Caspian Journal of Internal Medicine 2025 (Autumn); 16(4): 775-790
DOI: 10.22088/cjim.16.4.775

Original Article

One-year survival prediction models following ST-elevation myocardial infarction: A comparative analysis of the Cox Frailty Model and machine learning

Mansour Rezaei (PhD) [1,]
Maryam Montaseri (PhD) [2*]
Shayan Mostafaei (PhD) [3]
Armin Khayati [4]
Mohammad Taheri (PhD) [5]

1. Social Development and Health Promotion Research Center, Health Institute, Kermanshah University of Medical Sciences, Kermanshah, Iran
2. School of Health, Kermanshah University of Medical Sciences, Kermanshah, Iran
3. Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden
4. Department of Computer Science and Engineering, School of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran
5. Department of Computer Science and Engineering, School of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran

*** Correspondence:**
Maryam Montaseri, School of Health, Kermanshah University of Medical Sciences, Kermanshah, Iran

E-mail:
Maryam.Montaseri@Kums.ac.ir
Tel: +98 8318264447

Received: 27 Nov 2024
Revised: 15 Jan 2025
Accepted: 18 Jan 2025
Published: 19 Oct 2025

Abstract

Background: The aim of this study was developing and comparative analyzing prediction models using a Cox proportional hazards model with and without frailty, random survival forests (RSF) and survival support vector regression (SVR).

Methods: In this study, 2800 patients with STEMI have been used and two machine learning methods for survival analysis have been applied: RSF and SVR, then the Cox model with and without frailty has been employed. The main outcome was 1-year mortality after STEMI. In this study, 16 variables have missing data. After applying four multiple imputation via chained equations methods, the "Sample" algorithm was selected as the appropriate model with complete data and the modeling process was continued with this data and Hazard Ratio (HR) were calculated.

Results: Overall, 1628 (58.1%) patients received primary percutaneous coronary intervention and 737 (26.3%) received thrombolytic therapy. Based on the experimental results, between all the models, the Cox with frailty model performed the best, with the highest overall C-index (0.891) and time-dependent area under the curve (0.9134) and the least Brier score (0.0458). Ever smoking (HR= 1.46), systolic blood pressure (HR= 0.98), left ventricular ejection fraction (HR= 0.96), glomerular filtration rate (HR= 0.96), and reperfusion therapy (No reperfusion HR= 2.71) independently associated with 1-year mortality of STEMI patients.

Conclusion: The findings suggest that there are advantages in developing frailty models further than the fundamental Cox proportional hazards regression for estimating the likelihood of survival for STEMI patients to account for the unobserved heterogeneity in grouped observations.

Keywords: Survival analysis, Machine-learning, Myocardial Infarction, Frailty model.

Citation:
Rezaei M, Montaseri M, Mostafaei Sh, Khayati A, Taheri M. One-year survival prediction models following ST-elevation myocardial infarction: A comparative analysis of the Cox Frailty Model and machine learning. Caspian J Intern Med 2025; 16(4): 775-790.

Ischemic heart disease is the leading cause of death globally, responsible for 16.17% of all fatalities (1). With a global frequency of 197 million cases, it was the primary cause of 9.14 million deaths and 182 million years of life with a handicap in 2019 (2). Ischemic heart disease, which accounts for 26.28% of all fatalities in Iran, is the main cause of mortality. The acute coronary syndrome, which includes ST-segment elevation myocardial infarction (STEMI), maybe the initial sign of ischemic heart disease and is associated with high morbidity and death (3). There are regional variations in STEMI mortality rates and treatment methods both within and between nations, indicating room for improvement (4). Primary percutaneous coronary intervention (PPCI), thrombolysis, and pharmaco-invasive (i.e., thrombolysis followed by angiography and, if required, percutaneous coronary intervention [PCI]) are examples of contemporary reperfusion therapies that are commonly employed (5).

However, the most frequent cause of STEMI mortality in low- and middle-income countries (LMICs), where 80% of all cardiovascular deaths take place, is a lack of an adequate care system (5). There is no consensus on the factors predicting short-term mortality in cases of STEMI (6). Numerous health-system-level and individual-level variables can affect STEMI mortality, including time to treatment, effectiveness of the ambulance system, reperfusion strategy, in-hospital treatment, age, history of heart disease, renal function, number of afflicted coronary arteries, and known risk factors like diabetes mellitus, hypertension, dyslipidemia, and tobacco use (6, 7).

In the statistical field of survival analysis, the time until the occurrence of a certain event, such as the transition from being alive to being dead, serves as the random variable. This event reflects a qualitative change or the move from one categorical state to another. Death is the event that is most frequently researched in the field of biomedicine (8). One of the main objectives of survival analysis is to examine the time to the event of interest for patients with certain predictors. It offers useful details about factors that affect survival and can illuminate ways to extend patients' lives. Such analysis also helps clinicians create appropriate treatment plans for individuals with varied risk levels and more effectively allocate resources (9). When time-to-event statistics show that the event of interest may not be seen by all subjects, censorship is a common problem that needs to be properly dealt with. Subjects with unobserved event times are referred to as censored (10, 11). Models based on regression can be created using the parametric and semi-parametric techniques including the Cox proportional hazards model (12, 13). Biomedical research is increasingly using machine learning (ML) techniques to analyze data (10). But today's majority of machine learning techniques are created for uncensored data. To adapt current ML techniques to work with censored data, enormous amounts of work are required (14). Several ML methods, such as the survival tree model (15) and the support vector approach (12), have been modified to handle survival data.

It is necessary to account linear/and nonlinear relationships and complex interactions between biomarkers and survival time for better modelling. In addition, the support vector regression (SVR) algorithm has some limitations regarding non-linear relationships and complex interactions, may be its complexity to support large number of subjects (instances) and censoring handling. However, the random survival forests (RSF) model as a machine-learning algorithm is capable of modeling complex interactions and non-linear relationships in survival data. To this end, a Cox PH regression was applied with and without frailty and two ML methods such as RSF and SVR. Finally, C-index, Brier score and time-dependent area under the curve (AUC) for 1-years survival time of STEMI patients is compared among these models. The primary objective of modeling in this work was to predict survival and optimize the concordance index (C-index) and time-dependent AUC, irrespective of the method used for generating predictions. As a result, we did not run the proportional hazards test during the modeling phase (16).

Methods

Study setting, design, and participants: This registry-based cohort study with the code of ethics: IR.KUMS.REC.1400.272, included all adult patients (> 18 years) who presented with STEMI to Imam- Ali Hospital from July 1, 2016 to September 19, 2019. Imam-Ali Hospital is a cardiology training center in the city of Kermanshah, affiliated to the Kermanshah University of Medical Sciences, in Iran. This is the only 24/7 PPCI-capable hospital in the province. Diagnosis was made by cardiologists based on current guidelines (17). This is the only 24/7 PPCI-capable hospital in the province. The STEMI patients who were hospitalized more than 24-hours before referring to Imam-Ali Hospital were excluded from this registry. In this study, we also excluded patients with out-of-hospital cardiac arrest. The eligible patients were followed up 1 year after STEMI events. The sample size was 2800. Written informed permission was signed by each participant.

