Original Article Open Access

Developing in hospital mortality prediction model tools for patients with acute myocardial infarction (STEMI) using Yazd cardiovascular disease registry, YCDR data

Seyedeh Mahdieh Namayandeh[1], Mohsen Mohammadi[2]*, Masoud Mirzaei[3]*, Mohsen Askarishahi[4], Hamidreza Dehghani[5], Seyed Mahmoud Sadr Bafghi[6]

1- Afshar Research Development Center, Center for Healthcare Data Modeling, Departments of Biostatistics and Epidemiology, School of Public Health, Afshar Hospital, Shahid Sadoughi University of Medical Sciences, Yazd, Iran
2- Center for Healthcare Data Modeling, Departments of Biostatistics and Epidemiology, School of Public Health, Shahid Sadoughi University of Medical Sciences, Yazd, Iran
3- Yazd Cardiovascular Research Center, Non-Communicable Diseases Research Institute, Shahid Sadoughi University of Medical Sciences, Yazd, Iran
4- Center for Healthcare Data Modeling, Departments of Biostatistics and Epidemiology, School of Public Health, Shahid Sadoughi University of Medical Sciences, Yazd, Iran
5- Health Assessment Technology Center, Shahid Sadoughi University of Medical Sciences, Yazd, Iran
6- Yazd Cardiovascular Research Center, Department of Cardiology, Shahid Sadoughi University of Medical Sciences, Yazd, Iran

Correspondence:

Mohsen Mohammadi;
Center for Healthcare Data Modeling, Departments of Biostatistics and Epidemiology, School of Public Health, Shahid Sadoughi University of Medical Sciences, Yazd, Iran;
Email: mohsen2019@ymail.com

Masoud Mirzaei;
Yazd Cardiovascular Research Center, Non-Communicable Diseases Research Institute, Shahid Sadoughi University of Medical Sciences, Yazd, Iran;
Email:
masoud_mirzaie@hotmail.com

Received: 2024-12-05
Accepted: 2025-06-07

How to cite this article:
Namayandeh SM, Mohammadi M, Mirzaei M, Askarishahi M, Dehghani H, Sadr Bafghi SM. **Developing in hospital mortality prediction model tools for patients with acute myocardial infarction (STEMI) using Yazd cardiovascular disease registry, YCDR data.** ARYA Atheroscler. 2025; 21(5): 21-33.

DOI:
https://doi.org/10.48305/arya.2025.43144.3004

Abstract

BACKGROUND: An acute ST-elevation myocardial infarction (STEMI) is a medical event characterized by transmural myocardial ischemia that leads to myocardial injury or necrosis. This study was undertaken to develop, evaluate, and compare models for assessing the risk of hospital mortality in patients with acute myocardial infarction.

METHODS: The study made use of data from the Yazd Cardiovascular Diseases Registry (YCDR), which is a cohort study of inpatient records of ischemic heart disease in Yazd province, Iran. A total of 1,861 patients who had experienced a STEMI were included in the analysis. Decision tree analysis was conducted using R software with the *rpart* package. Additionally, to analyze the data and adjust for any confounding variables, logistic regression was performed using the *glm2* package. To compare the effectiveness of the two models, accuracy measures were used, and the Receiver Operating Characteristic (ROC) curve was applied.

RESULTS: In this study, three clinical, laboratory, and clinical-laboratory models were created. In a clinical-laboratory model, variables such as blood sugar (BS), triglycerides, HDL cholesterol, peak myocardial band (MBpick), CVA history, and low ejection fraction (EF) were found to increase the risk of in-hospital mortality in patients with ST-elevation myocardial infarction (STEMI). Conversely, higher levels of hemoglobin, low HDL-C, and previous myocardial infarction (MI) were associated with a protective effect against the risk of in-hospital mortality from acute myocardial infarction.

The performance of the models in terms of Receiver Operating Characteristic (ROC) curve was 86.5%, 79.5%, and 90.2% for logistic regression model in three different models: clinical, laboratory, and combined clinical-laboratory. The accuracy of these models was calculated to be 88.3%, 81.3%, and 93%, respectively. Important variables influencing the prediction of in-hospital mortality in STEMI patients included Killip class, triglycerides, blood sugar, creatinine levels, the need for treatment due to ventricular fibrillation or ventricular tachycardia (VF/VT), age, and hemoglobin (HB). In the ROC curve analysis of the decision tree algorithm across the clinical, laboratory, and combined clinical-laboratory models, the performance levels were 74.6%, 69.8%, and 81.7%, respectively. The accuracy of the decision tree was 93.0%, 92.5%, and 95.8%.

CONCLUSION: The findings of this study indicated that the decision tree algorithm had higher accuracy across all three models: clinical, laboratory, and combined clinical-laboratory compared to logistic regression. However, logistic regression showed higher sensitivity and better ROC curve performance than the decision tree algorithm.

Keywords: STEMI; Acute Myocardial Infarction; Ischemia; Prognostic Factors

Introduction

An acute ST-elevation myocardial infarction (STEMI) is a medical event characterized by transmural myocardial ischemia that leads to myocardial injury or necrosis[1]. This condition necessitates immediate hospitalization due to the urgent need for medical care. Upon arrival at the hospital, patients with suspected STEMI are given priority based on triage and are typically admitted to either the cardiac care unit or a specialized care department. Acute coronary syndromes are classified into two categories based on the presence or absence of ST-elevation myocardial infarction: STEMI and non-STEMI[2]. Non-STEMI cases tend to be less severe and are often associated with a less urgent approach to treatment[2].

Cardiovascular diseases are recognized as the leading cause of death and disability-adjusted life years lost worldwide[1]. While the age-adjusted mortality rate from chronic diseases has seen a decline in several developed countries in recent years, the past decade has witnessed a significant rise in the incidence of cardiovascular diseases in low- and middle-income countries[3,4].

