Original Article

# Prediction of the Fatal Acute Complications of Myocardial Infarction via Machine Learning Algorithms

Reza Ghafari, PharmD[1], Amir Sorayaie Azar, MS[2], Ali Ghafari, MS[3,4], Fatemeh Moradabadi Aghdam, PharmD[1], Morteza Valizadeh, PhD[5], Naser Khalili, MD[6], Shima Hatamkhani, PharmD, PhD[7,8*]

[1]Pharmacy Faculty, Urmia University of Medical Sciences, Urmia, Iran.
[2]Department of Computer Engineering, Urmia University, Urmia, Iran.
[3]Medical Physics and Biomedical Engineering Department, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran.
[4]Research Center for Evidence-Based Medicine, Tabriz University of Medical Sciences, Tabriz, Iran.
[5]Faculty of Electrical and Computer Engineering, Urmia University, Urmia, Iran.
[6]Department of Cardiology, School of Medicine, Urmia University of Medical Sciences, Urmia, Iran.
[7]Experimental and Applied Pharmaceutical Sciences Research Center, Urmia University of Medical Sciences, Urmia, Iran.
[8]Department of Clinical Pharmacy, Urmia University of Medical Sciences, Urmia, Iran.

## Abstract

*Background:* Myocardial infarction (MI) is a major cause of death, particularly during the first year. The avoidance of potentially fatal outcomes requires expeditious preventative steps. Machine learning (ML) is a subfield of artificial intelligence science that detects the underlying patterns of available big data for modeling them. This study aimed to establish an ML model with numerous features to predict the fatal complications of MI during the first 72 hours of hospital admission.

*Methods:* We applied an MI complications database that contains the demographic and clinical records of patients during the 3 days of admission based on 2 output classes: dead due to the known complications of MI and alive. We utilized the recursive feature elimination (RFE) method to apply feature selection. Thus, after applying this method, we reduced the number of features to 50. The performance of 4 common ML classifier algorithms, namely logistic regression, support vector machine, random forest, and extreme gradient boosting (XGBoost), was evaluated using 8 classification metrics (sensitivity, specificity, precision, false-positive rate, false-negative rate, accuracy, F1-score, and AUC).

*Results:* In this study of 1699 patients with confirmed MI, 15.94% experienced fatal complications, and the rest remained alive. The XGBoost model achieved more desirable results based on the accuracy and F1-score metrics and distinguished patients with fatal complications from surviving ones (AUC=78.65%, sensitivity=94.35%, accuracy=91.47%, and F1-score=95.14%). Cardiogenic shock was the most significant feature influencing the prediction of the XGBoost algorithm.

*Conclusion:* XGBoost algorithms can be a promising model for predicting fatal complications following MI.

**This paper should be cited as:** Ghafari R, Sorayaie Azar A, Ghafari A, Moradabadi Aghdam F, Valizadeh M, Khalili N, et al. Prediction of the Fatal Acute Complications of Myocardial Infarction via Machine Learning Algorithms. J Teh Univ Heart Ctr 2023;18(4):278-287.

*Corresponding Author:* **Shima Hatamkhani,** *Assistant Professor of Clinical Pharmacy (Pharmacotherapy), Experimental and Applied Pharmaceutical Sciences Research Center, Urmia University of Medical Sciences, Urmia, Pardis Nazlou, 11 km of Nazlou Road, Urmia, Iran. Tel: +98 44 32234897. Fax: +98 44 32229059. E-mail: hatamkhani.sh@umsu.ac.ir.*

## Introduction

Myocardial infarction (MI) is one of the most challenging issues in modern medicine, with a high rate of first-year mortality.[1] It is the leading cause of death in the United States owing to coronary artery disease, and its incidence remains high in all countries. According to the World Health Organization (WHO), an estimated 17.9 million people died from cardiovascular diseases in 2019, representing 32% of all global deaths, principally because of heart attack and stroke. The American Heart Association 2021 statistics reported that over 3.1% of adults over age 20 (over 880,0000 people) had experienced MI between 2015 and 2018.[2] During the acute and subacute phases of MI, about half of the patients develop complications that aggravate the condition and could lead to death. These complications negatively affect both short- and long-term survival.[3] Despite advances in the management of MI complications, rates are still high. These complications should be clinically recognized and treated expeditiously to prevent mortality and morbidity. The early detection of these complications requires a high clinical index of suspicion.[4]

Machine learning (ML) is a subfield of artificial intelligence (AI) science. Pattern recognition, rule-based reasoning techniques, and modeling big data that teach computers how to favor "good" outputs and reject "bad" ones are some of the core principles of ML methods in medicine.[4]

Diagnosing COVID-19, chronic kidney disease, urinary infections, pulmonary hypertension, influenza, skin lesions, and acromegaly; predicting the risk of severe complications after bariatric surgery;[5] estimating the prevalence of long-term complications in patients with type 2 diabetes,[6] and determining the risk of major adverse events following intravenous lead extraction for cardiac rhythm management[7] are some examples of the application of ML in medicine.[8, 9] Predicting fatal MI complications through prompt and critical preventive measures is crucial since competent experts are almost always unable to forecast all these problems.

