted and classified in Figure *1*.

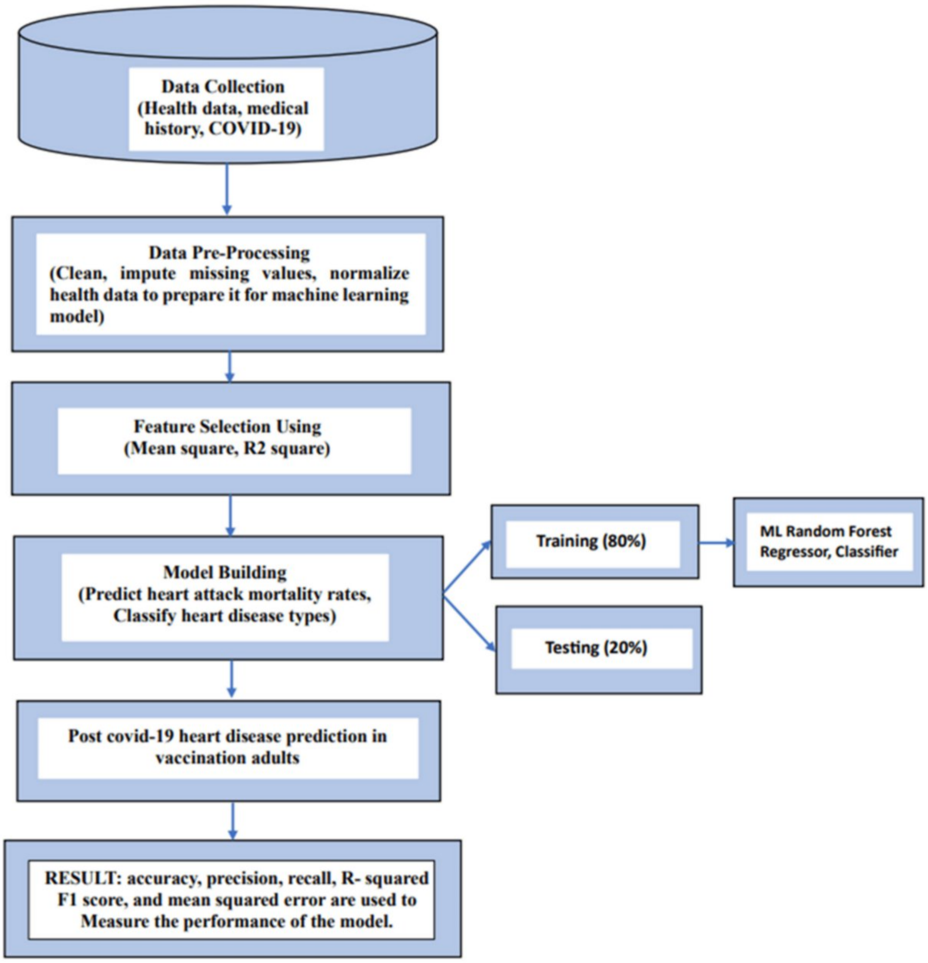

FIGURE 1: Proposed workflow of post-COVID-19 heart disease prediction in vaccinated adults.

Dataset Collection

We used a well-known public dataset for diagnosing heart disease [18], which consists of 13 important variables and 1,025 patient records (Table *1*). These characteristics give a multifaceted picture of cardiac health and include age, cholesterol, thallium heart scan results, and more. The PRS Legislative Research ("PRS") dataset [19] is used to forecast and classify the post-pandemic vaccine mortality rate status. The dataset's participants range in age from 29 to 77. These were divided into five bands for analytical clarity: ages 29-39, 40-49, 50-59, 60-69, and 70-77. This grouping provides a distinct demographic profile that makes it possible to identify post-COVID-19 mortality patterns and age-related trends in heart disease risk.

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal
63	1	3	145	233	1	0	150	0	2.3	0	0	1
37	1	2	130	250	0	1	187	0	3.5	0	0	2
41	0	1	130	204	0	0	172	0	1.4	2	0	2
56	1	1	120	236	0	1	178	0	0.8	2	0	2
57	0	0	120	354	0	1	163	1	0.6	2	0	2
57	0	0	140	241	0	1	123	1	0.2	1	0	3
45	1	3	110	264	0	1	132	0	1.2	1	0	3
68	1	0	144	193	1	1	141	0	3.4	1	2	3
57	1	0	130	131	0	1	115	1	1.2	1	1	3
57	0	1	130	236	0	0	174	0	0	1	1	2

TABLE 1: Sample heart disease identification information.

Source [18]

The definitions and descriptions of the dataset features (age, sex, cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, and thal) are outlined in Table 2.

Figure *2* highlights the unpredictable development of the COVID-19 pandemic in India from March 12, 2020, to June 30, 2022. The pandemic's changing trend is represented by the four lines: red for fatalities, yellow for recovered patients, green for active cases, and blue for confirmed cases. While the green line varies with infection surges, the blue line rises dramatically at each wave peak. The red line indicates a continuing risk of death, albeit at its lowest, while the yellow line follows, showing a slower recovery trajectory.

The purpose of this study is to investigate the relationship between the trajectory of COVID-19, immunization initiatives, and their combined impact on heart disease outcomes. The background for the present investigation is provided by this contextual image. We used two main sources to create a composite dataset to accomplish this:

• A well-known publicly available dataset on heart disease [18], which includes more than 1,025 patient records with comprehensive clinical information.

• The PRS Legislative Research COVID-19 vaccination and mortality database [19], which offers the most recent data on vaccination patterns and fatality rates. From the PRS Legislative Research COVID-19 database, the "Mortality Rate (1-100%)" variable was retrieved. It was calculated by dividing the total number of COVID-19-related deaths in each demographic category by the number of vaccinated individuals in that group and then multiplying the result by 100. To facilitate consistent cross-group analysis, the conversion into a percentage guaranteed a consistent scale, enabling the smooth integration of mortality rates with the clinical and demographic characteristics of the heart disease dataset. A systematic pipeline that harmonized feature definitions and aligned variable formats was used to integrate the datasets. The most frequent value in each demographic stratum was used to fill in categorical omissions, while median substitution was used to fill in gaps in numerical records. By aligning vaccination and death rates with the heart disease dataset's reporting intervals, temporal consistency was guaranteed. A Synthetic Minority Over-Sampling Technique (SMOTE) was used to address category imbalance in disease classes, producing a balanced training input without altering the underlying distributions.