Baseline variables: In this cohort, skilled doctors and nurses obtained demographic and clinical information from patient and/or attendant interviews, including past medical histories, the onset of symptoms, and transfers to Imam-Ali Hospital. Direct admission (self-presentation) or referral from other hospitals to Imam-Ali Hospital were both documented. It was determined how long it took from the start of a symptom to when the patient arrived at the hospital. Based on self-reports of confirmed diagnoses by healthcare members, a history of cardiovascular events (prior myocardial infarction, stroke, or chronic heart failure), coronary intervention (PCI or coronary artery bypass graft surgery), diabetes, and hypertension were documented. In addition to PPCI, pharmacoinvasive procedures, thrombolysis alone, and none (no reperfusion), the reperfusion therapies performed were documented. Before or after admission to Imam-Ali Hospital, thrombolysis was given. Hospital medical records were used to collect details regarding the admission procedure, hemodynamic status, electrocardiography data, medical

treatment, laboratory testing, etc. Systolic blood pressure (SBP) and heart rate (HR) were measured upon admission to Imam-Ali Hospital and divided into two groups based on the TIMI risk score categories (SBP: 100/ 100 mm Hg and HR: > 100/ 100 bmp, respectively) (18). The calculation of body mass index (BMI) involves dividing weight in kilograms by the square root of height in meters. Lipid profile and creatinine level was measured on the first day of admission. We defined high low-density lipoprotein cholesterol (LDL-C) as LDL-C > 160 mg/dL and low high-density lipoprotein cholesterol (HDL-C) as HDL-C < 40 mg/dL in men and HDL-C < 50 mg/dL in women (19). The CKD-EPT equation was used for the estimation of the glomerular filtration rate (GFR) (20). Qualified medical professionals verified the accuracy of all recorded data.

Death Event: Cardiovascular diseases are among the world's top causes of demise. The formation of plaque in the blood vessels, which restricts or blocks blood flow, renal system malfunction, which raises creatinine levels, low salt levels, changing ejection fraction, and other cardiac abnormalities can all be causes of heart failure. Acute myocardial infarction, gradual heart failure, sudden death, or other circulatory irregularities can all result in death depending on the severity of the aforementioned conditions. Depending on a person's gender, color, and ethnicity, death may take several forms.

Main outcome and follow-up period: The main outcome was all-cause of 1-year mortality after STEMI. The hospital records were used to collect data on in-hospital mortality. The follow-up period is extended from the date of STEMI diagnosis until death, loss-to-follow-up, or 1-year after STEMI, whichever occurs first. In this study, time means survival time of patients (days).

Data analysis:

Imputation methods: In most medical data sets, there is a problem with missing data. The most typical method is to eliminate the observations with missing data, which results in a complete participant analysis. When the group removed is a chosen subsample of the research population, that is, when the values are not completely missing at random, this method not only wastes data and loses power but also creates biased results (21, 22). It has been demonstrated that an ad hoc method can be used to replace missing data with a fixed value, such as the mean (in the case of data that is regularly distributed) or median of the observed values (in the case of skewed data). Due to the use of a single value to replace all missing data, this method may artificially reduce variance and weaken correlations with other variables (23). Use one of the several methods for imputing the missing data as an alternative. The approach of multiple imputations (MI) (24) is the most appealing of these since theoretical and simulation studies have demonstrated that it produces estimates with favorable statistical qualities, such as efficiency and validity when the appropriate model is specified for the imputation. Here, all of the imputation techniques are predicated on the premise that data are missing at random (25).

In multiple imputation, each missing value is replaced with the M possible value, and finally the M complete observation set is produced. The M value is usually chosen between 5 and 10. The probability of a value being missed may be influenced by observable data, which can offer insight into the missing values and serve as a foundation for imputation; nevertheless, this is independent of the unobserved data (22). Multiple Imputation via Chained Equations (MICE), one of the advancements in the field of analysis of missing data, offered a strategy to deal with missing data effectively. This approach uses partial observations to impute missing data and is based on likelihood. Even though this method yields more accurate estimates, it still necessitates knowledge of computer software, and explaining the processes to clinical audiences that might be challenging. The MICE technique's basic idea was to use the distribution of the seen data to estimate likely values for the missing data, then add random components to the predictions to account for uncertainty. Here, each missing value is replaced by one of five values to provide five sets of imputed data (26). To obtain a single Hazard Ratio (HR) and Confidence Interval (CI), estimates produced from imputed data sets were combined using Rubin's method. Each data set was analyzed separately (24). Missing data were imputed using the MICE package (27). The MICE method is one of the best ways to handle missing data, according to a thorough study (28); Additionally, literature-based research indicates that the MICE method is the most effective way to impute missing data (23).

In this study, 16 variables have missing data (figure 1) (table 1). The possible value of M was considered equal to five. According to the type of missing variables, which were of the quantitative and qualitative types, a total of four imputation algorithms with five models were developed. The only difference between them was the approach utilized to deal with missing data (table 2). In the development of all the models, the Cox regression model was fitted to the data, then the concordance index and Akaike information criterion (AIC) were compared. Finally, the algorithm with the highest value of concordance and the lowest value of (AIC) were selected as the algorithm with suitable complete data. After checking the results of table 2, model 2 of the “Sample” algorithm was selected as the appropriate model

with complete data, and the modeling process was continued with this data.

Statistical analysis: Python (Scikit Survival) and R 4.3.1 were used for all statistical analyses. Since we included all eligible patients in the experiment and used registry data, we did not compute the sample size. Qualitative factors were displayed as counts (percentages), whereas quantitative data were given as means (SD). Cox proportional-hazard modeling with and without frailty was performed to determine predictors of 1-year mortality. According to previously published mortality predictors, potential variables for the study were chosen based on the medical consultant's assessment of the approach (7, 18). First, a rank logarithm test was used to determine the factors affecting the patient survival time. All of the variables that showed significance in the aforementioned test were then included in the models, as were those that did not show significance but had a p-value of less than 0.25. These variables were age, Gender (female/male), Education (illiterate, under diploma, upper diploma), History of MI, History of CHF, History of stroke, History of PPCI, History of CABG, ever smoking, diabetes, hypertension, LDL-C, HDL-C, GFR, hemoglobin, BMI, MI type (anterior wall or left bundle branch block/ others), admission, the highest level of CK-MB, reperfusion therapy (primary PCI, thrombolysis alone and no-reperfusion), LVEF, and Systolic blood pressure. We reported HRs with 95% confidence intervals (95% CIs) using Cox proportional hazards models.