It has been demonstrated that concentrating on preventing premature death by thoroughly understanding the inflammatory response, and by preventing or reversing the failure of vital organs at an early stage, may result in an increased life expectancy for patients suffering from acute myocardial infarction[5]. The importance of accurate prognosis is crucial in medical conditions, as an error in its determination can result in the selection of inappropriate treatment, potentially leading to irreversible consequences. The risk score, which is used to predict outcomes, is composed of multiple factors. Each factor is assigned a specific weight that contributes to the overall score. This score is then used to estimate the likelihood of patient mortality or the need for hospital readmission[6]. These models are instrumental in assisting physicians to make more precise prognostic assessments regarding a patient's future health outcomes. Beyond their direct benefits to patient care, these models also serve a variety of other functions. They are valuable tools for evaluating the effectiveness of medical interventions, comparing different treatment options, forecasting the impact of public health policies, assessing health technologies, and analyzing the efficacy of treatments and laboratory tests. Furthermore, they provide a robust foundation for conducting new research. One well-established fact regarding patients suffering from acute myocardial infarction is the ability to identify individuals at a high risk of mortality and to predict potential outcomes based on their characteristics. Consequently, by developing models to pinpoint and forecast those at elevated risk for such conditions, we can significantly reduce mortality through timely and effective therapeutic interventions[7].

Research indicates that decision trees are effective tools for the early and precise diagnosis of heart attacks. The significance of prognosis in these circumstances cannot be overstated, as an error in determining the prognosis and consequently selecting an inappropriate treatment can have dire consequences. The risk score, which is used to predict patient outcomes such as mortality or readmission, is composed of various factors, each with its own assigned weight[8]. This analysis categorizes the primary group of individuals and continues to do so for subgroups. Given the absence of research in the country to develop, evaluate, and benchmark a model for assessing the risk of hospital mortality in patients with acute myocardial infarction, this study was undertaken.

Methods and Materials

Design and participants

To carry out this research, we utilized data from the Yazd Cardiovascular Diseases Registry (YCDR), a web-based and patient-centered registry[9], which focuses on patients with ischemic heart disease admission in some hospitals of Yazd Province.

We used YCDR data focused on a population of patients who experienced acute myocardial infarction in Yazd Province. Initiated in 2014, the registry project for ischemic heart disease patients in Yazd Province has been systematically gathering comprehensive data on individuals afflicted with ischemic heart conditions. By the end of 2018, records for approximately 3,000 patients had been compiled and were accessible. This study focused exclusively on patients who experienced a STEMI (ST-Elevation Myocardial Infarction) heart attack and involved a cohort of 1,861 individuals with myocardial infarction. The data collected from these patients encompassed a range of information, including demographic details, medical history of prior illnesses, medication usage, and risk factors such as hypertension, diabetes, tobacco and hookah smoking, elevated blood cholesterol levels, and a family history of cardiovascular disease. Additionally, the registry recorded details about revascularization procedures, biochemical test results, medications administered within the first 24 hours post-attack, symptoms presented by the patients, echocardiography findings, initial electrocardiogram results, angiography performed during hospitalization, complications that arose while hospitalized, and the status of the patients' health upon discharge.

The registry data underwent meticulous quality control before being cross-referenced with patient files for analysis at the software level.

Sample size

In this study, 1,861 patients who suffered from ST-Elevation Myocardial Infarction (STEMI) and entered in the registry from 2016 to 2018 were included.

Statistical analysis

First, a descriptive analysis was conducted. For quantitative variables, the mean and standard deviation (SD) were calculated, while for non-normally distributed variables, the median and interquartile range (IQR) were determined. Categorical variables were reported with frequency and percentage frequency. The study conducted three separate analyses using clinical data, laboratory data, and a combination of both clinical and laboratory data. For each analysis, logistic regression and decision tree models were applied to evaluate their effectiveness. For analytical analysis, a decision tree analysis was carried out using the R software with the 'rpart' package. This analysis involves segmenting and categorizing the main group of individuals and extends this process to subgroups too. Various criteria are applied to establish the threshold and to identify significant independent variables. In this research, the Gini index is utilized[17]. The relationship is strongest when the probability (p) is equal to 0.5. Consequently, the minimum value of the Gini index was selected as the threshold.

To evaluate the comparison criteria of the model, we employed a data-splitting method, allocating 60% for training and 40% for testing. Three different models clinical, laboratory, and combined clinical-laboratory were developed. In the logistic regression model, we adopted a stepwise approach for variable selection to ensure comparability with the decision tree method. Given that the decision tree method produced categorical outcomes, we utilized the class method available in the 'rpart' package in R. To regulate the decision trees, we set the complexity parameter to 0.01 for all models. The default criterion for selecting the optimal tree within the 'rpart' package was the Gini index. We compared the two models using accuracy measures and, ultimately, the Receiver Operating Characteristic (ROC) curve.

Ethics

The current study was approved by the Ethics Committee of Shahid Sadoughi University of Medical Sciences, code number: IR.SSU.SPH.REC.1400.184. Additionally, Yazd Clinical Data Repository (YCDR) has secured informed consent from all participants or their guardians for involvement in the study, ensuring adherence to the principles outlined in the Declaration of Helsinki.

Results

Out of a total of 3,247 registered patients with

Table 1. Baseline characteristics of qualitative variables of patients with acute myocardial infarction (STEMI) in Yazd province from 2016 to 2018.