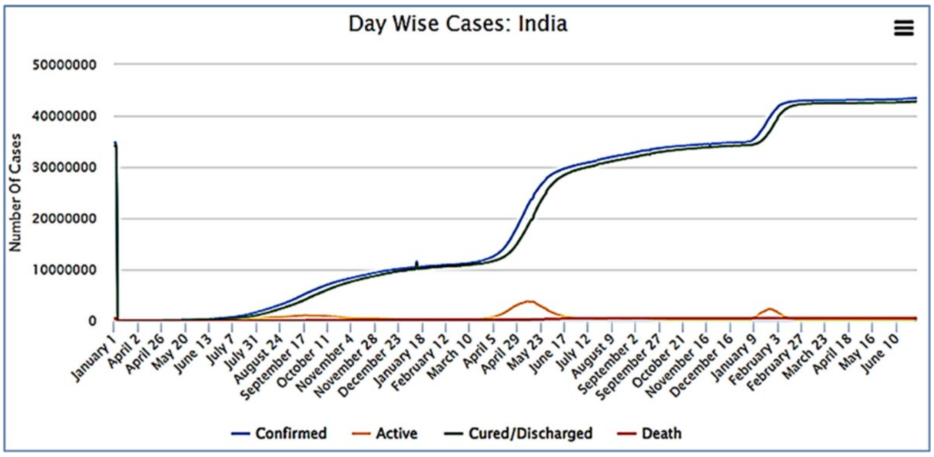

FIGURE 2: India's day-wise COVID-19 cases.

Source: [19]

With over 1,002 patient records and 18 essential variables, our extended dataset combines demographic, cardiac, and COVID-19 immunization data on a single, potent analytical basis. Table *2* lists these characteristics, which range from vaccination status and mortality rate to blood pressure and the type of chest discomfort.

Feature	Description	Feature	Description
Age	Patient's age (in years)	Oldpeak	ST depression due to exercise
Sex	Gender (1 = male, 0 = female)	Slope	ST-segment slope: upsloping, flat, or downsloping
Cp	Chest pain type (1–4): 1 = typical, 2 = atypical, 3 = non-anginal, 4 = none	Ca	Major vessels colored by fluoroscopy (0–3)
Trestbps	Resting blood pressure (mmHg)	Thal	Thalassemia: 3 = normal, 6 = fixed defect, 7 = reversible defect
Chol	Serum cholesterol (mg/dL)	Num	Heart disease severity (0 = none; 1–4 = increasing severity)
Fbs	Fasting blood sugar >120 mg/dL (1 = true, 0 = false)	Vaccine 1	First COVID-19 vaccine dose (1 = yes, 0 = no)
Restecg	ECG results: 0 = normal, 1 = ST-T abnormality, 2 = LV hypertrophy	Vaccine 2	Second COVID-19 vaccine dose (1 = yes, 0 = no)
Thalach	Max heart rate achieved	Mortality Rate	Death likelihood (1–100%)
Exang	Exercise-induced angina (1 = yes, 0 = no)	Classification	Heart disease type (0–10): 0 = None, 1 = Coronary Artery Disease, 2 = HF, 3 = Arrhythmia, 4 = Valve, 5 = Cardiomyopathy, 6 = Congenital, 7 = Pericarditis, 8 = Myocarditis, 9 = Hypertensive, 10 = Rheumatic

TABLE 2: Heart disease dataset correlates with COVID-19 vaccination and mortality data.

The mortality rate (%) was calculated by multiplying by 100 after dividing the number of COVID-19-related deaths by the total number of vaccinated people in the relevant demographic group.

Here are some salient features of the attributes:

• Lifestyle and demographics such as age, sex, height, and weight-based BMI, alcohol consumption, and smoking status.

• Clinical indicators include resting electrocardiogram (ECG) readings, thallium stress test results, and fasting blood sugar.

• Heart-related indicators like vessel imaging (ca), ST-segment slope (slope), and chest pain kinds (cp).

• Disease classification comprises 11 categories of heart disease, ranging from rheumatic heart disease to coronary artery disease.

With the use of this extensive feature set, our system can predict not only if an individual is at risk for heart disease, but also the sort of heart disease they may have, as well as the potential effects of COVID-19 immunization and mortality patterns. It is a data-driven advancement in predictive healthcare that connects cardiovascular risk assessment and pandemic response. This data is examined using Python's Pandas module and organized in CSV format to make it useful. This makes it simple to convert the data into a Data Frame, a versatile data structure perfect for machine learning applications. To improve model performance and lower noise, preprocessing eliminates redundant or unnecessary fields. To forecast adult vaccinated individuals' heart attack death rates and identify trends that may aid in averting further cardiac occurrences. Through the integration of current COVID-19 mortality and immunization data with conventional cardiovascular measurements, this work opens opportunities for customized post-pandemic risk prediction tools that may eventually help physicians make quicker, more informed, and potentially life-saving decisions.

Data Preprocessing and Feature Selection

Extensive preprocessing was performed on the raw dataset to guarantee precise predictions. The Pandas library in Python was utilized to eliminate redundant features, deal with missing data, and eliminate superfluous information. Then, to guarantee uniformity among all entries, the data was normalized. We used R-squared (R^2) and mean squared error (MSE) metrics for feature selection to improve model performance. These statistical metrics assisted in determining the most important characteristics associated with the categorization of cardiac disease and the prediction of mortality. This greatly increased the accuracy and efficiency of the model by keeping only the most predictive elements. Essentially, dependable machine learning requires clean, high-quality data, and Pandas offers the necessary resources to do this through efficient transformation, filtering, and validation systems.

Model Development

Several machine learning methods, such as SVM, GB, LR, and KNN, were assessed in conjunction with RF during the initial model selection stage. RF continuously produced better accuracy and generalization on our dataset, according to a comparative performance study. Because of this, the study highlights RF in its final reporting, but it acknowledges that a more comprehensive comparative evaluation of alternative models served as the basis for the strategy. As a standard practice to make sure the model performs well when applied to new data, the dataset was divided into 80% for training and 20% for testing to assess model performance. We have created two models for machine learning: The mortality rate is predicted by the RF Regressor, and the kind of cardiac condition is determined by RF Classifier.

An ensemble learning method [20] RF uses random subsets of characteristics and data to create several decision trees. It uses majority voting for classification and output averaging for regression to improve model stability and forecast accuracy. To produce a final forecast, the RF Regressor creates many decision trees (such as Trees #1, #2, and #3) with randomly chosen attributes, as illustrated in Figure *3*. Because of its versatility, scalability, and resilience, this approach was chosen as the best choice for examining intricate healthcare datasets. By using RF selection and bootstrap sampling, it efficiently manages high-dimensional data, minimizes overfitting, and has good generalization. To preserve only the most predictive elements, MSE and R^2 were employed during model construction. The model's high R^2 score of 0.92 and low MSE indicated that it was quite accurate at predicting mortality risk. About post-COVID cardiac risk prediction, this strategic design guarantees that the model is both interpretable and dependable for clinical decision-making.