Cox PH model: Conventional survival models, like the widely used Cox model, are usually developed from the hazard function, which is the instantaneous likelihood of the event of interest occurring within a restricted time frame (9). As a result, the hazard ratio (HR) is commonly used to express the treatment impact in traditional clinical studies (29). The Cox proportional hazards model directly produces HR, which is an estimator of relative risk and hazard rate decrease across different groups. Let $h(t \mid X)$ to the hazard/failure function at time t given the covariates X; the proportional hazards model (PHM) is expressed as:

$$h(t \mid X) = h_0(t) \exp(\beta X), \quad (1)$$

Where $h_0(t)$ is the baseline hazard/failure function. X is the vector of the covariates and β is the regression coefficient vector (30).

Frailty model: The frailty model aims to identify and address frailty using a method called the frailty model or the model with frailty. This model of conditional risk incorporates a multiplicative component. This phrase suggests that one patient may be more fragile than another, putting them at greater danger of passing away or having their illness worsen. Using an unobserved random variable ω, known as frailty, the fundamental idea is to introduce the dependence between the survival times $t_1, \ldots, t_d$ (31).

Regression models frequently employ random variables to represent unobserved dependence in the data. When used in survival analysis (also known as time-to-event outcome analysis), these models are called shared frailty models. Cluster members share an unobserved common risk, which is represented by the frailty, a cluster-specific random effect. According to these models, the failure times of cluster members are believed to be independent since the random frailty factor is thought to capture all within-cluster dependence given the reported variables and the unobserved frailty (32, 33).

In actuality, survival analysis areas frequently deal with correlated or clustered failure time data. Dependency on the observed failure time results from the subjects' shared environment. Frailty creates reliance among the correlated or clustered failure time data and is a useful tool for representing the random effect shared by patients in the same cluster (or group). The PHM frailty model (34) is:

$$h(t \mid X_{ij}, \omega_i) = \omega_i h_0(t) \exp(\beta X_{ij}), \quad (2)$$

Consider n independent clusters, with cluster i, i = 1, . . . , n, having mi ≥ 1 members. For member j of cluster i, let X_{ij} be a vector of covariates.

The random effect shared by the correlation between the outcomes within members of the ith cluster (group) is also modeled by the frailty factor, ω_i. Because the gamma frailty distribution is mathematically tractable, it has been used in previous publications and packages. However, a restrictive type of reliance is induced by the gamma hypothesis (32, 35). We used the Cox model without and with a gamma-frailty survival method, based on the maximized integrated log-likelihood and the Akaike information criterion (AIC).

Random survival forest: The Random Survival Forest (RSF) and survival trees are two further methods for handling the limited survival data. As it happens, random forests have grown to be a very effective, well-liked, and potent tool for survival analysis. One could think of the random forest as a nonparametric machine-learning technique (36).

The RSF approach uses a forest of survival trees to make predictions, extending Breiman's random forest method to right-censored survival data. Similar to classification and regression situations, RSF is an ensemble learner that is produced by averaging a tree base-learner. In survival situations, a binary survival tree serves as the base learner, and the ensemble learner is produced by averaging the cumulative hazard functions of each tree's Nelson-Aalen (37). RSF consists of four basic steps:

(1) Randomly select B bootstrap samples from the provided dataset. The remaining data is referred to as the out-of-bag (OOB) data since one-third of the training set's data is absent from the bootstrapping sample. In this study B=1,000.

(2) Create a survival tree for every sample by selecting a subset of the variables at random. Next, split the nodes into their child nodes using the candidate factors that maximize the survival difference between child nodes. Three criteria are used in this case to assess the survival difference: the log-rank statistic, the gradient-based Breier score, and the log-rank score. In this study, 3 candidate variables were randomly selected out of all 24 variables.

(3) Extend the tree to its maximum size while requiring that each terminal node have a specific amount of distinct unfiltered patients. In this study, the minimum final node size was equal to 15.

(4) Based on the Nelson-Aalen estimator, get the cumulative hazard function (STEMI) for each terminal node, and then obtain the ensemble of the OOB data by averaging the STEMI of each tree (38).

The variable importance (VIMP) for x is the difference between the prediction error for the initial ensemble and the prediction error for the new ensemble produced by randomizing the assignments for x (39). According to (39, 40) positive values denote variables with the capacity for prediction (important values) while zero or negative values denote variables that are not capable of prediction (not important values).

For the RSF technique in this work, the two node splitting criteria (log-rank splitting and random) were utilized (table 3).

Support vector regression: Support Vector Regression (SVR) is a supervised machine learning technique used for regression tasks (41). SVR has also been applied to censored regression issues, such as survival analysis (12, 42). The "support vectors" in SVR are the data points that define the margin and are closest to the regression line. When choosing the regression model, these factors are very important. To put it briefly, SVR is a regression technique that looks for a regression model that strikes a compromise between fitting the data and preventing overfitting by finding a margin around the projected values. Through the selection of kernel functions, it may be tailored to different problem domains and is especially helpful when working with non-linear connections. The core principle of SVR is to use a regularization parameter to minimize the $\mathcal{E}$-insensitive loss function, max (0,| f (x_i)- y_i |-$\mathcal{E}$). In this case, the projected value and the actual value of i^{th} subject, and also the allowable margin of the error are represented by the letters f (x_i), y_i, and $\mathcal{E}$, respectively (38). In the SVR model, the clinical factors x, as a feature vector, explain the overall survival time y. Then, y is tried to be stimated as a function f of its mapped feature vector i.e., y=f(φ(x)) + ∈, where φ(.) is referred to as the feature map function. Calculating the feature map itself is uncommon because it typically indicates a higher-dimensional space. As a result of Mercer's theorem (43), the kernel k(x_i, x_j)=φ(x_i)^T φ(x_j) directly calculates the inner product of associated feature vectors of patients i and j, in the new maping space without computing the mapping vector for each one separately. The kernel is simply used to carry out the complete process of training the model and producing predictions. The kernel plays a significant role in constructing SVR models, and various types of kernels are available. The linear kernel is also given as k(x_i, x_j)= $x_i^T x_j$(44).

We performed support vector analysis for datasets with survival outcomes by package ‘survivalsvm’. The package offers three different methods: The regression method formulates the inequality requirements of the support vector issue while accounting for censoring. The goal of the ranking approach's inequality restrictions is to maximize the concordance index for similar observation pairs. Regression and ranking restrictions are combined into a single model in the hybrid technique (40).

Harrell’s concordance index: The ratio of the concordance pairs in the data to all the pairs is known as Harrell's concordance index. It determines discrimination and forecasts the probability that, out of two patients chosen at random, the patient with the higher expected risk will have longer intervals without events (45). If the projected score is close to 0.5, choosing the patient who will experience no events for the longest is no more accurate than flipping a coin. The ideal value is one. In survival analysis, it is frequently used to assess risk models.