Characteristics	Total (n= 1861) N (%)	Dead (n=103) N (%)	Alive (n=1758) N (%)	P-value
Sex				
Male	1418 (76.66)	54 (52.4)	1364(78.1)	0.040
Female	433 (23.34)	49 (47.6)	384 (21.9)	
Marital Status				
Single	117 (6.3)	15 (14.6)	102 (5.8)	<0.001
Married	1744 (93.7)	88 (85.4)	1656 (94.2)	
Job Status				
Employed	275(20.00)	17 (20.73)	258 (19.94)	
Unemployed	749 (54.40)	48 (58.54)	701 (54.17)	0.760
Housekeeper	352 (25.60)	17 (20.73)	335 (25.89)	
Admission by season				
Spring	422 (32.86)	5(20.0)	417(33.1)	
Summer	252 (19.6)	4(16.0)	248(19.7)	
Autumn	326 (25.38)	15(60.0)	311(24.7)	<0.001
Winter	284 (22.70)	1(4.0)	283(22.5)	
CABG history				
Yes	83 (4.57)	5(5.3)	78(4.5)	0.610
No	1731 (95.43)	89(94.7)	1642(95.5)	
PCI history				
Yes	195 (10.97)	4(4.7)	191(11.3)	0.360
No	1581 (89.03)	82(95.3)	1499(88.7)	
CVA history				
Yes	33 (1.80)	5(5.5)	28(1.6)	0.060
No	1797 (98.20)	86(94.5)	1711(98.4)	
Other MI history				
Yes	193 (11.46)	2(2.5)	191(11.9)	0.160
No	1490 (88.54)	77(97.5)	1413(88.1)	
Smoking Status				
Smoker	551 (29.61)	13(12.6)	538(30.6)	<0.001
non-smoker	1310 (70.39)	90(87.4)	1220(69.4)	
Diabetes Mellitus history				
Yes	561 (31.16)	41(42.7)	520(30.5)	0.550
No	1239 (68.84)	55(57.3)	1184(69.5)	

Table 1. Baseline characteristics of qualitative variables of patients with acute myocardial infarction (STEMI) in Yazd province from 2016 to 2018.

Characteristics	Total (n= 1861) N (%)	Dead (n=103) N (%)	Alive (n=1758) N (%)	P-value
Family history of MI				
Yes	458 (36.06)	5(20.0)	453(36.3)	0.090
No	812 (63.94)	20(80.0)	792(63.7)	
Illness severity				
Killip I	1210 (94.31)	14(45.3)	1196(93.2)	<0.001
Killip II	37 (2.89)	4(13.1)	33(2.6)	
Killip III	26 (2.02)	3(14.6)	23(1.8)	
Killip IV	10 (77.0)	4(21.2)	6(0.5)	
Hypertension Status				
Yes	811 (63.11)	17(70.8)	794(63.0)	0.280
No	474 (36.69)	7(29.2)	467(37.0)	

Data are presented as number (%). A P-value of <0. 05 was considered statistically significant and calculated using the Chi-square test.
CBAG: Coronary Artery Bypass Graft, PCI: Percutaneous Coronary Intervention, CVA: Cerebrovascular Accident, MI: Myocardial Infarction, PCI Percutaneous Coronary Interventio.
*Some variables were missing, but in insignificant numbers.

myocardial infarction, 1,861 patients with STEMI were included in the study.

1,429 men and 432 women participated in this study. One finding is that the ratio of men to female in the alive population is higher than that of the dead (P = 0.04). Based on Table 1, the highest occurrence of heart attacks in the surviving group happened during the seasons of spring, autumn, winter, and summer, respectively. This difference is statistically significant (P < 0.001)

Furthermore, we observed that the prevalence of certain factors such as a history of PCI (percutaneous coronary intervention), other previous heart attacks (MI), and family history of heart attacks (MI) was higher in the surviving group compared to the deceased group. However, these differences were not statistically significant (P > 0.05). According to Table 1, the prevalence of smoking in the living population was higher than in the dead, and this difference was statistically significant (P < 0.001). Also, the percentage of single people in the dead population was higher than the living population, and this difference was statistically significant (P < 0.001).

On the other hand, we found that the prevalence of individuals with high illness severity at arrival to hospital (Killip IV) was higher in the deceased group compared to the surviving group. This difference was statistically significant (P < 0.001). (In the description of some variables in Table 1 of the Messing data, it is clear that due to their small number, no problems arose in the estimates and analyses.)

According to Table 2, the average age was significantly higher in the deceased group compared to the alive group (75.7 vs 64.4) (P < 0.001). However, the average levels of blood sugar, cholesterol, LDL, HDL, CPKMax, and MB pick were higher in the deceased group compared to the alive group. Nevertheless, this difference was not statistically significant in both groups (P > 0.05).

Logistic regression Model
Clinical Model
The results of logistic regression, as shown in Table 3, indicate that in the clinical model for STEMI patients, certain variables increase the risk of in-hospital mortality. These variables include sex (men), EF<40, and smoking. Each of these factors was associated with an increased risk of death (P < 0.05). Conversely, having a family history of CVD and being female were associated with a decreased risk of in-hospital mortality (P < 0.05). In fact, it has been shown that these variables (female and family history of CVD) have a protective effect against hospital mortality due to acute heart attack.

Laboratory Model
In the laboratory model, factors such as BS, MBpick, and creatinine increased the risk of hospital mortality from acute heart attack, and HB and HDL cholesterol had a protective effect on the risk of hospital death from acute heart attack (P < 0.05).

Clinical-Laboratory Model
In the combined clinical-laboratory model, previous MI, BS, mean blood pressure (BP), creatinine, and CVA history, as well as a reduced EF, were associated with an increased risk of in-hospital mortality following an acute myocardial infarction. Conversely, higher levels of hemoglobin and a low level of HDL-C were found to be protective (Table 3).

Decision tree Model
Clinical Model
The analysis of applying a decision tree model to laboratory data revealed that serotonin was identified as the most significant input variable for prediction. The resulting decision tree had a depth of four, as depicted in Figure 2.A and 2.B. Key factors affecting the prediction of in-hospital mortality for patients with STEMI include the Killip class, BMI, age, and the need for treatment of ventricular fibrillation or ventricular tachycardia (VF/VT), as shown in Figure 2-A. The decision tree's ROC curve performance in the clinical model was 74.6%, and its accuracy was determined to be 93%, as presented in Table 4 and Figure 1.A.