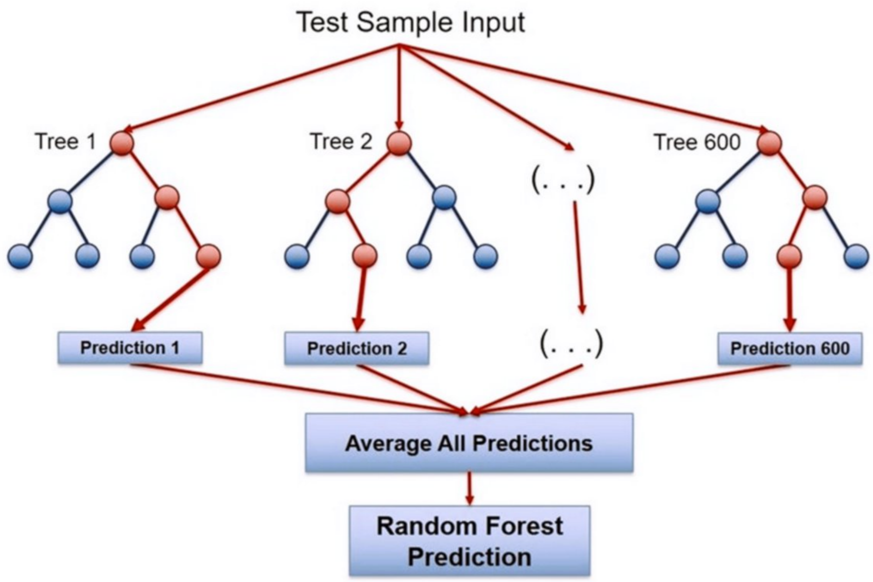

FIGURE 3: Random forest regression process.

Source: [20]

The RF Classifier model [21] uses several decision trees, each trained on distinct feature subsets of the original dataset, as shown in Figure *4*. A class, such as Class C, D, or B, is individually predicted by each tree, and the class with the highest votes is chosen as the final output by majority voting. This ensemble method produces a more robust and dependable model for categorizing different forms of heart disease by reducing overfitting and improving prediction accuracy.

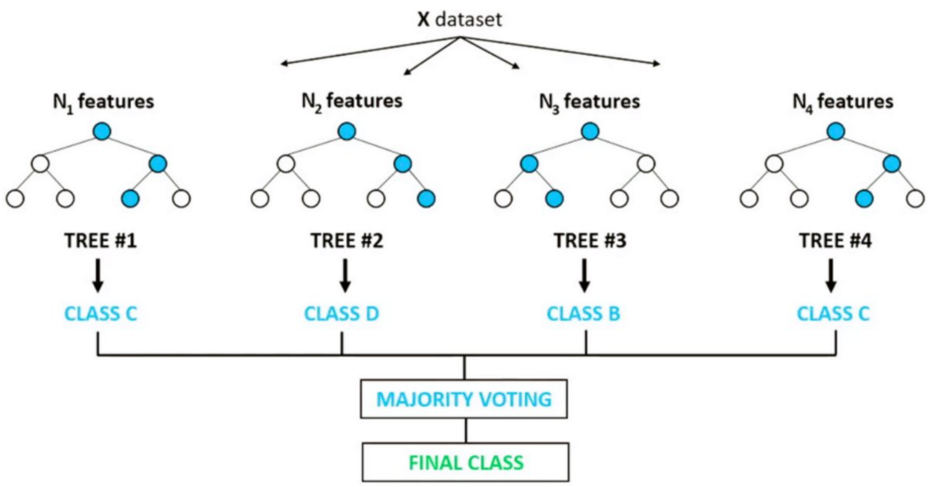

FIGURE 4: Random forest classifier process.

Source: [21]

Model training and testing: Using the training part of the dataset, the RF Regressor and Classifier were trained. To make sure the models functioned consistently across different data segments, we used K-fold cross-validation, a reliable method that enhances generalization and helps avoid overfitting. With this approach, the data is split up into K subsets, and the model is trained on K−1 of them and validated on the remaining one. This process is repeated until all the subsets have been tested. The testing dataset was then used to assess the model's performance, and false negatives (FN), true positives (TP), false positives (FP), true negatives (TN), and confusion matrices tracking these outcomes were used to analyze the results. To determine accuracy, conventional assessment formulas were used. To facilitate seamless integration with the web-based prediction interface, the Django Python framework was used to implement all training, testing, and assessment procedures.

Evaluation Metrics

When evaluating the performance of machine learning models, evaluation metrics are essential since they allow one to compare expected values with the actual results from the testing set. A confusion matrix includes the following quantities that are used for testing:

• The quantity of accurate positive forecasts is known as TP.

• The quantity of erroneously optimistic forecasts is measured as FP.

• Similarly, the number of negative samples that were correctly classified as negative is described as TN.

• The number of TP cases that were mistakenly forecasted as negative is known as FN.

To ensure that models are dependable and effectively generalize to new data, the metrics selected are contingent upon the job type, be it regression or classification. The metrics listed below were used to assess classification models (such as those predicting the different heart disease forms).

• Accuracy: Assesses the overall predictive accuracy of the model.

Accuracy = (TP + TN)/(TP + TN + FP + FN)

• Precision: Assesses how effectively the model forecasts favorable results. One way to describe it is as the ratio of all positive forecasts to precisely predicted positive observations.

Precision = TP/(TP + FP)

• Recall/Sensitivity: The model's recall quantifies its capacity to locate each pertinent instance within the dataset. It is the proportion of all truly positive observations to all accurately projected positive observations.

Recall = TP/(TP + FN)

• F1-score: When there is an unequal class distribution and you need to strike a balance between recall and precision, the F1-score comes in handy.

F1 Score = 2 × ((Precision × Recall)/(Precision + Recall))

• Specificity: Specificity is calculated by dividing the total number of individuals without the disease by the number of true negatives. It shows the proportion of true negatives that a test correctly detects.

Specificity = TN/(TN + FP)

Mortality Rate Prediction

Heart Disease Impact After the Pandemic: The incidence of heart disease has significantly increased since the pandemic, mostly because of COVID-19 aggravating pre-existing cardiovascular disorders. The virus has been shown to directly affect the cardiovascular system, causing inflammation and stress. Additionally, the pandemic delayed the detection and treatment of heart patients and redirected medical resources into handling COVID-19 cases, upsetting the healthcare system. During and after the epidemic, these delays made things worse for those with heart problems. Machine learning models were developed to forecast the death rates from heart attacks and to categorize patients into different types of heart illness, such as myocarditis, arrhythmias, coronary artery disease, and others. Through a Django-based online interface, real-world patient data was entered, allowing the models to be used practically and in real time.