Survival performance is gauged by Harrell's concordance index. It specifically considers censoring the individuals and does not depend on picking a specific time for the model's evaluation. When calculating the error rate, Harrell's concordance index, or C, is used as the input. Error rates range from 0 to 1, where 1 represents perfect accuracy and 0.5 represents a procedure that performs no better than random guessing (39). The random Survival Forest package was used by R 4.3.1 to analyze the data. RSF also extracted 1000 bootstrap samples from the generated data, developed a tree for every bootstrapped data set, and divided a predictor according to a survival-splitting criterion. 1,000 replications of each approach were used to generate concordance error rates, which were then averaged to determine the results.

Time-dependent area under the curve (AUC): A model's overall discriminative performance is evaluated by the C-index, whereas the time-dependent AUC compares the projected probabilities with the actual binary survival status and the probability estimate of a death outcome of censored observations at a period of interest. Perfect discrimination is represented by a value of 1, while random guessing is represented by a value of 0.5 for the time-dependent AUC and the C-index (46).

Brier score: The average squared distances between the observed survival status and the anticipated survival probability are represented by the Brier score, which is always a number between 0 and 1, with 0 being the greatest possible value. It is used to assess the accuracy of a predicted survival function at a specific time t (47). All analyses were conducted using the full case data. Statistical significance is defined as P- value < 0.05.

Table 1. Variables sorted by percentage of missing

Variables	Percentage	Variables	Percentage
HDL (mg/dL)	0.0639	**GFR_CKD (IU/L)**	0.0025
Education	0.0600	**Diabetes**	0.0018
LDL (mg/dL)	0.0468	**Peak CKMB**	0.0018
LVEF	0.0354	**Hypertension**	0.0007
BMI (kg/m^2)	0.0296	**Gender**	0.0000
CHF	0.0036	**Ever Smoking**	0.0000
History of PPCI	0.0032	**Perfusiontherapy**	0.0000
History of CABG	0.0032	**Anterior MI**	0.0000
Early Hb (g/dl)	0.0032	**Admission**	0.0000
History of MI	0.0029	**Outcome**	0.0000
History of Stroke	0.0029	**Time (days)**	0.0000
Sys BP (mm Hg)	0.0029	**Age (years)**	0.0000

HDL: high-density lipoprotein, LDL: low-density lipoprotein, LVEF: left ventricular ejection fraction, BMI: body mass index, CHF: congestive heart failure, PPCI: primary percutaneous coronary intervention, CABG: coronary artery bypass surgery, Hb: hemoglobin, MI: myocardial infarction, Sys BP: systolic blood pressure, GFR_CKD: glomerular filtration rate- chronic kidney disease, CKMB: creatine kinase-MB.

Table 2. Imputation methods based on the quantitative and qualitative variable type in the MICE function

Methods	Symbol	M=5	Concordance(se)	AIC
Predictive mean matching	PMM	Model1	0.813 (0.0121)	4826.953
		Model2	0.8094 (0.01228)	4839.06
		Model3	0.8107 (0.01208)	4826.833
		Model4	0.8087 (0.01223)	4839.405
		Model5	0.8119 (0.01202)	4811.239
Random sample from observed values	Sample	Model1	0.8176 (0.01214)	4419.951
		Model2	0.8218 (0.01214)*	4402.058**
		Model3	0.8125 (0.01218)	4487.099
		Model4	0.8132 (0.01241)	4437.555
		Model5	0.8181 (0.01199)	4433.494

Methods	Symbol	M=5	Concordance(se)	AIC
Classification and regression trees	CART	Model1	0.8076 (0.01242)	4843.33
		Model2	0.8138 (0.01222)	4822.059
		Model3	0.8132 (0.01192)	4823.498
		Model4	0.8066 (0.01216)	4851.607
		Model5	0.8147 (0.01212)	4819.393
Random forest imputations	RF	Model1	0.8062 (0.0125)	4843.204
		Model2	0.8062 (0.0123)	4844.868
		Model3	0.8149 (0.01193)	4770.057
		Model4	0.818 (0.01178)	4710.579
		Model5	0.812 (0.01202)	4819.87

* The highest value of concordance, ** The lowest value of Akaike information criterion (AIC).

Table 3. Hyperparameters for random survival forests

Hyperparameter	Value	
Number of trees	500	
No. of variables tried at each split	3	
Splitting rule	Log-rank /Random	
(OOB) Requested performance error	Training	0.1709299
	Test	0.1942646

Out-of-Bag (OOB)