Laboratory Model
The analysis of applying a decision tree model to the experimental data revealed that the variable MBpick was identified as the most significant input variable for prediction. The resulting decision tree had a depth of six, as depicted in Figure 2-A, B. Key variables that affect the

Table 2. Baseline characteristics of quantitative variables of patients with acute myocardial infarction (STEMI) in Yazd province from 2016 to 2018.

Variables	Dead (n= 103)	Alive (n=1758)	P-value
Age(years)	75.7 ± 12.0	64.4 ± 13.6	<0.001
BMI (kg/m^2)	25.9 ± 3.8	28.2 ± 13.7	0.700
BS (mg/dl)	220.9 ± 88.2	176.7 ± 45.1	0.440
HB (mg/dl)	12.8 ± 1.8	13.5 ± 2.1	0.350
Total Cholesterol (mg/dl)	169.3 ± 35.2	167.8 ± 31.2	0.920
HDL-C (mg/dl)	43.8 ± 4.3	41.0 ± 8.4	0.690
LDL-C (mg/dl)	111.3 ± 30.2	103.5 ± 20. 2	0.880
CPKMax (U/L)	922.05 ± 100.0	513.5 ± 99.1	0.090
MBpick (ng/mL)	111.0 ± 40.1	57.1 ± 9.6	0.120

Data are presented as mean ± SD. A P-value of <0. 05 was considered statistically significant and calculated using an independent sample t-test.
BS: Blood sugar, HB: Hemoglobin, HDL-C: High Density Lipoprotein Cholesterol, LDL-C: Low Density Lipoprotein Cholesterol, BMI: Body Mass Index, CPKMax: Creatine Phosphokinase Maximum
MBpick: Peak Level of Creatine Kinase–Myocardial Band

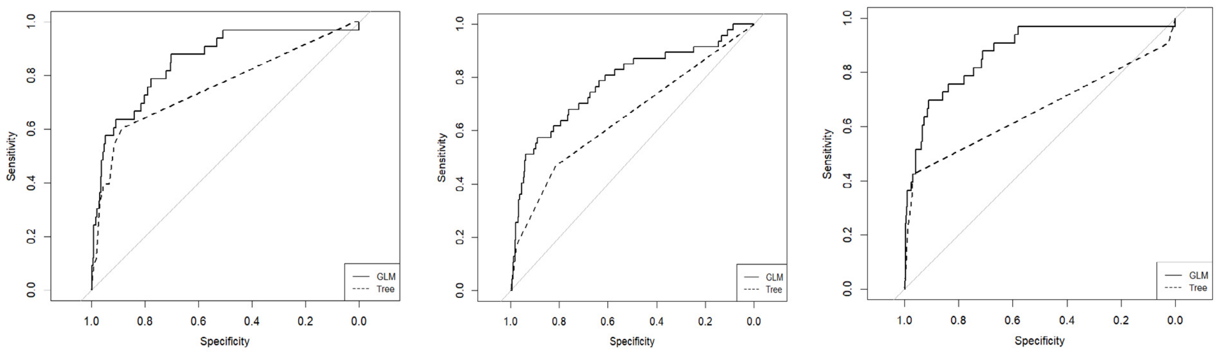
Table 3. Factors predicting hospital mortality in patients with myocardial infarction (STEMI) based on three clinical, laboratory and clinical-laboratory models

	Clinical		Laboratory		Clinical- Laboratory	
Variables	***OR*** **(CI)**	**P-value**	**OR (CI)**	**P-value**	**OR (CI)**	**P-value**
Sex (men)	.36(0.2-0.5)	<0.001	-	-	0.38 (0.1-1.8)	0.550
Total Cholesterol (mg/dl)	-	-	0.99 (0.98 - 1.01)	0.250	-1.03(0.8-1.12)	0.030
LDL – C (mg/dl)	-	-	1.01 (1.00 - 1.88)	0.060	1.93 (0.9 -2.9	0.080
HDL – C (mg/dl)	-	-	0.95 (0.90 - 0.99)	<0.001	0.83 (0.6 - 0.9)	0.040
BS (mg/dl)	-	-	1.06 (1.01 - 1.09)	<0.001	1.05 (0.9 -2.5)	0.080
HB (mg/dl)	-	-	0.80 (0.72-0.88)	<0. 001	0.72 (04 - 0.9)	0.030
Mbpick (ng/mL)	-	-	1.02 (1.01 - 1.03)	<0.02	1.01 (0.6 - 1.9)	0.280
CPKMax (U/L)	-	-	1.00 (1.00-1.00)	0.430	1.02 (0.6- 0.9	0.550
Creatinine (mg/dl)	-	-	1.34 (1.15 - 1.57)	<0.001	0.83 (0.6 - 0.9)	0.040
Family History of CVD (Yes)	.29 (0.16-0.59)	0.001	-	-	0.35 (0.1 - 1.9)	0.26
Previous MI (Yes)	.20 (0.06-0.62)	0.006	-	-	.80 (0.42-1.03)	0.003
CABG History (Yes)	2.01 (0.67-6.07)	0.210	-	-	0.98 (0.50 - 2.20)	0.35
PCI History (Yes)	1.15 (0.44-3.02)	0.76	-	-	-	-
CVA History (Yes)	1.17 (0.35-3.88)	0.790	-	-	5.6 (2.2-20.3)	-
Smoker (Yes)	3.55 (1.62-7.78)	0.005	-	-	1.30 (0.9 - 3.5)	0.300
EF<40	5.20 (2.1-8.3)	0.002	-	-	3.61 (1.20 - 9.80)	004
AIC (Yes)	524.07(255.002-6850.000)	0.112			482.41(321.25-675.20)	0.700
Killip Class	-	-	-	-	-	-
I (Reference)	-	-	-	-	-	
II	2.2 (0.7-3.8)	0.480	-	-	2.96 (0.31 -9.27)	0.34
III	3.38 (0.9-6.2)	0.200	-	-	3.32 (0.5 - 11.2)	0.070
IV	3.38 (1.5-4.8)	0.020	-	-	2.96 (0.31 - 9.27)	0.030