Cardiovascular Health With COVID-19 Vaccination: The COVID-19 vaccine has significantly improved cardiovascular health. The major benefit is that severe infections are strongly associated with several issues, including myocarditis, pericarditis, and acute coronary syndrome. People who have had vaccinations, particularly those who already have cardiac issues, are less likely to develop the disease because the infection is less severe.

Furthermore, hospitalization has decreased because of widespread immunization, which has improved access to care for non-COVID illnesses, including heart disease, and lessened the strain on healthcare systems. In addition, vaccination encourages healthier habits and reduces stress brought on by pandemics, which indirectly improves heart health. The COVID-19 vaccination has, in general, been instrumental in reducing heart disease-related death rates.

Factors influencing mortality: The risk of death in people with heart disease is greatly influenced by several factors:

• Age: Because of a larger chance of comorbidities and a decline in physiological resilience, aging inherently increases susceptibility.

• Comorbidities: Disorders like diabetes and high blood pressure hurt cardiovascular health and raise the chance of death.

• Cardiac Disease Severity: Higher death rates are associated with more advanced illnesses such as myocardial infarction and heart failure.

• Lifestyle Factors: Smoking, a poor diet, and a lack of exercise all influence the development and cause of disease.

• COVID-19 Infection: Severe consequences may arise from an active infection that exacerbates pre-existing heart conditions.

• Immunization Status: Those who are not immunized are more likely to experience serious COVID-19 consequences, which can exacerbate heart problems.

When combined, these factors provide a thorough understanding of the complex relationship between cardiac patient mortality and other factors. Real-time health indicators, updated treatments, and latest clinical research are all factors that improve the predicted accuracy of models. Individual risk assessments and clinical decisions are supported by these models when they are appropriately trained and verified using reliable datasets. For example, early intervention may be necessary if a patient has biomarkers that indicate high risk. Patient outcomes are much enhanced, and more effective care management is made possible by this data-driven approach.

Survival analysis: Survival analysis sheds light on how long it will take for an event, like death or relapse, to occur. Two important statistical methods were applied: under the presumption of proportional hazard ratios, the Cox Proportional Hazards Model evaluates the impact of variables (such as age and comorbidities) on survival time, and Kaplan-Meier Estimator is a non-parametric technique that accounts for censored data (such as patients lost to follow-up) while estimating survival probability across time. In medical research, these instruments are crucial for assessing prognosis and developing customized treatment plans.

Classification of cardiac illness: The study also aimed to classify patients according to the kind of heart disease they would acquire. The model employs machine learning to find trends that relate health information, such as immunization status, to the probability of cardiac disorders, including coronary artery disease, heart failure, ventricular disease, arrhythmias, myocarditis, and other types (such as hypertensive heart disease, cardiomyopathy, and pericarditis). This categorization uses data-driven prediction to promote better patient care, targeted interventions, and early diagnosis.

Implementation and Deployment

The Django framework, which offered a scalable and reliable environment for real-time predictions via an easy-to-use web interface, was used to deploy the learned machine learning models. This implementation streamlined clinical decision-making using data-driven insights by enabling medical staff to enter fresh patient data and obtain instant diagnostic classifications and mortality risk assessments. A scalable and secure data architecture was created to support this feature. To ensure patient security and data integrity, this involved putting in place safe databases and data warehouses by healthcare laws like HIPAA (Health Insurance Portability and Accountability Act). Sensitive data was protected with advanced encryption techniques and role-based access controls, while system security and transparency were guaranteed by regular audits and monitoring practices.

A structured pipeline comprising feature engineering, training, validation, and data preparation was used in the model creation process. Key predictive variables were chosen using methods like MSE and R2 after raw datasets were cleaned and standardized. K-fold cross-validation was used to train and validate models like RF, LR, and Neural Networks to guarantee accuracy and generalizability. These models were incorporated into the current healthcare systems to offer real-time decision support after they were optimized. Clinical professionals were able to obtain forecasts during consultations due to integration with Electronic Health Record (EHR) systems and diagnostic tools. At the point of care, actionable insights, such as identifying high-risk patients based on comorbidities and vaccination status, were supplied with ease. Initial evaluations of the platform's usability conducted with active doctors emphasized its quick output creation and easy navigation. Data entry redundancy was reduced, according to participants, by interaction with current EHR operations. Interactive visual dashboards and patient-specific longitudinal tracking were suggested as ways to improve interpretability and facilitate follow-up care.

Monitoring and Maintenance

The user interface was created to facilitate real-time communication with the installed predictive models, making it possible for medical practitioners to use them effectively and conveniently. The front-end, which was created using common web technologies like HTML, CSS, and JavaScript, enables users to examine model outputs, enter patient data, and analyze predictions about mortality risk and diagnostic classification. A Django-based backend framework that controls server-side functionality, API routing, and real-time communication with the machine learning models is closely linked with this interface. It also manages database operations for input and output record storage and retrieval, guaranteeing safe data management that complies with privacy regulations in the healthcare industry.

The effectiveness and dependability of the machine learning models are maintained by regularly monitoring their performance after deployment. Key visualization and assessment tools like confusion matrices for classification accuracy, scatter plots for regression error distribution, ROC curves and AUC scores for assessing classification quality, and precision-recall curves, which are especially crucial when there is class imbalance, are used to gain insights into performance. These tools direct required improvements and offer practical insights into the model's capacity for generalization. Regular retraining of the model using newly collected or updated patient data is necessary to maintain its relevance and accuracy over time. This allows for adjustments to treatment regimens, population demographics, and new medical trends. Additionally, the models are modified to consider the latest clinical findings, keeping them in line with accepted diagnostic and treatment guidelines. Performance deterioration or data drift can be identified with continuous assessment using metrics like recall, accuracy, precision, and F1-score. To further improve the system, it is essential to incorporate patient and clinician feedback. Feedback of this kind aids in identifying problems with model interpretability, user interface experience, or forecast dependability. Predictive models can continue to be useful instruments for assisting evidence-based decision-making and improving patient outcomes throughout time if healthcare organizations have a strong structure in place for monitoring and ongoing improvement.

Results

To assess the effectiveness of the suggested predictive approach, a few tests were carried out by merging COVID-19 vaccination and mortality data with conventional heart disease datasets. There were more than 1,000 patient records in the combined dataset. Python was used to construct the system, utilizing machine learning tools like Pandas and Scikit-learn. The dataset was cleaned, noise was eliminated, and non-informative features were eliminated using data preprocessing techniques. To keep the best predictive variables, MSE and R^2 metrics were used to guide feature selection. After that, the dataset was divided into training and testing subsets in an 80:20 ratio.