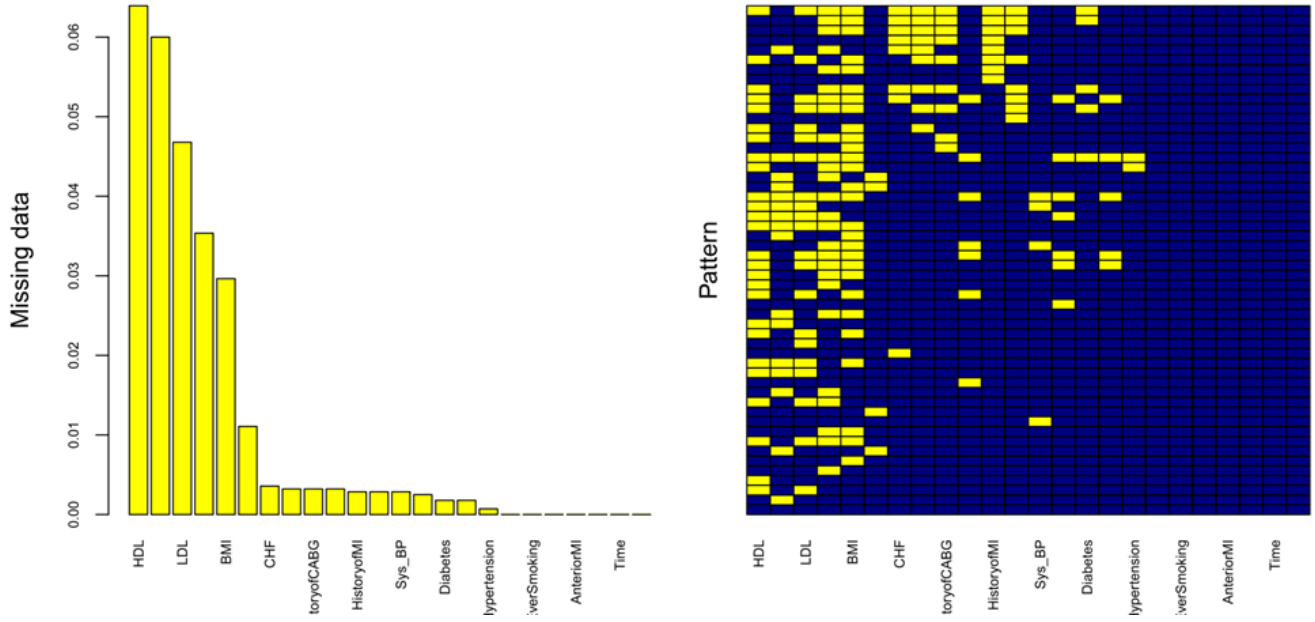


Figure1. The Structure of Missing data in this study

Results

Of the 2800 patients, the mean age of the patients was 60.76±12.34 years and 2167 (77.4%) were males. Also, 2079 (73.9%) were directly admitted to Imam-Ali Hospital and 731 (26.1%) were referred from other hospitals. Overall, 1628 (58.1%) patients received PPCI and 737 (26.3%) received thrombolytic therapy. The median duration of follow-up was 178.5 days. The only quantitative variable was reported as the mean (standard deviation), and the other qualitative variables were reported as the frequency with percentage (table 4). We used the Cox PH regression with and without the frailty correction on typical patient hazard ratios (95% CI) and contrasted the results with those derived from machine learning techniques. As shown in table 5, the hazard ratio for ever smoking, admission, systolic blood pressure, LVEF, GFR, reperfusion therapy (No reperfusion) and Anterior MI was significant in the Cox PH model with frailty. The frailty model has also improved the model due to the relevance of the variance of the random effect (0.59). In this study, we split the dataset into 70% (1960 randomly selected patients) for the training set, 30% (840 randomly selected patients) for the test set. In the train and test set, the important variables in the construction of the decision tree are shown with positive values. By comparing the error rate value obtained for the test set and the training set, and due to the smallness and closeness of these values (table 3), we can report the results of the decision tree well. Figure 2 & 3 shows the most important variables in the construction of the decision tree for the subject and test datasets in STEMI patients, which are almost the same. Ever smoking, LVEF, systolic blood pressure, GFR, and reperfusion therapy in training and test set was the most important. Finally, for the survival support vector regression, we split the dataset into 70% (1960 randomly selected patients) for the training set, 30% (840 randomly selected patients) for the test set and applied a regression approach. The kernel is a collection of numerical functions used in SVR computations. Information can be accepted as information by the kernel, which can then transform it into the required structure. Diverse piece functions are used in different SVR computations. They can serve a variety of purposes. The SVR kernels that we have examined here are linear, RBF, polynomial, and clinical. The results of the 4 methods are shown in table 6. Of the 4 methods, the Cox with frailty model performed the best, with the highest overall C-index and mean AUC and the least Brier score. The RSF models, over 1 year's follow-up, achieved a mean C-index of 0.8641 for testing. This means that the data is well-trained and executed with high accuracy in the test phase. In survival SVR between four kernels, the RBF kernel had a better concordance index and mean AUC than others.

Table 4. Baseline characteristics of STEMI patients according to vital status.

	Patient Died (n = 305)	Patient Alive (n = 2464)	Patient Missing (n = 31)
Quantitative Variables			
Age (years)	68.12 (12.74)	59.79 (11.95)	64.98 (14.1)
Body mass index (kg/m2)	25.28 (4.22)	26.35 (4.09)	25.12 (4.33)
Systolic blood pressure (mm Hg)	121.58 (34.77)	135.49 (29.48)	135.90 (27.70)
LVEF	30.54 (10.70)	38.47 (9.29)	35.33 (9.46)
Early Hemoglobin (g/dL)	14 (2.1)	14.83 (1.74)	14.65 (2.38)
LDL-C (mg/dL)	102.19 (35.08)	104.37 (30.68)	95.06 (25.25)
HDL-C (mg/dL)	41.34 (11.07)	41.37 (9.02)	39.81 (5.66)
Peak CK-MB (IU/L)	135.39 (137.83)	125.14 (117.05)	112.65 (87.31)
GFR (mL/min per 1.73m2)	53.73 (19.08)	70.23 (17.25)	64.80 (19.66)
Time (days)	69.67 (107.53)	365	-

		Patient Died (n = 305)	Patient Alive (n = 2464)	Patient Missing (n = 31)
Qualitative Variables				
Gender	Female	114 (37.4)	513 (20.8)	6 (19.4)
	Male	191 (62.6)	1951 (79.2)	25 (80.6)
Ever Smoker	Yes	120 (60.7)	1234 (50.1)	17 (54.8)
	No	185 (39.3)	1230 (49.9)	12 (45.2)
Education	Illiterate	155 (61)	693 (29.2)	1 (20)
	Under Diploma	67 (26.4)	998 (42.1)	3 (60)
	Upper Diploma	32 (12.6)	682 (28.7)	1 (20)
Diabetes mellitus	Yes	92 (30.7)	490 (19.9)	7 (22.6)
	No	208 (69.3)	1974 (80.1)	24 (77.4)
Hypertension	Yes	177 (58.4)	1011 (41)	16 (51.6)
	No	126 (41.6)	1453 (59)	15 (48.4)
History of MI	Yes	39 (13.1)	291 (11.8)	5 (16.2)
	No	258 (86.9)	2173 (88.2)	26 (83.8)
History of CHF	Yes	17 (5.7)	75 (3)	2 (6.5)
	No	280 (94.3)	2387 (97)	29 (93.5)
History of PPCI	Yes	22 (7.4)	149 (6)	3 (9.7)
	No	274 (92.6)	2315 (94)	28 (90.3)
History of Stroke	Yes	37 (12.5)	106 (4.3)	5 (16.1)
	No	260 (87.5)	2358(95.7)	26 (83.9)
History of CABG	Yes	12 (4)	78 (3.2)	1 (3.2)
	No	285 (96)	2385 (96.8)	30 (96.8)
Direct admission to hospital	Yes	191 (62.6)	1852 (75.2)	26 (83.9)
	No	114 (37.4)	612 (24.8)	5 (16.1)
Anterior MI/LBBB	Yes	81 (26.6)	1942 (78.8)	10 (32.3)
	No	224 (73.4)	522 (21.2)	21 (67.7)
Reperfusion therapy	Primary PCI	123 (40.3)	1488 (60.4)	17 (54.8)
	Thrombolysis alone	63 (20.7)	665 (27)	9 (29)
	No reperfusion	119 (39)	311 (12.6)	5 (16.2)

HDL: high-density lipoprotein, LDL: low-density lipoprotein, LVEF: left ventricular ejection fraction, CHF: congestive heart failure, PPCI: primary percutaneous coronary intervention, CABG: coronary artery bypass surgery, Hb: hemoglobin, MI: myocardial infarction, Sys BP: systolic blood pressure, GFR_CKD: glomerular filtration rate- chronic kidney disease, CKMB: creatine kinase-MB. LBBB: left bundle branch block.