Blood tests and electrocardiographic (ECG) signals are among the tools for MI diagnosis. Nonetheless, post-MI blood enzyme levels take time to rise. This time delay may suspend the diagnosis of MI. Very few studies have investigated the use of AI in the field of MI. Most of these studies have used only information related to blood tests and ECG and paid little attention to other clinical information. Some ML models using standard 12-lead ECG signals or different troponin levels in different ages, sexes, and times have been proposed to improve the risk assessment of this disease in patients.[10,11]

The present study aimed to establish an ML model with numerous features to predict mortality following MI complications during the first 72 hours of admission. We used clinical and comprehensive factors that have not been investigated to this extent. Applying such models may help clinicians implement appropriate interventions to attenuate the risk of MI-induced mortality.

## Methods

The study protocol complied with the Declaration of Helsinki and was approved by the Ethics Committee of Urmia University of Medical Sciences (Code: IR.UMSU. REC.1401.128).

We applied an MI complications database in the Krasnoyarsk Interdistrict Clinical Hospital named after I.S. Berzon (Russia) in 1992-1995.[12, 13] It contains the demographic and clinical records of 1699 patients at the time of hospital admission as 111 input features and 12 complications, including fatal complications during the first 72 hours after admission, based on 2 output classes: dead due to known complications of MI and alive, with 7.6% of missing values (Tables 1 & 2).[14] Missing values in the data set are replaced by their mean value. The clinical data were conduction characteristics on ECG at the time of hospital admission; time, type, and extent of MI; laboratory parameters; underlying diseases; patients' signs and symptoms; and administered drugs at the hospital. Complications were considered to be myocardial rupture, atrial fibrillation, supraventricular tachycardia, ventricular tachycardia, ventricular fibrillation, third-degree atrioventricular (AV) block, Dressler syndrome, pulmonary edema, chronic heart failure, MI relapse, and post-infarction angina. Fatal complications were considered to be cardiogenic shock, pulmonary edema, myocardial rupture, congestive heart failure progression, thromboembolism, asystole, and ventricular fibrillation. As an attempt to address the class imbalance, samples with fatal complications in Class I were weighted 7 times higher than those in Class 0.

Recursive feature elimination (RFE) is a methodology employed within ML to meticulously curate a subset of the most pivotal features of an initially extensive array. The process entails training a model employing the complete feature set and assigning priority rankings to each feature based on their significance. Subsequent stages involve an iterative elimination of the least crucial features while concurrently retraining the model. This sequential process persists until a specified quantity of features is attained, or a predetermined performance threshold is achieved. RFE significantly enhances model performance, diminishes the risk of overfitting, and amplifies model explicability by concentrating solely on the most pertinent features germane to a given problem.[15] We applied the RFE method and reduced the number of features to 50.

Table 1. Input Features of the Patients at the Time of Hospital Admission

| Demographic Characteristics | Age, sex, and obesity |
| --- | --- |
| Conduction Characteristics on ECG at the Time of Hospital Admission | First-degree AV block, type 1 second-degree AV block (Mobitz I/Wenckebach), type 2 second-degree AV block (Mobitz II/Hay), third-degree AV block, LBBB (anterior branch), LBBB (posterior branch), incomplete/complete LBBB, incomplete/ complete RBBB, ECG rhythm at the time of admission to the hospital (sinus; heart rate=60-90 bpm), atrial fibrillation rhythm, atrial rhythm, idioventricular rhythm, sinus rhythm with a heart rate above 90 bpm (tachycardia), sinus rhythm with a heart rate below 60 bpm (bradycardia), premature atrial contractions, frequent premature atrial contractions, premature ven-tricular contractions, frequent premature ventricular contractions, paroxysms of atrial fibrillation, persis-tent form of atrial fibrillation, paroxysms of supraventricular tachycardia, paroxysms of ventricular tachycardia, ventricular fibrillation, and sinoatrial block |
| Type and Extent of MI | Anterior MI (left ventricular) (ECG changes in leads V1–V4), lateral MI (left ventricular) (ECG changes in leads V5–V6, I, and AVL), inferior MI (left ventricular) (ECG changes in leads III, AVF, and II), posterior MI (left ventricular) (ECG changes in leads V7–V9 and reciprocity changes in leads V1–V3), and right ventricular MI |
| Laboratory Parameters | ALT, AST, CPK, WBC, ESR, hypokalemia (< 4 mmol/l), serum potassium, hypernatremia (>150 mmol/l), and serum sodium |
| Underlying Diseases | Underlying cardiovascular diseases: chronic heart failure (including the functional class of angina pectoris in the last year), coronary heart disease in recent weeks or days before admission to the hospital, essential hypertension, symptomatic hypertension (including the duration of arterial hypertension), and premature atrial contractions<br>Other underlying diseases: diabetes mellitus, obesity, thyrotoxicosis, chronic bronchitis, obstructive chronic bronchitis, bronchial asthma, and pulmonary tuberculosis |
| Signs and Symptoms | History of cardiovascular signs and symptoms: exertional angina pectoris, incomplete or complete LBBB, incomplete or complete RBBB, premature ventricular contractions, paroxysms of atrial fibrillation, persistent form of atrial fibrillation, ventricular fibrillation, ventricular paroxysmal tachycardia, first-degree AV block, third-degree AV block, LBBB (anterior branch), quantity of MI in the anamnesis<br>Signs and symptoms at emergency admission: systolic and diastolic blood pressure<br>Signs and symptoms at ICU admission: systolic and diastolic blood pressure, pulmonary edema, cardiogenic shock, paroxysms of atrial fibrillation, paroxysms of supraventricular tachycardia, paroxysms of ventricular tachycardia, ventricular fibrillation, pain relapse in the first hours[§] |
| Administered Drugs | At the emergency department: opioids, NSAIDs, and lidocaine<br>At the ICU: liquid nitrates, lidocaine, β-blockers, calcium channel blockers, anticoagulants (heparin), acetylsalicylic acid, ticlopidine, pentoxifylline, opioids[§], and NSAIDs[§]<br>Fibrinolytic therapy: with celiasum (750k IU)/celiasum (1m IU)/celiasum (3m IU)/ streptase/celiasum (500k IU)/ celiasum (250k IU)/streptodecase (1.5m IU) |
| Other Parameters | Time elapsed from the beginning of the attack of coronary heart disease to hospital admission and ob-serving arrhythmias in the anamnesis |