An RF Classifier was used to classify the different forms of heart disease, and an RF Regressor was used to forecast the mortality rates following COVID-19. The model's performance was assessed using several regression and classification measures, such as ROC-AUC, F1-score, accuracy, precision, recall (sensitivity), and specificity. Scatter plots, ROC curves, precision-recall curves, and confusion matrices were among the visualization methods used to visually convey the results and evaluate the behavior of the model. To enable interface-based result presentation and real-time input handling, the entire system was deployed using the Django web framework. The performance comparisons between the proposed RF Regressor and Classifier models and other state-of-the-art methods are presented in Table *3*. With an MSE of 62.75, the RF Regressor demonstrated a modest average difference between the death rates that were predicted and those that were observed. Strong predictive accuracy is demonstrated by the model's high R^2 value of 0.93, which shows that 93% of the variance in mortality outcomes was correctly captured. With 92% accuracy, the RF Classifier showed remarkable performance in the classification domain, demonstrating its capacity to accurately categorize most test cases. In clinical settings where false positives might result in needless procedures, a precision score of 0.92 guarantees that 92% of positive forecasts were accurate. With a recall score of 0.92, the model's capacity to detect genuine positive cases and reduce missed diagnoses is demonstrated. The F1-score of 0.92, which strikes a compromise between precision and recall, particularly important for datasets with class imbalance, further reflects these values. A remarkable ROC-AUC value of 1.00 indicated near-perfect classification accuracy, and the classifier also showed a high specificity of 0.99, demonstrating its efficacy in accurately recognizing non-cardiac instances.

Algorithms	MSE	R-Squared	Accuracy	Precision	Recall/Sensitivity	F1-score	Specificity	ROC-AUC
RF Algorithm [9]	–	–	53.8%	–	92.8%	–	94.6%	–
RF Algorithm [11]	–	–	97.2%	94.5%	100%	97.3%	94.5%	99.9%
Proposed RF Regressor and Classifier	62.57%	0.93	92%	92%	92%	92%	99%	100%

TABLE 3: Proposed post-COVID-19 heart attack mortality rate prediction comparison performance with the other benchmark models.

AUC-ROC, Area Under the Receiver Operating Characteristic Curve; MSE, Mean Squared Error; RF, Random Forest

As seen in Figure 5, the suggested model consistently performs better than other benchmark models in almost every important aspect. Strong generalization and diagnostic capabilities are demonstrated by the suggested method, which has consistent scores of 92% in Accuracy, Precision, Recall, and F1-Score, 99% in Specificity, and 100% in ROC-AUC. Particularly in a post-pandemic clinical setting where accuracy and early identification are crucial, this graphical representation highlights the model's durability and dependability. The model's suitability for precise mortality risk prediction and heart disease classification is confirmed by these findings taken together, which makes it a useful instrument for improving decision support in healthcare systems.

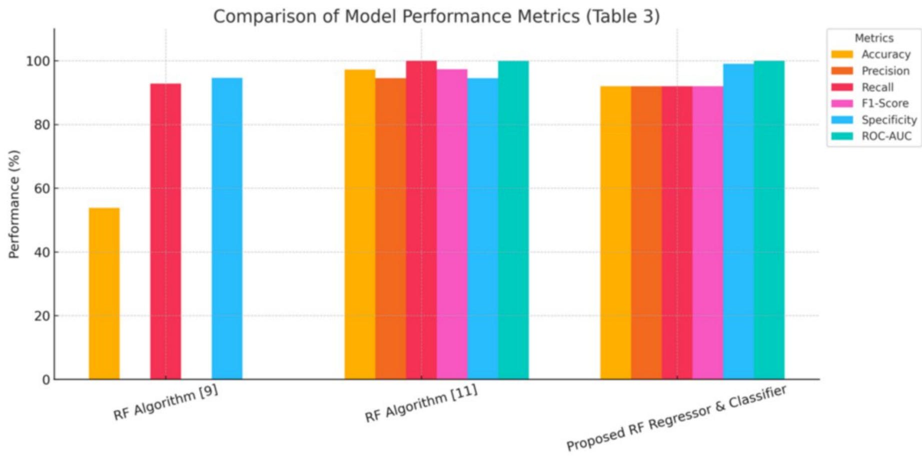

FIGURE 5: Performance comparison of post-COVID-19 heart disease prediction models.

AUC-ROC, Area Under the Receiver Operating Characteristic Curve; RF, Random Forest

Actual vs. predicted mortality rate

A scatter plot contrasting actual death rates with those forecasted by the RF Regressor is shown in Figure 6. The x-axis displays the actual fatality rate, while the y-axis displays the equivalent model forecast. Each blue dot represents a patient record. The ideal situation, where projected values and actual values coincide exactly, is represented by the red dashed line (y = x). Most data points closely match this line, suggesting that the model is performing well and that there is little difference between the actual and anticipated values. Although there are a few outliers that show a small overestimation or underestimate, most fall within a small range surrounding the line. This visual trend demonstrates the model's strong prediction accuracy, which is further supported by the high R^2 value of 0.93 and the low MSE of 62.75. The figure demonstrates the model's overall efficacy in precisely predicting mortality risk in individuals with heart disease who have recovered from COVID-19.

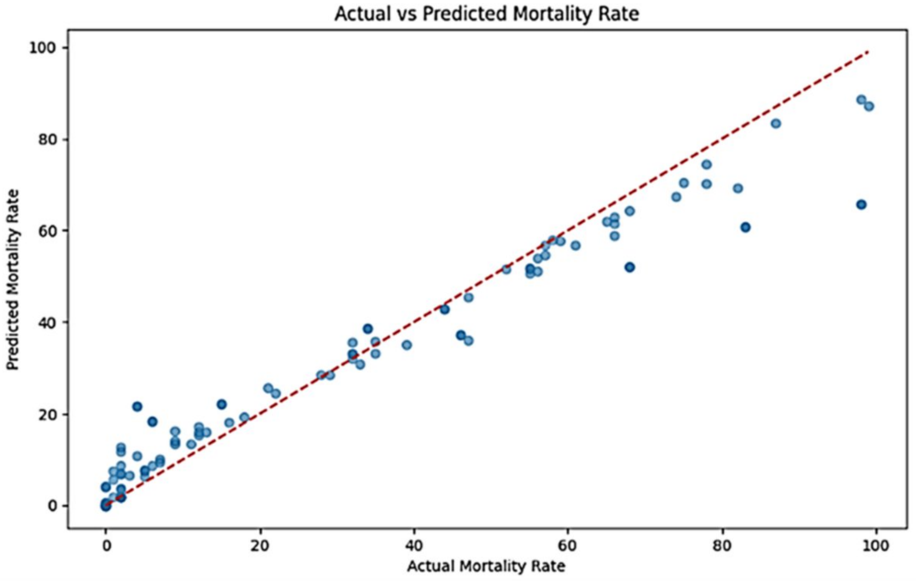

FIGURE 6: Actual vs. predicted mortality rate.