Data are presented as OR (95% CI) and obtained from Regression logistic. A P-value of <0. 05 was considered statistically significant.
EF: Ejection Fraction. CBAG: Coronary Artery Bypass Graft, PCI: Percutaneous Coronary Intervention, CVA: Cerebrovascular Accident, BS: Blood sugar, HB: Hemoglobin, HDL-C: High Density Lipoprotein Cholesterol, LDL-C: Low Density Lipoprotein Cholesterol, MI: Myocardial Infarction, CPKMax: Creatine Phosphokinase Maximum
MBpick: Peak Level of Creatine Kinase–Myocardial Band, EF: Ejection Fraction, AIC: Arrhythmia Induced Cardiomyopathy

prediction of in-hospital mortality for patients with STEMI include MBpick, HDL, creatinine, BS, CPKMax, HB, total cholesterol, and LDL-C, as shown in Figure 3-A, B. The decision tree model's performance, as measured by the ROC curve in the laboratory setting, was 69.8%, and its accuracy was determined to be 92.5%, as presented in Figure 1-B.

Figure 1.



A B C

Figure 1. Comparison of Roc curve performance of logistic regression and decision tree models based on the data of A: clinical model B: laboratory model C: clinical-laboratory model

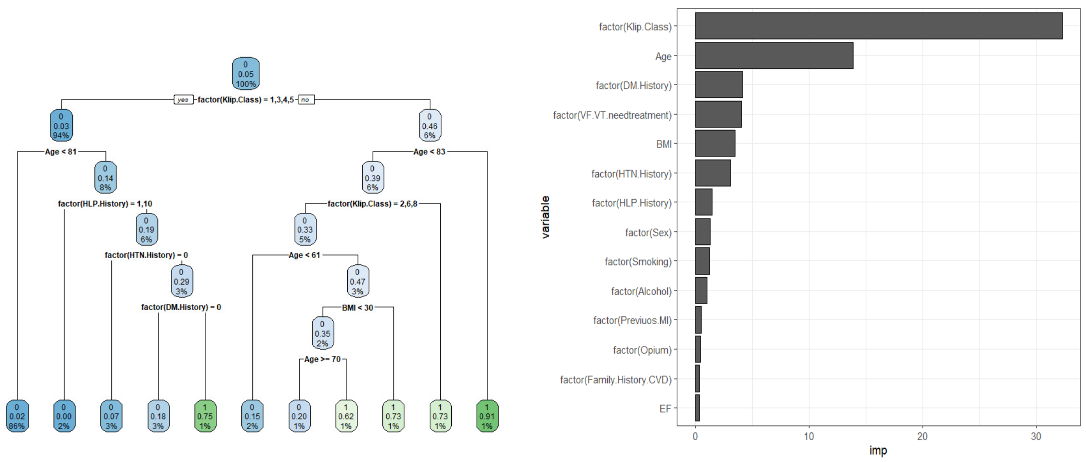


Figure 2. A: The decision tree of hospital mortality of acute myocardial infarction (STEMI) patients in the clinical model B: Rating of important variables influencing the prediction of in-hospital mortality of acute myocardial infarction (STEMI) patients
A (clinical model).

Clinical-Laboratory Model

The analysis of applying a decision tree model to the experimental data revealed that the most significant predictor was the Killip class, which was chosen as the key input variable. A decision tree with a depth of four levels was constructed, as depicted in Figure 2.

The critical factors affecting the prediction of in-hospital mortality for patients with STEMI include Killip class, TG, BS, creatinine levels, ventricular fibrillation or ventricular tachycardia (VF/VT) requiring treatment, age, and HB, in that order (refer to Figure 2-A, B). The decision tree's ROC curve performance in the combined clinical-laboratory model was 81.7%, and its accuracy was determined to be 95.8% (as shown in Table 4 and Figure 1.C). According to the area under the ROC curve, logistic regression models outperformed the decision tree in all three models: clinical, laboratory, and combined clinical-laboratory. However, the decision tree's accuracy was superior to that of logistic regression models (refer to Table 4 and Figure 4).

Discussion

In this study, the risk of mortality from heart attacks was predicted using a dataset from Yazd Province, Iran. Three models incorporating clinical and laboratory variables were employed:

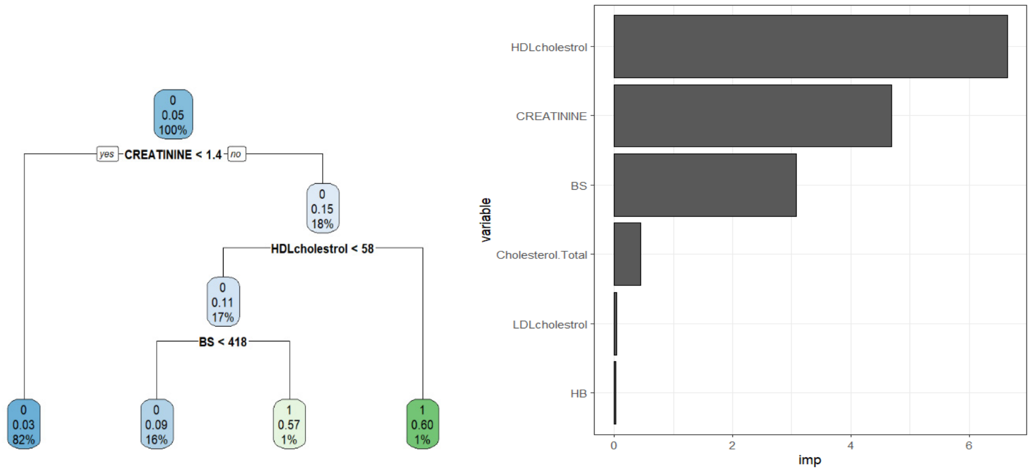


Figure 3. A: The decision tree of hospital mortality of acute myocardial infarction (STEMI) patients in the laboratory model
B: Rating of important variables influencing the prediction of in-hospital mortality of acute myocardial infarction (STEM) patients (laboratory model).