[§]Also determined at 24, 48, and 72 hours after admission

MI, Myocardial infarction; ECG, Electrocardiogram; AV, Atrioventricular; NSAIDs, Non-steroidal anti-inflammatory drugs; LBBB, Left bundle branch block; RBBB, Right bundle branch block; ALT, Alanine transaminase; AST, Aspartate aminotransferase; CPK, Creatine phosphokinase; WBC, White blood cell; ESR, Erythrocyte sedimentation rate

Table 2. Patients' Outcomes as Output Features

| | n (%) |
| --- | --- |
| Alive | 1428 (84.04) |
| Dead due to Fatal Complications | |
| Cardiogenic shock | 110 (6.47) |
| Pulmonary edema | 18 (1.05) |
| Myocardial rupture | 54 (3.17) |
| Progress of congestive heart fail-ure | 23 (1.35) |
| Thromboembolism | 12 (0.70) |
| Asystole | 27 (1.58) |
| Ventricular fibrillation | 27 (1.58) |

AI enables systems to learn automatically. The principal focus of ML algorithms is to develop programs that access data and use them for learning. The process of ML algorithms begins by observing and working with data to find the desired pattern and make informed decisions based on the provided samples. The main goal of ML methods is to extend learning beyond trained examples.[5]

ML methods are divided into 3 groups: supervised learning, unsupervised learning, and reinforcement learning.[5] In supervised learning, a machine is trained using labeled data. In other words, in this type of learning, the data are already labeled with correct answers. In unsupervised learning, the machine is trained using unlabeled data. In this method, the learning algorithm is not told what the data represent. In reinforcement learning, similar to unsupervised learning, the data used for learning are not labeled. In this method, when a question is asked for data, the result is graded.

In the current study, the supervised learning approach was considered for classification. Four common ML classifier algorithms were used: logistic regression, support vector machine (SVM), random forest, and extreme gradient boosting (XGBoost). Each of these classifiers applies different classification methods.

Logistic regression: Logistic regression is a very simple, basic, and useful algorithm for classification. It uses a linear equation with independent predictors to predict a value. The predicted value can be between negative infinity and positive infinity. In this study, the output of the algorithm was considered to be the class variable (alive or dead). Therefore, the output of the linear equation was transferred to a range of 0 to 1. To transfer the predicted value between 0 and 1, we utilized the sigmoid function.[5]

$$z = \theta_0 + \theta_1.x_1 + \ldots + \theta_n.x_n \tag{1}$$

$$h = g(z) = \frac{1}{1+e^{-z}} \tag{2}$$

$$Cost(h_\theta(x), y) = \{-\log\log(h_\theta(x)),$$
$$y = 1 - \log\log(1 - h_\theta(x)), \ y = 0 \tag{3}$$

The output of Eq. (1) was given to the function g, Eq. (2), which returned the transfer value in the range of 0 to 1. The sigmoid function became asymptotic to y=1 for large positive values of x and became asymptotic to y=0 for large negative values of x. For the prediction of class values, a logarithmic loss function was used to calculate the cost of misclassification, Eq. (3), where x was the input vector, and $\theta$ was the coefficient vector.

SVM: The SVM algorithm is one of the most powerful ML models. This algorithm can be used for linear or nonlinear classification and regression problems. SVM is one of the most popular ML models, especially since this algorithm is suitable for classifying small or medium-sized data sets. SVM creates a boundary between classes called "a hyperplane". The goal of this algorithm is to maximize the margin between classes [5] by minimizing Eq. (4)

$$C\|w\|^2 + \frac{1}{n}\sum_{i=1}^{n} max(0, 1 - y_i(w.x_i - b)) \tag{4}$$

where n is the number of the samples and C is the balance value between the extent of decision boundaries. SVM can draw nonlinear boundaries using kernel tricks. As a result, the data samples are extended into the feature space with higher dimensions so that they can be linearly separated.[5]

Random forest: The random forest algorithm is an ensemble algorithm that uses decision trees for its simple algorithms. This algorithm, one of the most common ML algorithms, is used for both classification and regression problems. Since a decision tree algorithm can easily perform classification operations on the data, several decision trees are employed in the random forest algorithm. A set of decision trees together produces a forest, and this forest can make better decisions than a single tree. Finally, the random forest algorithm can select the class with the most votes among the decision trees by voting and place it as the final class for the classification problem.[5]

XGBoost: XGBoost is a family of tree-structured algorithms. Tree-based models are frequently implemented in this method, considered a collective classification method. In other models, the models are merged, so that the final operation takes place. However, this algorithm takes a smart approach. In this algorithm, instead of training all the models separately from each other, the boosting process trains the models one after the other. Each new model is trained with the aim of correcting the errors caused by previous models. Models are added sequentially until there is no further development possible. The advantage of this iterative method is that the added models seek to correct the mistakes made by other models. In the standard ensemble classification method, where models are trained individually, all models may make the same mistakes. Gradient boosting refers to a method in which new models are trained with the aim of predicting the residuals of previous models.[16]

In this study, 8 classification metrics (sensitivity, specificity, precision, false-positive rate, false-negative rate, accuracy, F1-score, and AUC) were chosen to evaluate the performance of ML models in predicting fatal MI complications according to the database of patients after being hospitalized. These 8 metrics were applied to compare the performances of the models (positive instances: patients who died due to fatal complications of MI and negative instances: living patients).