By comparing the real and predicted heart disease categories, Figure 7 illustrates the model's classification ability. The classifier's predicted labels are shown on the y-axis, while the actual classification labels are shown on the x-axis. Perfect categorization would be achieved if every data point fell precisely on the red dashed diagonal line (y = x). There is a remarkable degree of agreement between the projected and actual values in this plot, as evidenced by the large number of blue dots that cluster along this line.

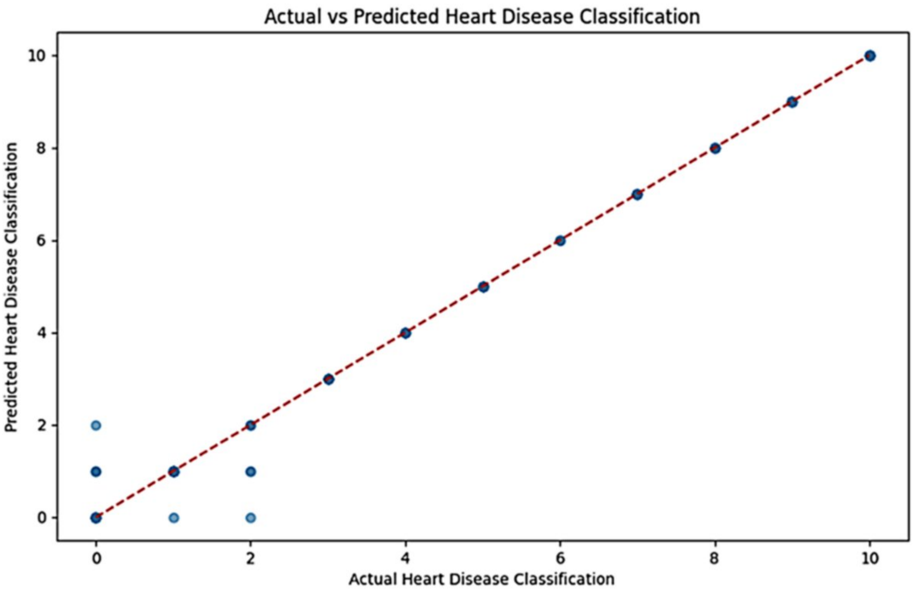

FIGURE 7: Actual vs. predicted heart disease classification.

A small percentage of instances show a modest deviation, suggesting that the model misidentified the type of heart illness a patient had. These variances point to possible areas for model improvement. The classifier's high accuracy, precision, and recall (92%), however, are supported by the good overall alignment of data points, indicating that it successfully distinguishes between various cardiac states. This plot, together with Figure 6, confirms the model's dependability for both regression and classification tasks in post-pandemic clinical settings.

Confusion matrix

The confusion matrix used to assess the effectiveness of the suggested categorization model for heart disease is shown in Figure 8. Each column in this matrix indicates the anticipated class, while each row gives the actual class label. The off-diagonal elements indicate misclassifications, whereas the diagonal elements display the number of cases of each heart disease group that were correctly identified. Higher values are shown by lighter hues, which correlate to the frequency of classifications. The matrix shows that the model performs well in most categories, especially in classes 3 through 10, where predictions and actual labels closely match. However, a few cases were incorrectly placed into adjacent categories, indicating some uncertainty among the lower classes, notably classes 0, 1, and 2. These off-diagonal values point to locations where clinically comparable illnesses need to be better distinguished. The classifier's advantages and disadvantages are generally visualized understandably via the confusion matrix, which also offers suggestions for future feature engineering or model modifications that could improve classification accuracy even further.

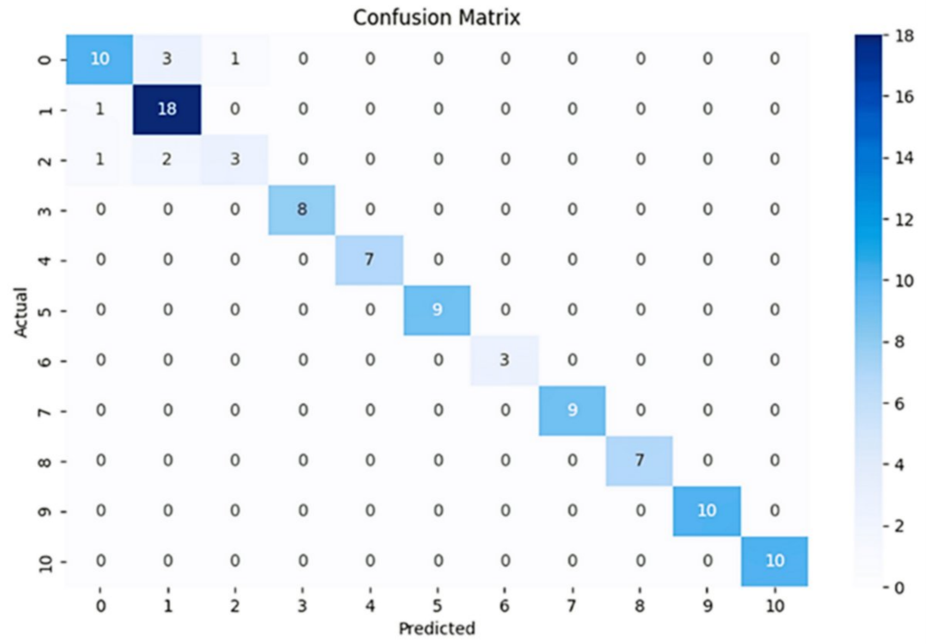

FIGURE 8: Confusion matrix for performance evaluation.

Vaccination status

The distribution of people into "unvaccinated" and "vaccinated" groups according to their COVID-19 vaccination status is depicted in the bar graph in Figure 9. The y-axis displays the total number of people in each category, while the x-axis depicts these groupings. The bars' heights reflect a balanced sample because both groups are almost the same size, with each group having about 500 members. This nearly equal distribution reduces the possibility of bias resulting from unequal group sizes, which is crucial for fair comparisons in research assessing health outcomes like death or infection rates. The x-axis of the graph indicates vaccination status, whereas the y-axis shows the total number of COVID-19 deaths among vaccinated and unvaccinated people. It shows a significantly greater proportion of deaths among those who had vaccinations, which seems to go against generally acknowledged public health evidence indicating immunizations, lower mortality, and serious consequences. A higher overall vaccination rate, variations in age or health status between groups, or methods of data collection and reporting could all have an impact on this seeming oddity. To properly understand these findings, more research into the study's methodology and population characteristics is also necessary.