Table 5. Hazard ratio (HR) of factors affecting mortality in STEMI patients for the Cox PH model without and with frailty

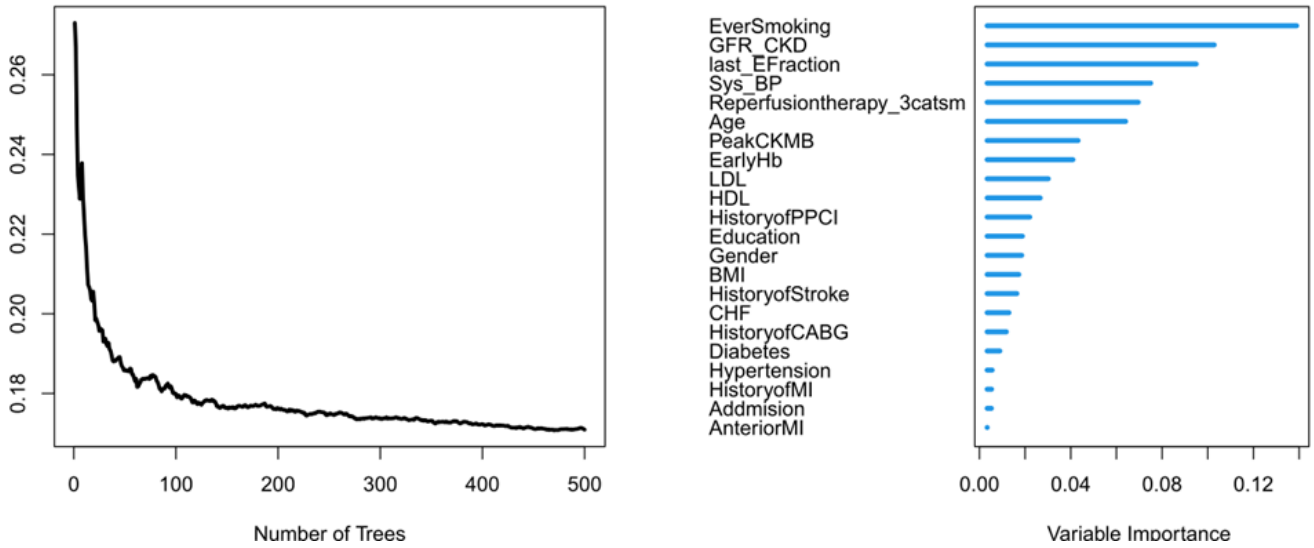
Factors		Cox PH without Frailty (95% CI)	Cox PH with Frailty (gamma distribution) (95% CI)
Age (year)		1.01 (1.00, 1.02)	1.01 (0.99, 1.02)
Gender		1.16 (1.00, 1.31)	1.21 (0.86, 1.69)
BMI (kg/m^2)		0.98 (0.96, 1.00)	0.97 (0.94, 1.01)
Ever Smoking		1.40* (1.27, 1.53)	1.46* (1.11, 1.93)
Diabetes		1.24 (0.99, 1.77)	1.27 (0.95, 1.70)
Hypertension		1.27 (0.98, 1.74)	1.28 (0.97, 1.69)
Admission		1.51* (1.09, 1.94)	1.50* (1.15, 1.97)
Education (Reference: Illiterate)	Under Diploma	0.69 (0.55, 1.13)	0.68 (0.51, 1.07)
	Upper Diploma	0.65 (0.38, 1.12)	0.62 (0.35, 1.09)
History of MI		0.80 (0.50, 1.18)	0.77 (0.60, 1.00)
History of CHF		0.90 (0.53, 1.66)	0.94 (0.64, 1.16)
History of Stroke		1.42* (1.13, 1.61)	1.45 (0.96, 1.62)
History of PPCI		1.38 (0.84, 2.50)	1.48 (0.87, 1.62)
History of CABG		1.10 (0.58, 2.27)	1.15 (0.79, 1.41)
Systolic blood pressure		0.98* (0.97, 0.98)	0.98* (0.98, 0.99)
LDL (mg/dL)		1.00 (0.99, 1.01)	1.00 (0.99, 1.01)
HDL (mg/dL)		0.99 (0.98, 1.01)	0.99 (0.98, 1.01)
Peak CKMB		1.00 (0.99, 1.01)	1.00 (0.99, 1.01)
LVEF		0.97* (0.95, 0.97)	0.96* (0.96, 0.97)
Early Hemoglobin		0.98 (0.91, 1.05)	0.98 (0.94, 1.01)
GFR_CKD		0.97* (0.96, 0.97)	0.96* (0.96, 0.97)
Reperfusion therapy (Reference: PPCI)	Thrombolysis	1.02 (0.77, 1.40)	1.01 (0.86, 1.17)
	No reperfusion	2.49* (1.99, 3.67)	2.71* (2.35, 2.63)
Anterior MI		1.40* (1.13, 2.04)	1.52* (1.26, 1.54)
Significance of the frailty variance		-	0.00205*

*statistical signification (p-value <0.05). HDL: high-density lipoprotein, LDL: low-density lipoprotein, LVEF: left ventricular ejection fraction, CHF: congestive heart failure, PPCI: primary percutaneous coronary intervention, CABG: coronary artery bypass surgery, Hb: hemoglobin, MI: myocardial infarction, Sys BP: systolic blood pressure, GFR_CKD: glomerular filtration rate- chronic kidney disease, CKMB: creatine kinase-MB.

Table 6. Comparison of evaluation criteria for each model

Models		Concordance Index	Brier Score	Mean AUC
Cox model		0.8164	0.05921	0.8898
Cox model with Frailty		0.891*	0.0458**	0.9134*
Survival SVR	**Linear Kernel**	0.8375	-	0.8857
	RBF Kernel	0.8489	-	0.8975
	Polynomial Kernel	0.8423	-	0.8921
	Cilinical Kernel	0.8352	-	0.8859
Random Survival Forest (testing)		0.8641	0.05318	0.8984

Survival support vector regression does not have access to the Brier score. Because only predicted a risk score, not a probability. * The highest overall C-index and mean AUC. ** The least Brier score. RBF: radial basis function kernel, or RBF kernel.