Accuracy: Accuracy can be inferred as how close the measured value is to the desired value[9, 16]:

$$Accuracy = \frac{True\ positive + True\ negative}{True\ positive + False\ positive + False\ negative + True\ negative} =$$

$$\frac{Correct\ lethal\ complication\ detection}{Total\ complication\ detection} \tag{5}$$

Sensitivity or Recall: It is the fraction of positive responses that have been correctly identified[9, 16]:

$$Sensitivity = \frac{True\ positive}{True\ positive + False\ negative} \tag{6}$$

Specificity: The ability of the test to correctly diagnose patients without the listed disease or condition[9, 16]:

$$Specificity = \frac{True\ negative}{True\ negative + False\ positive} \tag{7}$$

Precision (PPV): It indicates how close the measurement values are after consecutive value measurements [9, 16]:

$$Precision = \frac{True\ positive}{True\ positive + False\ positive} \qquad (8)$$

False-Positive Rate: it corresponds to the proportion of negative data points that are mistakenly considered as positive, with respect to all negative data points [9, 16]:

$$False\ Positive\ Rate = \frac{False\ positive}{True\ negative + False\ positive} \qquad (9)$$

False-Negative Rate: signifies how many positive class samples your model predicted incorrectly [9, 16]:

$$False - Negative\ Rate = \frac{False\ negative}{True\ negative + False\ negative} \qquad (10)$$

F-Score (F1-Score): It is used to combine the 2 criteria of specificity and accuracy, and it is a measure of the accuracy of the test [9, 16]:

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \qquad (11)$$

AUC: It is the area under the receiver operating characteristic (ROC) curve that shows the overall performance of the models. [9,16]

In this study, ML explanatory and interpretable methods were used to identify the salient features influencing the decision-making of ML models. Explain Like I'm 5 (ELI5) is an interpretability method used to explain the predictions of ML models and their interpretation. It is an easy method in that it can identify salient features and support agnostic methods. [18]

The effects of these 4 models (in this method, taking into account the classification results, different clauses are adopted in the final decision) and the effects of selecting and extracting more discriminant features confer more desirable results. Therefore, finding those crucial features and reporting them can benefit other researchers.

The final decision about the effects of models was made based on the accuracy of the different classification methods. Additionally, finding those optimal features can be advantageous to other researchers.

After the implementation of the mentioned models, the 5-fold cross-validation method (20% of the data in each round for testing the algorithm and validating and the rest for training the program) was used to check the efficiency of the models and the effects of using more appropriate features. The k-fold cross-validation method divided the data set into K number of subsets. Then, each evaluation partition was considered the test data set, and the rest was considered the training set. This approach was repeated for K times, so that each partition was utilized as a test set once. Next, the prediction error was calculated for each partition, and their average was considered the total cross-validation error. At each stage, the test data were given to the machine, and the output was compared with the desired values of the classes that were available. Further, performance evaluation criteria, composed of accuracy, sensitivity, specificity, F1-score, precision, negative predictive value, false-positive rate, and false-negative rate, were calculated, and the quantitative evaluation was carried out based on them.

The data were analyzed using NumPy, SciPy, Matplotlib, Pandas, and Scikit libraries in Python programming software, version 3.7.

## Results

The present study evaluated 1699 patients with confirmed MI (62.65% males and 37.35% females) with a mean age of 61.86±11.26 years. Within 3 days of admission, 15.94% of the study population experienced fatal complications, and the rest remained alive (Table 2). Among all the patients, 663 (39%) did not experience any complications.

Table 3 shows the diagnostic performance of the applied models. Different classification metrics provide valuable but different insights into the performance of the classifiers. Confusion matrices are presented in Figure 1. The XGBoost model achieved better results based on the accuracy and F1-score metrics and distinguished patients with fatal complications from survived ones (AUC=78.65%, sensitivity=94.35%, accuracy=91.47%, and F1-score=95.14%). Via the ELI5 method, the most decisive features influencing the prediction of the XGBoost algorithm are shown in Figure 2. Additionally, Table 4 presents the distribution of the most influential features regarding the target class.