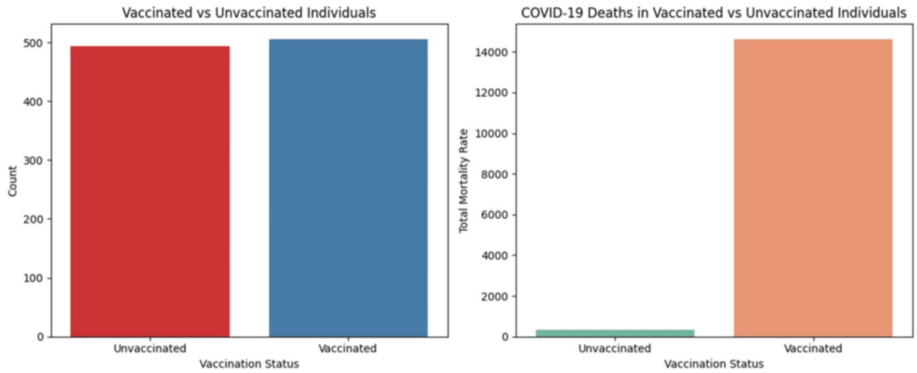

FIGURE 9: Analysis of COVID-19 death rates in vaccinated and unvaccinated people.

ROC curve

An ROC curve that assesses a classification model's performance across several classes is shown in Figure *10*. While the real positive rate is displayed on the y-axis, the false positive rate is represented on the x-axis. A class is represented by each colored line, and the legend shows the AUC. High classification performance is indicated by the majority of curves being grouped in the upper-left corner. While the lowest AUC is still high at 0.97, other classes have a perfect AUC of 1.00. Curves much above the dashed diagonal line, which stands for random chance, demonstrate the model's exceptional capacity to discriminate between classes with a high number of TP and a low number of FP.

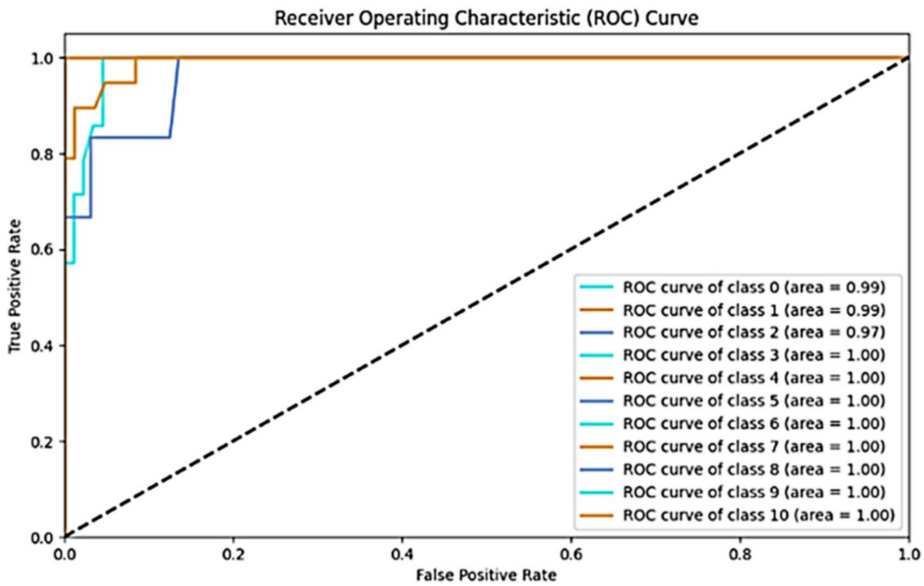

FIGURE 10: Receiver operating characteristic (ROC) curve.

Precision-Recall Curve for Each Class

The precision-recall curve for a multi-class classification model presents the precision versus recall performance for each class in Figure *11*. Every line represents a distinct class, with the x-axis representing recall and the y-axis representing precision. One measure of performance is the AUC. Classes 3 through 10 have flawless classification, meaning they have no FP or FN, with AUC values of 1.00. On the other hand, the AUCs of Classes 0 (0.93), 1 (0.97), and 2 (0.82) are marginally lower, suggesting that those predictions should be improved. All things considered, the curves grouped close to the top right corner show a good model with robust classification abilities in many classes, but they also point out classes where recall and precision might be improved.

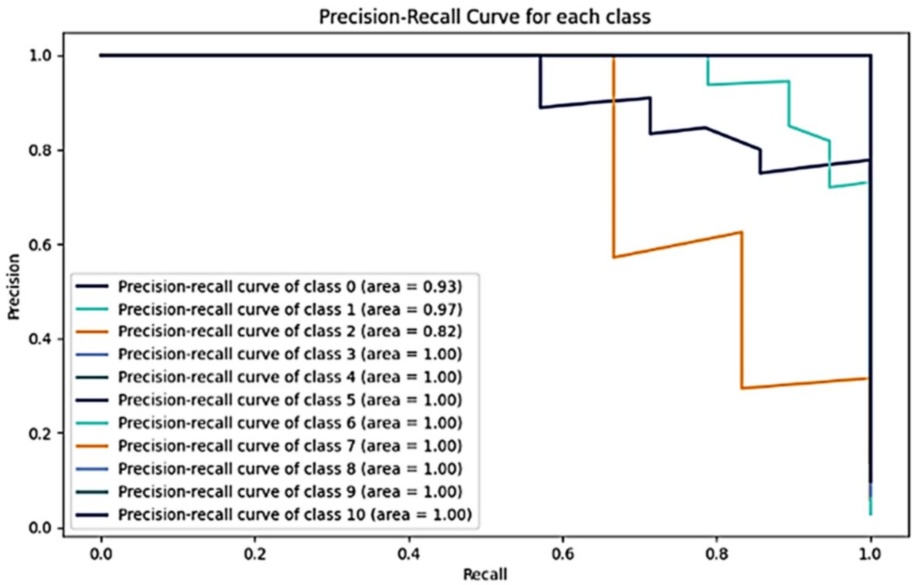

FIGURE 11: Precision-recall curve for each class.

Discussion

The experimental investigation demonstrates the remarkable efficiency of the RF Regressor and Classifier in detecting various forms of heart disease and forecasting the mortality rates from heart attacks in vaccinated patients post-COVID-19. The RF Classifier demonstrated its efficacy in reducing false positives and negatives by achieving remarkable metrics, such as 92% accuracy, precision, recall, and F1-score. With an R^2 value of 0.93 and an MSE of 62.57%, the regression model likewise demonstrated good performance, explaining 93% of the variation in death rates. The precision-recall curves and confusion matrix showed nearly error-free classification, particularly for disorders like cardiomyopathy, myocarditis, and congenital heart disease, while ROC curves demonstrated great discriminative power with AUC values as high as 1.00. All things considered, the findings support the model's stability and applicability for actual medical applications, including post-COVID-19 cardiovascular risks in vaccinated populations.