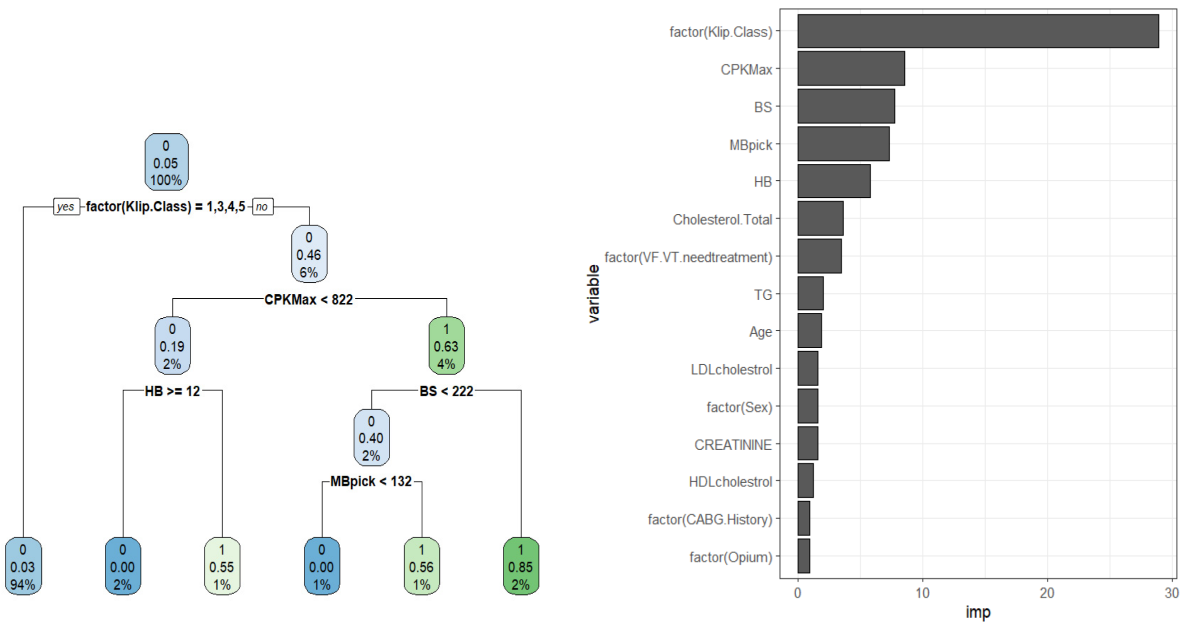


Figure 4. A: The decision tree of hospital mortality of acute myocardial infarction (STEMI) patients in the clinical-laboratory model
B: Rating of important variables influencing the prediction of in-hospital mortality of acute myocardial infarction (STEM) patients (clinical-laboratory model).

Table 4. Accuracy and AUC Model of Logistic Regressions and Decision Tree in Patients with Myocardial Infarction (STEMI) Based on three Clinical, Laboratory and Clinical-Laboratory Models

Model	Logistic Regression			Decision Tree		
Parameters	**Clinically**	**Laboratory**	**Clinically-Laboratory**	**Clinically**	**Laboratory**	**Clinically-Laboratory**
Accuracy	88.3%	81.3%	93.0%	93.0%	92.5%	95.8%
AUC	86.5%	79.5%	90.2%	74.6%	69.8%	81.7

AUC: the Area Under the ROC Curve

two logistic regression models and one decision tree model. Our findings indicated that the cumulative incidence of mortality due to acute myocardial infarction (MI) was 5.5%, accounting for 103 cases.

Consistent with our results, various studies reported similar variables affecting the risk of hospital mortality in STEMI patients. Hospital mortality of acute myocardial infarction in our study (5.5%) was lower than Cretu's study (7.1%)[10]. The findings of a study showed that in-hospital mortality was significantly higher in STEMI patients over 65 years old, in women, in diabetic patients, in patients with a history of MI, and with an advanced Killip class at the time of admission[10]. Another study reported the death rate in STEMI patients as 10.5%, and advanced age and uncontrolled diabetes increased the risk of death in patients with myocardial infarction[10].

Gong et al. reported that Killip class was among the variables that increased the risk of death in STEMI patients[11], and the maximum creatine kinase activity was one of the most important independent predictors for increasing the risk of hospital mortality in patients with acute coronary syndrome complicated by cardiogenic shock[12]. Gao et al. reported an in-hospital mortality rate of 8.13% in STEMI patients. Also, the findings of this study showed that ejection fraction and low-density lipoprotein cholesterol are protective factors affecting in-hospital mortality, and Killip class led to an increase in the risk of in-hospital mortality in STEMI patients[13].

Similar to our in-hospital mortality rate in STEMI patients, another study reported an overall in-hospital mortality rate in STEMI patients of 5.8%, and these patients who died were older, had diabetes and chronic renal failure, had a lower left ventricular ejection fraction, and had Killip class III or IV[14].

Also, one of the important issues in cardiovascular diseases is predicting the occurrence and outcome of cardiovascular diseases, which helps doctors to make more accurate health decisions for their patients. Early diagnosis of the disease can help people make lifestyle changes and, if necessary, ensure effective medical care. Recently, machine learning (ML) techniques have been used to reduce and understand cardiac symptoms[15]. There are various data mining techniques that can be used to identify and prevent mortality among patients with acute myocardial infarction.

In this study, we used the decision tree data mining technique and examined its performance against classical statistical methods such as logistic regression in all three clinical, laboratory, and clinical-laboratory models. The decision tree is a non-parametric method that has been used in the prediction of many diseases, including cardiovascular diseases[16]. This model is a suitable method for data mining when the data have high elongation and skewness or when our qualitative variables are high[17]. One of the important advantages of the decision tree model is its high interpretability, and therefore the decision tree has been introduced as one of the appropriate models for the interpretability of study results. However, one of the problems of the decision tree is the decision method based on only one variable in each step of the algorithm. Logistic regression is able to determine the impact of each of the independent variables on the desired outcome and has good interpretability, but it is strongly affected by the collinearity between the independent variables.