Table 3. Performance of the Applied Models at the Time of Admission

| Algorithm | Sensitivity | Specificity | Preci-sion | False-Positive Rate | False-Negative Rate | Accuracy | F1-Score | AUC (%) |
|---|---|---|---|---|---|---|---|---|
| Logistic regression | 92.83 | 66.6 | 96.28 | 33.33* | 7.17 | 90.29 | 64.53 | 73.14 |
| Support vector ma-chine | 89.54 | 80.00 | 98.98 | 20.00 | 10.46* | 89.12 | 94.02 | 77.12 |
| Random forest | 91.28 | 84.21* | 98.99* | 15.79 | 8.72 | 90.88 | 94.98 | 67.68 |
| XGBoost | 94.35* | 69.23 | 95.95 | 30.77 | 5.65 | 91.47* | 95.14* | 78.65* |

*The favorable metrics among all models
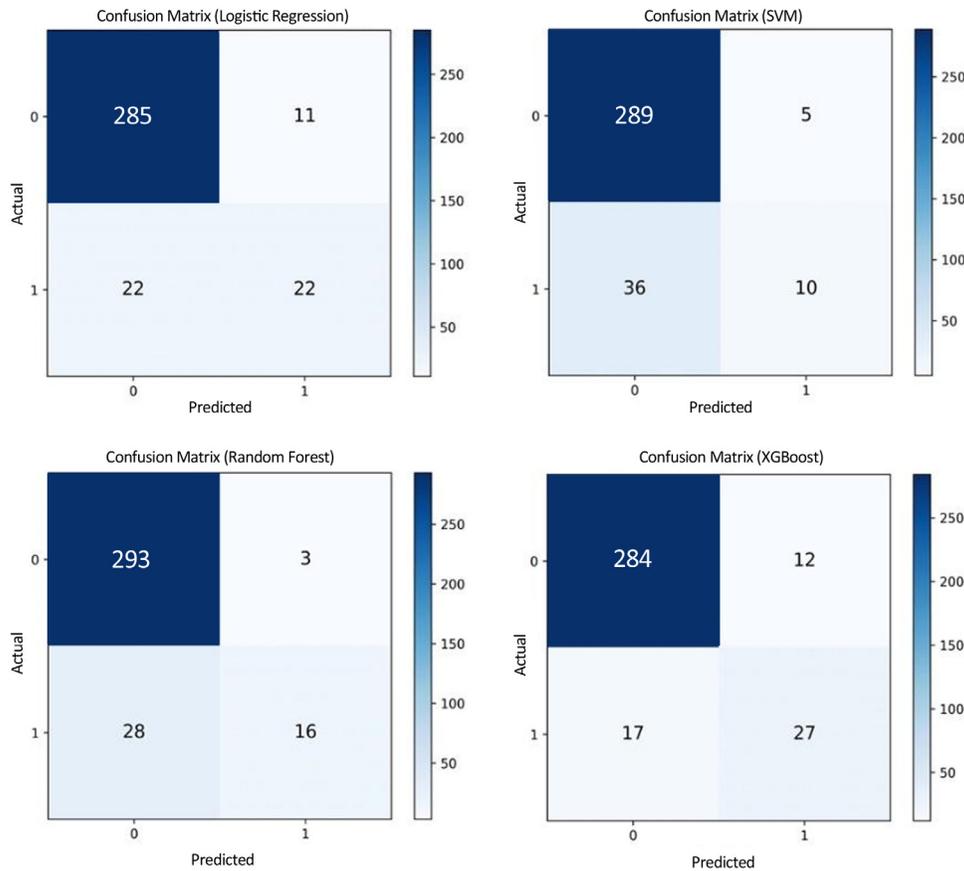
Figure 1. The images present the confusion matrices of the classifiers at the time of admission.

Table 4. Distribution of the Samples Considering the Most Notable Features and Target Class

| Feature | States | Outcome | |
|---|---|---|---|
| | | Alive | Dead |
| Cardiogenic shock at the time of admission to the ICU (K_SH_POST) | | | |
| Yes | | 3 | 43 |
| No | | 1414 | 224 |
| Missing value (N/A)* | | 12 | 3 |
| Frequent premature ventricular contractions on the electrocardiogram at the time of admission to the hospital (n_r_ecg_p_04) | | | |
| Yes | | 50 | 19 |
| No | | 1272 | 243 |
| Missing value (N/A) | | 107 | 8 |
| Third-degree atrioventricular block on the electrocardiogram at the time of admission to the hospital (n_p_ecg_p_06) | | | |
| Yes | | 12 | 15 |
| No | | 1308 | 249 |
| Missing value (N/A) | | 109 | 6 |
| Presence of chronic heart failure in the anamnesis (ZSN_A) | | | |
| I stage | | 98 | 5 |
| IIÐ-R stage | | 11 | 16 |
| IIÐ-L stage | | 18 | 11 |
| IIB (IIÐ-R & IIÐ-L) stage | | 6 | 13 |
| No | | 1292 | 175 |
| Missing value (N/A) | | 4 | 50 |
| Pulmonary edema at the time of admission to the ICU (O_L_POST) | | | |
| Yes | | 67 | 43 |
| No | | 1351 | 226 |
| Missing value (N/A) | | 11 | 1 |