Conclusions

To predict post-COVID-19 heart attack mortality in vaccinated patients, we developed an RF-based framework for classifying cardiac diseases. The classifier achieved 92% accuracy, with good precision, recall, and F1-scores. The regression model explained 93% of the variance in mortality (R^2 = 0.93, MSE = 62.57%). Its effectiveness, particularly in identifying cardiomyopathy, myocarditis, and congenital heart disease, is demonstrated by its practically perfect ROC-AUC and precision-recall measures. Together, these results highlight the model's usefulness in improving post-pandemic cardiovascular risk assessment. Although the results are promising, the size and diversity of the dataset’s sources limit the current work's applicability to populations with varying health characteristics. Despite the harmonization procedure, some inconsistencies may be introduced by differences in the methods used to get the data from different sources. Assessment of long-term post-COVID cardiovascular outcomes was limited by the lack of longitudinal follow-up data and external validation on independent cohorts.

Future research can improve the model's therapeutic relevance and adaptability by using larger and more varied datasets, possibly developed using federated learning. Further prediction robustness improvement can be explored by integration of multimodal inputs, including wearable data, genomics, imaging, and ECGs. An interpretable, lightweight pipeline can be developed that works with mobile and IoT platforms to facilitate real-time use. Finally, clinical validation can be achieved by benchmarking against cutting-edge deep-learning techniques and conducting trials in various healthcare contexts, guaranteeing end users' performance and transparency.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Uttam U. Deshpande, Mahendra M. Dixit, Aishwarya Jamakandi

Acquisition, analysis, or interpretation of data: Uttam U. Deshpande, Ramchandra Alias Ameet Chate, Sowmyashree H. Srinivasaiah, Aishwarya Jamakandi

Drafting of the manuscript: Uttam U. Deshpande, Mahendra M. Dixit, Ramchandra Alias Ameet Chate

Critical review of the manuscript for important intellectual content: Uttam U. Deshpande, Sowmyashree H. Srinivasaiah, Aishwarya Jamakandi

Supervision: Uttam U. Deshpande

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. World Health Organization: World health statistics 2021: monitoring health for the SDGs, sustainable development goals. (2021). Accessed: June 6, 2025: https://www.who.int/publications/i/item/9789240027053.
2. Rahebi J: Fishier mantis optimiser: a swarm intelligence algorithm for clustering images of COVID-19 pandemic. International Journal of Nanotechnology. 2023, 20:25-49. 10.1504/ijnt.2023.131111
3. Sameni R, Clifford GD: Review of fetal ECG signal processing issues and promising directions. The Open Pacing Electrophysiology & Therapy Journal. 2010, 3:4-20. 10.2174/1876536x01003010004
4. Al-Safi H, Munilla J, Rahebi J: Patient privacy in smart cities by blockchain technology and feature selection with Harris Hawks Optimization (HHO) algorithm and machine learning. Multimedia Tools and Applications. 2022, 81:8719-8743. 10.1007/s11042-022-12164-z
5. Choi WY: Mortality rate of patients with COVID-19 based on underlying health conditions. Disaster Medicine and Public Health Preparedness. 2021, 16:2480-2485. 10.1017/dmp.2021.139
6. Hridoy KM, Akash SI, Dipti FA, Hasan M, Ovi JA, Al-Imran M: Heart disease prediction using machine learning algorithms. 2023 4th International Conference on Big Data Analytics and Practices (IBDAP), Bangkok, Thailand. 2023, 1-6. 10.1109/IBDAP58581.2023.10271997
7. Akhtar Z, Trent M, Moa A, Tan TC, Fröbert O, MacIntyre CR: The impact of COVID-19 and COVID vaccination on cardiovascular outcomes. European Heart Journal Supplements. 2023, 25:A42-A49. 10.1093/eurheartjsupp/suac123
8. Jayagopal PB, Abdullakutty J, Sridhar L, et al.: Acute decompensated heart failure (ADHF) during COVID-19 pandemic-insights from South India. Indian Heart Journal. 2021, 73:464-469. 10.1016/j.ihj.2021.06.014
9. Weissler EH, Naumann T, Andersson T: The role of machine learning in clinical research: transforming the future of evidence generation. Trials. 2021, 22:537. 10.1186/s13063-021-05489-x
10. Zakariaee SS, Naderi N, Ebrahimi M, Kazemi-Arpanahi H: Comparing machine learning algorithms to predict COVID-19 mortality using a dataset including chest computed tomography severity score data. Scientific Reports. 2023, 13:11343. 10.1038/s41598-023-38133-6
11. Ali MM, Al-Doori VS, Mirzah N, et al.: A machine learning approach for risk factors analysis and survival prediction of heart failure patients. Healthcare Analytics. 2023, 3:100182. 10.1016/j.health.2023.100182
12. Alqahtani MS, Abbas M, Alsabaani A, et al.: The potential impact of COVID-19 virus on the heart and the circulatory system. Infection and Drug Resistance. 2022, 15:1175-1189. 10.2147/idr.s351318
13. Maini E, Venkateswarlu B, Maini B, Marwaha D: Machine learning-based heart disease prediction system for Indian population: An exploratory study done in South India. Medical Journal Armed Forces India. 2021, 77:302-311. 10.1016/j.mjafi.2020.10.013
14. Karthick K, Aruna SK, Samikannu R, Kuppusamy R, Teekaraman Y, Thelkar AR: Implementation of a heart disease risk prediction model using machine learning [RETRACTED]. Computational and Mathematical Methods in Medicine. 2022, 6517716:14. 10.1155/2022/6517716
15. Singh A, Kumar R: Heart disease prediction using machine learning algorithms. 2020 International Conference on Electrical and Electronics Engineering (ICE3), Gorakhpur, India. 2020, 452-457. 10.1109/ICE348803.2020.9122958
16. Khdair H, Dasari NM: Exploring machine learning techniques for coronary heart disease prediction. International Journal of Advanced Computer Science and Applications (IJACSA). 2021, 12:28-36. 10.14569/IJACSA.2021.0120505
17. Kumar Sahoo G, Kanike K, Das SK, Singh P: Machine learning-based heart disease prediction: A study for home personalized care. 2022 IEEE 32nd International Workshop on Machine Learning for Signal Processing (MLSP), Xi'an, China. 2022, 01-06. 10.1109/MLSP55214.2022.9943373
18. Heart Disease Dataset. (2019). Accessed: June 6, 2025: https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset.
19. Number of Cases. (30 June 2022). Accessed: June 6, 2025: https://prsindia.org/covid-19/cases.
20. Blakely L, Reno MJ, Broderick RJ: Evaluation and Comparison of Machine Learning Techniques for Rapid QSTS Simulations. Sandia National Laboratories, Albuquerque, New Mexico; 2018. 10.2172/1734485
21. Kirasich K, Smith T, Sadler B: Random forest vs logistic regression: Binary classification for heterogeneous datasets. SMU Data Science Review. 2018, 1:9.