Our findings showed that the ROC curve was higher in the logistic regression model and the accuracy of the decision tree model was higher. Specificity in all three clinical, laboratory, and clinical-laboratory models was higher in the decision tree than in logistic regression, and sensitivity in all three models in logistic regression was higher than in the decision tree. Other studies also compared the performance of decision tree and logistic regression. Raj et al. used logistic regression model and decision tree to predict cardiovascular diseases. Their findings showed that the decision tree algorithm is more accurate than logistic regression for predicting heart disease[18]. Khan et al. proposed machine learning algorithms—logistic regression, KNN classifier, RF, SVM, decision tree, and

Gaussian Naïve Bayes—for the classification of cardiovascular diseases. Using the UCI Cleveland dataset, the accuracy of the logistic regression model was 85.71%[19].

To predict cardiovascular diseases, Reveli et al. used two logistic regression models and a decision tree. Their findings showed that the logistic regression model (92.2%) performed better than the decision tree (86.6%)[16]. Ambrish et al. reported that with the increase of the test ratio (from 50:10 to 90:10), the accuracy of the logistic regression model in predicting cardiovascular disease cases increased from 81.58% to 87.10%[20]. In another study, the logistic regression model showed the best performance against other machine learning algorithms, including the decision tree, with an accuracy of 86.51%[21]. Karthick et al. reported that among machine learning algorithms for predicting cardiovascular diseases, logistic regression—after the random forest algorithm—had the best performance with 80.32% accuracy compared to support vector machine (SVM), Gaussian Naive Bayes, LightGBM, and XGBoost[15]. The accuracy of our CART model is higher than other studies by Shah et al.[22], Li et al.[23], and Tiwari et al.[18], who predicted heart disease using the decision tree model. Ozcan et al., using electronic file information of 1,190 patients and a decision tree model to predict cardiovascular diseases, showed that the CART model has a good prediction accuracy of 87%, and the sensitivity, specificity, and accuracy of this model are 85%, 90%, and 88%, respectively[24]. In another study, Doppala et al. reported the accuracy of the decision tree model for predicting cardiovascular diseases as 82%, and other parameters of the model as 79% sensitivity, 85% specificity, and 83% accuracy[25].

Sensitivity and specificity are inversely related, meaning that as sensitivity increases, specificity decreases and vice versa. Some studies reported better performance of decision tree or logistic regression model in predicting cardiovascular diseases. These studies were different in terms of sample size, age-sex structure, risk factors including underlying diseases and the most attributed to the test, which has led to different reports. In the present study, although the logistic regression model was less accurate than the decision tree model, it showed a higher sensitivity. When the sensitivity of the test is high, it is more likely to give a true positive result and correctly diagnose the disease if it is present, that is, the logistic regression model has more power to correctly diagnose death in STEM patients than the decision tree, which is really possible.

In our study, the sensitivity of the decision tree model was low. A test with low sensitivity is more likely to generate a large number of false negatives and may not detect the disease in people if it is present. Therefore, the decision tree was not able to correctly identify sick people from healthy people, but the specificity of the decision tree was higher than that of the logistic regression model.

Conclusion

In this article, two techniques decision tree algorithm and logistic regression—were used to predict the risk of hospital mortality in patients with acute myocardial infarction. The findings of this study showed that the accuracy of the decision tree algorithm in all three models—clinical, laboratory, and clinical-laboratory was higher than that of logistic regression, but the sensitivity and level of ROC curve performance in logistic regression were higher compared to the decision tree algorithm.

limitations

Among the limitations of this study is the lack of sufficient information on some variables, including chronic kidney disease, which has been reported in many studies as an independent predictor of increased risk of in-hospital mortality in acute myocardial infarction patients. Among the strengths of this study are the large sample size and the quality control of the registry information. One of the limitations of this study is the underestimation of variables such as alcohol and opiate use.

Acknowledgements
The researchers, particularly A. Professor Mohammad Hossein Soltani, wish to extend their heartfelt gratitude to the management and staff of the Registry of Cardiovascular Diseases (YCDR). They also express their appreciation to the referee of this study proposal, Dr. Hamid Reza Dehghan, Assistant Professor of Medical Informatics, for his invaluable contribution in compiling and supporting the registry software for the first time in Yazd Province. Additionally, the School of Public Health at Yazd University of Medical Sciences is acknowledged for the generous support throughout this project.

Conflict of interests
The authors declare no conflict of interest.

Funding
Shahid Sadooghi University of Medical Sciences.

Author's Contributions
Study Conception or Design: SMN, MMo, MM, SMSB
Data Acquisition: HD, MMo
Data Analysis or Interpretation: MA, MMo, HD
Manuscript Drafting: SMN, MMo
Critical Manuscript Revision: MMi, SMSB
All authors have approved the final manuscript and are responsible for all aspects of the work.