*The missing values are filled with the mean value of that specific feature, although not considered by the machine learning models for training.
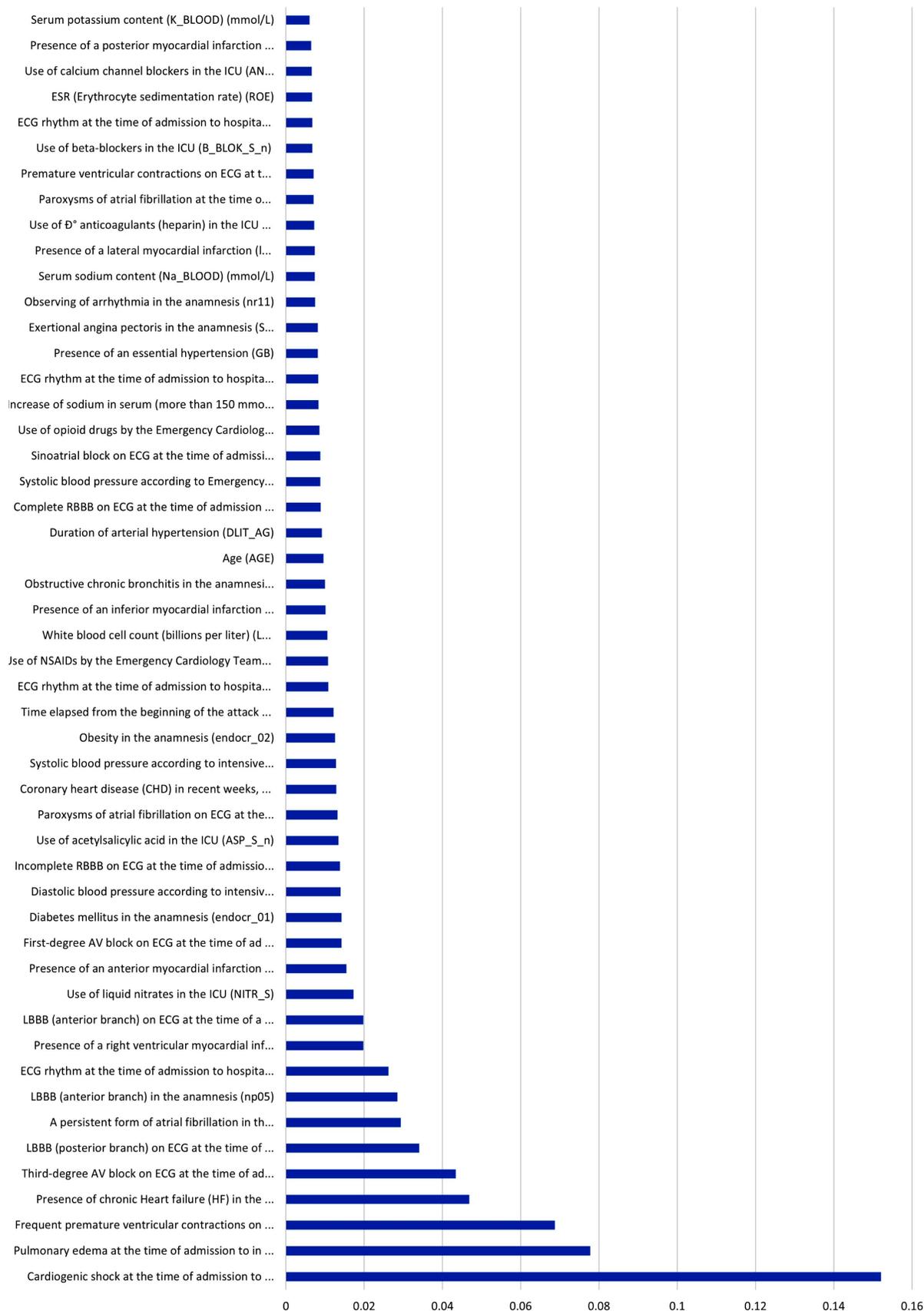
Figure 2. The image presents the most notable features of the XGBoost model.

## Discussion

AI and ML can support medical professionals with data-driven processes in the healthcare system. Although, as an emerging technology, there have always been running debates on the pros and cons of its application in healthcare systems, the numerous profits of ML outweigh a few possible losses. It has the advantages of flexibility and scalability compared with traditional biostatistical methods, which makes it deployable for many tasks, such as risk stratification, diagnosis and classification, and survival predictions using various data types. Medical application of ML requires data pre-processing, model training, and system refinement regarding the actual clinical problem.[17] Compared with risk assessment guidelines that require the manual calculation of scores, ML-based prediction of disease outcomes can be utilized to save time and improve prediction accuracy.

In this study, we evaluated 4 common ML algorithms in terms of 8 metrics using bulk data from 1699 patients. Because of the importance of the initial time of patient admission, we considered the data set at the time of hospital admission to identify acute fatal complications. A multitude of variables might determine the rate of complications following MI. Knowing how to use this knowledge to forecast probable outcomes is crucial. As observed from this data set, no single element can provide all the necessary predictability information. However, including all prominent features improves forecasts. In the literature, the improvement in model performance seems very promising and revolutionary, but that is not the case in the business world. Nevertheless, improving measurement metrics (eg, increasing AUC, minimizing log loss, and improving recall and specificity by decreasing the false positive and the false negative) can augment the model's performance.

ML has been used for predicting mortality in various clinical settings, including COVID-19 and heart failure, with accuracy rates of 90% and 80%, respectively.[18] ML also showed promising results in the diagnosis of chronic MI using non-enhanced cine magnetic resonance imaging, with an AUC of 94%.[19]

The results of the current study showed that the proposed models accurately predicted the acute fatal complications of MI. The XGBoost classifier achieved more desirable results based on the accuracy and F1-score metrics. We compared 4 standard common models to help other researchers avoid redundant analysis and select the best model.

No study in the literature has employed ML to detect the fatal complications of MI. In 2017, Mansoor et al[20] assessed 2 ML approaches (logistic regression and random forest) to predict the possible all-cause in-hospital mortality risks of third-degree AV block at the time of admission in 9637 women hospitalized with a diagnosis of ST-elevation MI using the United States National Inpatient Sample Data, collected in 2011 and 2013. The reported results were a mean accuracy rate of 0.88 for both data-logistic regression models and a mean accuracy rate of 0.88 for random forest models. Since the data set only included women, the findings can be applied only to female patients.