Figher2. Error rate and variable importance in the training dataset

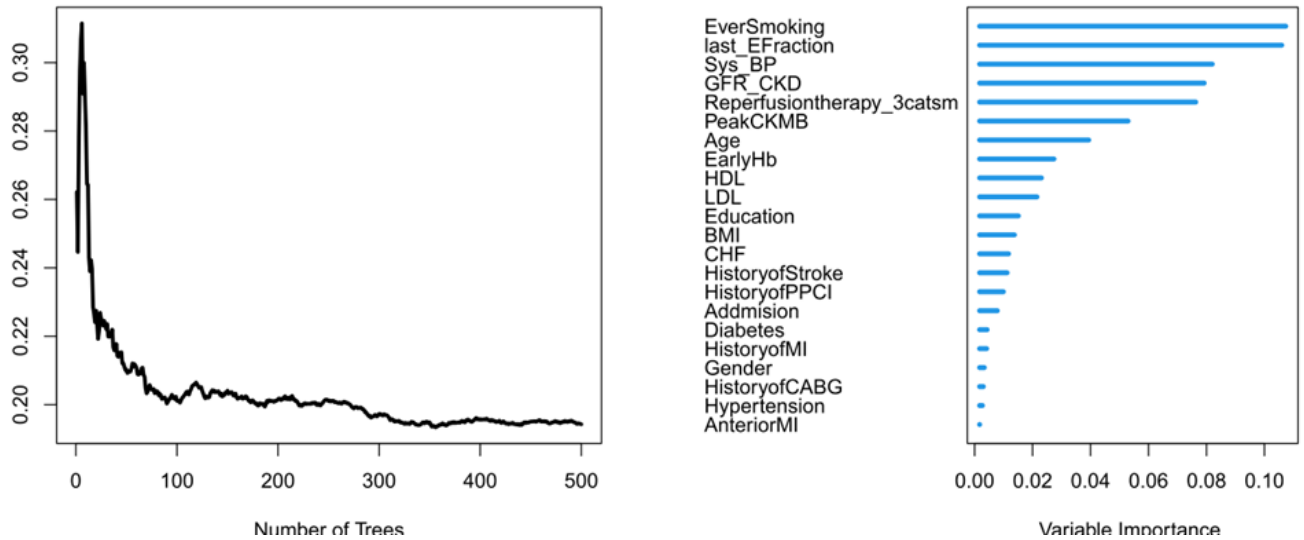


Figure 3. Error rate and variable importance in the test dataset

Discussion

The Cox PH regression, unlike machine learning models, considers the impact of censored records, such as participants whose hazard behavior is unknown because they were fired at the beginning of the data collection period (left censoring) or dropped from the study or sample (right censoring). The impact of clinical, behavioral, and demographic factors on predicted time to death has primarily been modeled in prior research employing the Cox PH regression. These investigations, however, made the assumption that there was no unobserved heterogeneity arising from correlated data that would affect the likelihood of the "hazard" occurring (48). To account for this violation, shared frailty correction has been implemented in more recent research (49). Support vector regression is a subset of data mining techniques that has the ability to work with high-dimensional data and also does not require a priori testing of events. This technique, with changes in the objective function and constraints, also has the ability to work with censored data such as survival data and is called survival support vector regression. On the other hand, survival tree is a new method for analyzing survival data, which aimed to divide individuals into groups that are homogeneous in terms of survival rate.

In this study, both RSF and SVR methods for survival analysis were considered and compared with the Cox PH model with and without frailty using the STEMI datasets. Performance improvements were partially significant for the Cox model with frailty when compared to the RSF and survival SVR models. Analysis of more intricate and nonlinear relationships between high-dimensional variables, such as genetic data, can be facilitated by machine learning techniques. The assumption of Cox proportional hazards model is that the hazard function for each individual is proportional to the basine hazard, h0(t). This assumption suggests that the covariate vector determines the hazard function in its entirety; nevertheless, this assumption may be broken by unseen covariates. The issue is that the unobservable individual-level characteristics cause heterogeneity in the data. The assumption of proportional hazards is broken since our model is unable to account for individual-level factors; this issue can be resolved by employing a frailty model. One way to explain the unaccounted-for heterogeneity is through frailty models (50). Prior research on machine learning's application to CVD risk prediction primarily used data from China, Europe, and the United States. We conducted a registry-based cohort study at Imam-Ali Hospital in Kermanshah, Iran, which comprised consecutive STEMI patients from 2016 to 2019. Both with and without the frailty correction, which accounts for the constraints of the Cox estimate, we estimated patient survival using the Cox PH regression. A comparison was made between the results of the two Cox PH regression models, one with and one without the frailty correction. The results demonstrate that the frailty adjustment improved the performance of the basic Cox PH model, which was statistically significant in our study. Cox PH, either with or without shared frailty, is frequently utilized in the healthcare industry in general and in heart failure investigations in particular. Abrahantes and Legrand (51) provide an overview of time-to-event models and analyses of different frailty multiplier distributions. In the same institution, Gasperoni, and Ieva (52) used the Cox PH with frailties that are typical of heart failure patients. Toenges and Mütze (53) exploited shared frailty to explain the relationship between the two analyzed events hospitalization for heart failure and coronary artery mortality. Reese, Roman (54) modeled the time to cardiovascular illness among American Indians who were monitored for up to 20 years in a recent publication. Using shared frailty, they adjusted for participant family ties. In the beginning, the Cox PH regression was created to predict the time until death for actuarial calculations (55). Since then, a lot of studies (49, 56, 57) have employed it for survival analyses where mortality was the target variable.

The largest Australian study to create machine learning-based risk prediction models for both cardiovascular mortality and hospitalization for Ischemic Heart Disease (IHD) was conducted by H. Wang et al. They compared various machine learning algorithms, such as survival methods (SVM, Cox regression, random survival forest, and neural network) and traditional classification methods (SVM, logistic regression, random forest, and random forest). The optimal model for cardiovascular mortality, after examining various data re-sampling techniques, ratios, and classification approaches, was a Cox survival regression with an L1 penalty, utilizing a re-sampled case/non-case ratio of 0.3 through the under-sampling of non-cases.Harrel's and Uno's concordance indices for this model were 0.900 and 0.898, respectively. At a re-sampled case/non-case ratio of 1.0, a Cox survival regression with L1 penalty was the most effective model for IHD hospitalization, with Uno's and Harrel's concordance indices of 0.711 and 0.718, respectively (58). For survival analysis, Kim et al. used two machine learning techniques, RSF and SVM, and evaluated how well they predicted outcomes using the two datasets. They, after comparing the three approaches, it was determined that the Cox model, RSF, and SVM performed better with mixed datasets than with unmixed datasets. The C-index and 1-year time-

dependent areas under the curve for the Cox model were 0.644, 0.6 respectively (38). Based on the feature importance analysis, we were able to determine which characteristics contributed most to the prediction of a higher risk of cardiovascular death. Ever smoking, systolic blood pressure, LVEF, GFR and reperfusion therapy are the most important variables that predict mortality in STEMI patients. In other studies, these variables could be independently associated with a 1-year mortality of STEMI patients (59-62). Our results show how methodological research could advance healthcare by developing better models. The frailty correction is a less prevalent survival analysis technique than the Cox PH regression, which is possibly the most widely employed in medicine. In the STEMI domain, there are even fewer scientific and practical similarities between these two models. The results indicate that advancing frailty models beyond the basic Cox proportional hazards regression offers benefits in estimating survival probabilities for STEMI patients and potentially other chronic conditions, addressing the unobserved heterogeneity in grouped data. The Cox PH assumptions should be further examined in future research, and suggestions for improvements to the model's performance should be made. Comparing the outcomes of the present models to other algorithms, like Kernel learning, might be another future direction.