References
1. Alpert JS, Thygesen K, Antman E, Bassand JP. Myocardial infarction redefined--a consensus document of The Joint European Society of Cardiology/American College of Cardiology Committee for the redefinition of myocardial infarction. J Am Coll Cardiol. 2000 Sep;36(3):959-69. https://doi.org/10.1016/s0735-1097(00)00804-4
2. Braat SH, Gorgels AP, Bär FW, Wellens HJ. Value of the ST-T segment in lead V4R in inferior wall acute myocardial infarction to predict the site of coronary arterial occlusion. Am J Cardiol. 1988 Jul 1;62(1):140-2. https://doi.org/10.1016/0002-9149(88)91380-x
3. Murray, Christopher J. L, Lopez, Alan D, World Health Organization, World Bank & Harvard School of Public Health. (1996). The Global burden of disease : a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020: summary / edited by Christopher J. L. Murray, Alan D. Lopez. World Health Organization.
4. Yusuf S, Reddy S, Ounpuu S, Anand S. Global burden of cardiovascular diseases: part I: general considerations, the epidemiologic transition, risk factors, and impact of urbanization. Circulation. 2001 Nov 27;104(22):2746-53. https://doi.org/10.1161/hc4601.099487
5. Acharya D. Predictors of Outcomes in Myocardial Infarction and Cardiogenic Shock. Cardiol Rev. 2018 Sep/Oct;26(5):255-66. https://doi.org/10.1097/crd.0000000000000190
6. Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. BMJ. 2013 Feb 5;346:e5595. https://doi.org/10.1136/bmj.e5595
7. Fahimfar N, Khalili D, Sepanlou SG, Malekzadeh R, Azizi F, Mansournia MA, et al. Cardiovascular mortality in a Western Asian country: results from the Iran Cohort Consortium. BMJ Open. 2018 Jul 5;8(7):e020303. https://doi.org/10.1136/bmjopen-2017-020303
8. Tsien CL, Fraser HS, Long WJ, Kennedy RL. Using classification tree and logistic regression methods to diagnose myocardial infarction. Stud Health Technol Inform. 1998;52 Pt 1:493-7.
9. Mohammadi M, Namayandeh SM, Mirzaei M, Shahi MA, Sadr SM, Dehghan H. The most important predictors in hospital mortality of patients with acute ST elevation myocardial infarction (STEMI)-using Yazd Cardiovascular Diseases Registry, YCDR data. 2024. https://doi.org/10.21203/rs.3.rs-3829808/v1
10. Tiwari A, Chugh A, Sharma A. Ensemble framework for cardiovascular disease prediction. Comput Biol Med. 2022;146:105624. https://doi.org/10.1016/j.compbiomed.2022.105624
11. Gong M, Liang D, Xu D, Jin Y, Wang G, Shan P. Analyzing predictors of in-hospital mortality in patients with acute ST-segment elevation myocardial infarction using an evolved machine learning approach. Comput Biol Med. 2024 Mar;170:107950. https://doi.org/10.1016/j.compbiomed.2024.107950
12. Szabo GT, Ágoston A, Csato G, Racz I, Barany T, Uzonyi G, et al. Predictors of Hospital Mortality in Patients with Acute Coronary Syndrome Complicated by Cardiogenic Shock. Sensors (Basel). 2021 Feb 1;21(3):969. https://doi.org/10.3390/s21030969
13. Gao N, Qi XY. Risk factors for in-hospital death in acute ST-segment elevation myocardial infarction after emergency percutaneous coronary intervention: a multicenter retrospective study. Ann

Palliat Med. 2021 Nov;10(11):11756-66. https://doi.org/10.21037/apm-21-2722
14. Falcao FJdA, Alves CMR, Barbosa AHP, Caixeta A, Sousa JMA, Souza JAM, et al. Predictors of in-hospital mortality in patients with ST-segment elevation myocardial infarction undergoing pharmacoinvasive treatment. Clinics (Sao Paulo). 2013 Dec;68(12):1516-20. https://doi.org/10.6061/clinics/2013(12)07
15. Karthick K, Aruna SK, Samikannu R, Kuppusamy R, Teekaraman Y, Thelkar AR. Implementation of a Heart Disease Risk Prediction Model Using Machine Learning. Comput Math Methods Med. 2022 May 2;2022:6517716. https://doi.org/10.1155/2022/6517716
16. Yadav P, Jaiswal K, Patel S, Shukla D. Intelligent heart disease prediction model using classification algorithms. IJCSMC. 2013; 3(08): 102-7.
17. Harper PR. A review and comparison of classification algorithms for medical decision making. Health Policy. 2005 Mar;71(3):315-31. https://doi.org/10.1016/j.healthpol.2004.05.002
18. Raj KS, Thinakaran K. Prediction of Heart Disease using Decision Tree over Logistic Regression using Machine Learning with Improved Accuracy. Cardiometry. 2022(25): 1514-9. https://doi.org/10.18137/cardiometry.2022.25.15141519
19. Khan Z, Mishra DK, Sharma V, Sharma A, editors. Empirical study of various classification techniques for heart disease prediction. In: 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA); 2020. Piscataway, NJ: IEEE; 2020. https://doi.org/10.1109/ICCCA49541.2020.9250852
20. Ambrish G, Ganesh B, Ganesh A, Srinivas C, Dhanraj, Mensinkal K. Logistic regression technique for prediction of cardiovascular disease. Glob Transit Proc. 2022;3(1):127-30. https://doi.org/10.1016/j.gltp.2022.04.008
21. Muppalaneni NB, Ma M, Gurumoorthy S, Kannan R, Vasanthi V. Machine learning algorithms with ROC curve for predicting and diagnosing the heart disease. Soft computing and medical bioinformatics. 2019: 63-72. https://doi.org/10.1007/978-981-13-0059-2_8
22. Shah D, Patel S, Bharti SK. Heart disease prediction using machine learning techniques. SN Computer Science. 2020; 1: 1-6. https://doi.org/10.1007/s42979-020-00365-y
23. Li JP, Haq AU, Din SU, Khan J, Khan A, Saboor A. Heart disease identification method using machine learning classification in e-healthcare. IEEE access. 2020; 8: 107562-82. https://doi.org/10.1109/ACCESS.2020.3001149
24. Ozcan M, Peker S. A classification and regression tree algorithm for heart disease modeling and prediction. Healthcare Analytics. 2023; 3: 100130. https://doi.org/10.1016/j.health.2022.100130
25. Doppala BP, Bhattacharyya D, Janarthanan M, Baik N. A Reliable Machine Intelligence Model for Accurate Identification of Cardiovascular Diseases Using Ensemble Techniques. J Healthc Eng. 2022 Mar 8;2022:2585235. https://doi.org/10.1155/2022/2585235