Patients who recover from an MI encounter an elevated risk of subsequent cardiovascular events, including an increase in mortality. Acute MI can be complicated by a variety of pathophysiologic causes categorized as mechanical, arrhythmic, ischemic, inflammatory, and embolic complications.[21] Patients may undergo a risk assessment to identify those at high risk for both short- and long-term unfavorable consequences. Numerous risk assessment tools have been developed, such as the thrombolysis in myocardial infarction (TIMI) risk score and the Global Registry of Acute Coronary Events (GRACE) risk model, which apply a limited set of variables.[22] An invasive electrophysiology study (EPS) is performed only under very particular circumstances for risk classification.[23] As these tools are derived from clinical trial data with a restricted population, they may not consider real features and all aspects required for predicting cardiac outcomes. These pitfalls may be fulfilled to some extent by ML. ML can uncover the complex effect of each variable and relationships between variables presented as simple ones in the database.

Based on our results, the most prominent features identified in the prediction made by the best algorithm of this study (the XGboost classifier) were cardiogenic shock and a history of complete left bundle branch block. These characteristics could be incorporated into risk score assessments in practice or hospital information systems for use by physicians at the bedside of patients. On the other hand, these models can be improved by using a database specific to each country, region, and race. This notion is supported by previous clinical findings in the literature that introduce cardiogenic shock as the leading cause of in-hospital death in patients with acute MI.[24] Cardiogenic shock is a consequence of heart dysfunction, presenting as myocardial tissue hypoxia and necrosis following systemic hypoperfusion. About 10% of patients experience cardiogenic shock right after an acute MI, with the shock linked to roughly 40% of 30-day death rates.[25] Patients with acute MI may arrive at the hospital in cardiogenic shock, or the condition may develop later on.[26] Despite advances in treatment, the only therapy shown to reduce mortality significantly in cardiogenic shock in a randomized trial was the emergency revascularization of the infarct-related artery.[27, 28]

This study has some limitations. The applied database was set up between 1990 and 1998, when reperfusion strategies had not been introduced yet. Future research should compare ML mortality prediction with other risk assessment tools.

Our study presents an initial step in testing the capacity for ML predictive models suitable for clinical decision-making and risk avoidance vis-à-vis patients presenting to the hospital with acute MI.

## Conclusion

The findings of this study offer a novel viewpoint on reducing the fatal outcomes of MI. XGBoost algorithms could be a promising model for predicting fatal complications following MI. The minimal financial burden of employing such models makes it easier for clinicians to reduce patient morbidity and mortality.

## Acknowledgments

## References

1. Reeder GS. Identification and treatment of complications of myocardial infarction. Mayo Clin Proc. 1995;70:880-884.
2. Virani SS, Alonso A, Aparicio HJ, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, Chamberlain AM, Cheng S, Delling FN, Elkind MSV, Evenson KR, Ferguson JF, Gupta DK, Khan SS, Kissela BM, Knutson KL, Lee CD, Lewis TT, Liu J, Loop MS, Lutsey PL, Ma J, Mackey J, Martin SS, Matchar DB, Mussolino ME, Navaneethan SD, Perak AM, Roth GA, Samad Z, Satou GM, Schroeder EB, Shah SH, Shay CM, Stokes A, VanWagner LB, Wang NY, Tsao CW; American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2021 Update: A Report From the American Heart Association. Circulation 2021;143:e254-e743.
3. Kutty RS, Jones N, Moorjani N. Mechanical complications of acute myocardial infarction. Cardiol Clin 2013;31:519-531, vii-viii.
4. Montrief T, Davis WT, Koyfman A, Long B. Mechanical, inflammatory, and embolic complications of myocardial infarction: An emergency medicine review. Am J Emerg Med 2019;37:1175-1183.
5. Cao Y, Fang X, Ottosson J, Näslund E, Stenberg E. A Comparative Study of Machine Learning Algorithms in Predicting Severe Complications after Bariatric Surgery. J Clin Med 2019;8:668.
6. Munna MTA, MM Alam, Allayear SM, K Sarker, Ara SJF. Prediction Model for Prevalence of Type-2 Diabetes Complications with ANN Approach Combining with K-Fold Cross Validation and K-Means Clustering. In: Arai K, Bhatia R, eds. Advances in Information and Communication. FICC 2019. Lecture Notes in Networks and Systems, vol 69. Cham: Springer; 2020. p. 1031–1045.
7. Mehta VS, O"brien H, Elliott MK, Sidhu BS, Gould J, Razavi R, Niederer S, Rinaldi CA. Machine learning based major complication prediction for patients undergoing transvenous lead extraction. EP Europace 2021;23(Supplement_3):euab116.511
8. Bhavsar KA, Abugabah A, Singla J, AlZubi AA, Bashir AK, et al. A comprehensive review on medical diagnosis using machine learning. Computers, Materials & Continua 2021;67:1997–2014.
9. Sorayaie Azar A, Ghafari A, Ostadi M, Babaei Rikan S, Ghafari R, Farajpouri M, Sheikhzadeh P. Covidense: Providing a suitable solution for diagnosing Covid-19 lung infection based on deep learning from chest X-ray images of patients. Frontiers in Biomedical Technologies 2021;8:131-142.
10. Than MP, Pickering JW, Sandoval Y, Shah ASV, Tsanas A, Apple FS, Blankenberg S, Cullen L, Mueller C, Neumann JT, Twerenbold R, Westermann D, Beshiri A, Mills NL; MI3 Collaborative. Machine Learning to Predict the Likelihood of Acute Myocardial Infarction. Circulation 2019;140:899-909.
11. Baloglu UB, Talo M, Yildirim O, Tan RS, Acharya UR. Classification of myocardial infarction with multi-lead ECG signals and deep CNN. Pattern Recognition Letters 2019;122:23-30.
12. Golovenkin SE, Gorban A, Mirkes E, Shulman VA, Rossiev DA, Shesternya PA, et al. Myocardial infarction complications Database. University of Leicester. Dataset 2020. https://doi.org/10.25392/leicester.data.12045261.v3 (01 August 2022).
13. Golovenkin SE, Gorban AN, Mirkes EM, Shulman VA, Rossiev DA, Shesternya PA, et al. Complications of myocardial infarction: a database for testing recognition and prediction systems. https://www.stat.rice.edu/~scottdw/stat541/PROJECTS/uci-data/MI-Complications/MI-desc.pdf. (01 August 2022)
14. Golovenkin SE, Bac J, Chervov A, Mirkes EM, Orlova YV, Barillot E, Gorban AN, Zinovyev A. Trajectories, bifurcations, and pseudo-time in large clinical datasets: applications to myocardial infarction and diabetes data. Gigascience 2020;9:giaa128.
15. Darst BF, Malecki KC, Engelman CD. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. BMC Genet 2018;19(Suppl 1):65.
16. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. Indian J Ophthalmol 2008;56:45-50.
17. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. Lancet Oncol 2019;20:e262-e273.
18. MAM Hasan, Shin J, Das U, Yakin Srizon A. Identifying Prognostic Features for Predicting Heart Failure by Using Machine Learning Algorithm. ICBET '21: Proceedings of the 2021 11th International Conference on Biomedical Engineering and Technology March 2021; p. 40-46.
19. Zhang N, Yang G, Gao Z, Xu C, Zhang Y, Shi R, Keegan J, Xu L, Zhang H, Fan Z, Firmin D. Deep Learning for Diagnosis of Chronic Myocardial Infarction on Nonenhanced Cardiac Cine MRI. Radiology 2019;291:606-617.
20. Mansoor H, Elgendy IY, Segal R, Bavry AA, Bian J. Risk prediction model for in-hospital mortality in women with ST-elevation myocardial infarction: A machine learning approach. Heart Lung 2017;46:405-411.
21. Mullasari AS, Balaji P, Khando T. Managing complications in acute myocardial infarction. J Assoc Physicians India. 2011;59 Suppl:43-48.
22. Chen YH, Huang SS, Lin SJ. TIMI and GRACE Risk Scores Predict Both Short-Term and Long-Term Outcomes in Chinese Patients with Acute Myocardial Infarction. Acta Cardiol Sin 2018;34:4-12.
23. Thomas KE, Josephson ME. The role of electrophysiology study in risk stratification of sudden cardiac death. Prog Cardiovasc Dis 2008;51:97-105.
24. Samsky MD, Morrow DA, Proudfoot AG, Hochman JS, Thiele H, Rao SV. Cardiogenic Shock After Acute Myocardial Infarction: A Review. JAMA 2021;326:1840-1850.
25. Harjola VP, Lassus J, Sionis A, Køber L, Tarvasmäki T, Spinar J, Parissis J, Banaszewski M, Silva-Cardoso J, Carubelli V, Di Somma S, Tolppanen H, Zeymer U, Thiele H, Nieminen MS, Mebazaa A; CardShock Study Investigators; GREAT network. Clinical picture and risk prediction of short-term mortality in cardiogenic shock. Eur J Heart Fail 2015;17:501-509.
26. Webb JG, Sleeper LA, Buller CE, Boland J, Palazzo A, Buller E, White HD, Hochman JS. Implications of the timing of onset of cardiogenic shock after acute myocardial infarction: a report from the SHOCK Trial Registry. SHould we emergently revascularize Occluded Coronaries for cardiogenic shocK? J Am Coll Cardiol 2000;36(3 Suppl A):1084-1090.
27. Collet JP, Thiele H, Barbato E, Barthélémy O, Bauersachs J, Bhatt DL, Dendale P, Dorobantu M, Edvardsen T, Folliguet T, Gale CP, Gilard M, Jobs A, Jüni P, Lambrinou E, Lewis BS, Mehilli J, Meliga E, Merkely B, Mueller C, Roffi M, Rutten FH, Sibbing D, Siontis GCM; ESC Scientific Document Group. 2020 ESC

Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation. Eur Heart J 2021;42:1289-1367.

28. Amsterdam EA, Wenger NK, Brindis RG, Casey DE Jr, Ganiats TG, Holmes DR Jr, Jaffe AS, Jneid H, Kelly RF, Kontos MC, Levine GN, Liebson PR, Mukherjee D, Peterson ED, Sabatine MS, Smalling RW, Zieman SJ; ACC/AHA Task Force Members; Society for Cardiovascular Angiography and Interventions and the Society of Thoracic Surgeons. 2014 AHA/ACC guideline for the management of patients with non-ST-elevation acute coronary syndromes: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation 2014;130:2354-2394.