Assuming that this expanded collection is now possible due to the widespread use of health information systems in hospitals that offer thorough and longitudinal data, we selected the 22 most significant features for this paper—more than are often included in other studies. Future studies can extend these techniques to other chronic diseases and simulate the frailty correct utilizing the competing risks of several events, including death and readmissions. Because only one hospital's patients were included in the study, its external validity is constrained. Additional datasets from other hospitals should be reproduced to improve the results' generalizability. Furthermore, hospitalized patients may not be representative of the typical STEMI patient because they are typically more seriously unwell. Therefore, more acute individuals are more affected by the outcomes. Future studies should integrate community and hospital patient records to provide a more comprehensive picture of patients' health. This will make it possible to apply the frailty model to different dataset stratifications.

Acknowledgments

We wish to thank all the Imam-Ali Hospital staff, especially the Cardiovascular Research Center, Dr. Mahdi Nalini, without their contribution, this work would not have been accomplished.

Funding: The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not- for- profit sectors.

Ethics approval: All methods were carried out in accordance with the relevant guidelines and regulations. Based on the latest version of the Declaration of Helsinki (2013), we have complied with all the issues, and accordingly, ethics approval was obtained from the Research Ethics Committee of Kermanshah University of Medical Sciences (ID IR.KUMS.REC.1400.272) in July 2021.This study used deidentifed data from Imam- Ali Hospital and followed all ethical considerations instituted by the providers of the data. Informed consent was obtained from all participants for the use of their data for research purposes at the time of their screening test.

Conflict of interests: None declared.

Authors' contribution: All the authors contributed to the design of the work, participated in the study, and drafted the manuscript. MM searched for the studies and collected and analyzed the data. AK, SM, and MT were in charge of data management and contributed to the analysis and interpretation of the data for the work. MR acted as a guarantor. All authors read and approved the final manuscript.

References

1. Salehi N, Motevaseli S, Janjani P, et al. Reperfusion therapy and predictors of 30-day mortality after ST-segment elevation myocardial infarction in a university medical center in western Iran. Arch Iran Med 2021; 24: 796-803.
2. Roth GA, Mensah GA, Johnson CO, et al. Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study. J Am Coll Cardiol 2020; 76: 2982-3021.
3. Lee J, Reyes F, Budzikowski AS. Acute ST-segment elevation myocardial infarction (STEMI). In: Mieszczanska, H., Budzikowski, A. (eds) Cardiology Consult Manual. Springer, Cham 2018; pp: 107-19.
4. Bradley EH, Sipsma H, Brewster AL, Krumholz HM, Curry L. Strategies to reduce hospital 30-day risk-standardized mortality rates for patients with acute myocardial infarction: a cross-sectional and longitudinal survey. BMC Cardiovasc Disord 2014; 14: 1-8.
5. Chandrashekhar Y, Alexander T, Mullasari A, et al. Resource and infrastructure-appropriate management of

ST-segment elevation myocardial infarction in low-and middle-income countries. Circulation 2020; 141: 2004-25.
6. Wilson PWF, D'Agostino RB Sr. No one size fits all: Scoring risk of in-hospital death after myocardial infarction. J Am Coll Cardiol 2016; 68: 636-8.
7. Ibanez B, James S, Agewall S, et al. Scientific Document Group; et al. 2017 ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation: The Task Force for the management of acute myocardial infarction in patients presenting with ST-segment elevation of the European Society of Cardiology (ESC). Eur Heart J 2018; 39: 119-77.
8. Salerno S, Li Y. High-dimensional survival analysis: Methods and applications. Annu Rev Stat Appl 2023; 10: 25-49.
9. Singh R, Mukhopadhyay K. Survival analysis in clinical trials: Basics and must know areas. Perspect Clin Res 2011; 2: 145-8.
10. Wang Y, Flowers CR, Li Z, Huang X. CondiS: A conditional survival distribution-based method for censored data imputation overcoming the hurdle in machine learning-based survival analysis. J Biomed Inform 2022; 131: 104117.
11. Montaseri M, Yazdani Cherat J, Espahbodi F, Mousavi SJ. Five-year survival rate in hemodialysis patients Attending Sari Imam Khomeini Hospital. J Mazandaran Uni Med Sciences 2013; 23: 78-85. [in Persian]
12. Shivaswamy PK, Chu W, Jansche M. A support vector approach to censored targets. InSeventh IEEE international conference on data mining. In Seventh IEEE international conference on data mining (ICDM) 2007; pp: 655-60.
13. Montaseri M, Charati JY, Espahbodi F. Application of parametric models to a survival analysis of hemodialysis patients. Nephrourol Mon 2016; 8: e28738.
14. Štajduhar I, Dalbelo-Bašić B, Bogunović N. Impact of censoring on learning Bayesian networks in survival modelling. Artif Intell Med 2009; 47: 199-217.
15. LeBlanc M, Crowley J. Survival trees by goodness of split. J Am Stat Assoc 1993; 88: 457-67.
16. Stensrud MJ, Hernán MA. Why Test for Proportional Hazards? JAMA 2020; 323: 1401-2.
17. Thygesen K, Alpert JS, Jaffe AS, et al. Fourth universal definition of myocardial infarction (2018). J Am Coll Cardiol 2018; 72: 2231-64.
18. Morrow DA, Antman EM, Charlesworth A, et al. TIMI risk score for ST-elevation myocardial infarction: a convenient, bedside, clinical score for risk assessment at presentation: an intravenous nPA for treatment of infarcting myocardium early II trial substudy. Circulation 2000; 102: 2031-7.
19. Detection NCEPEPo, Evaluation, Adults ToHBCi. Third report of the national cholesterol education program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III) final report. Circulation 2002; 106: 3143-421.
20. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. Ann Intern Med 2009; 150: 604-12.
21. Demissie S, LaValley MP, Horton NJ, Glynn RJ, Cupples LA. Bias due to missing exposure data using complete-case analysis in the proportional hazards regression model. Stat Med 2003; 22: 545-57.
22. Barzi F, Woodward M. Imputations of missing values in practice: results from imputations of serum cholesterol in 28 cohort studies. Am J Epidemiol 2004; 160: 34-45.
23. Baneshi MR, Talei A. Does the missing data imputation method affect the composition and performance of prognostic models? Iran Red Crescent Med J 2012; 14: 31-6.
24. Rubin DB. Multiple imputation for survey nonresponse. New York: John Wiley & Sons 1987; pp: 7.
25. Rubin DB. Inference and missing data. Biometrika 1976; 63: 581-92.
26. Van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. Stat Med 1999; 18: 681-94.
27. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. J Stat Softw 2011; 45: 1 - 67.
28. Ambler G, Omar RZ, Royston P. A comparison of imputation techniques for handling missing predictor values in a risk model with a binary outcome. Stat Methods Med Res 2007; 16: 277-98.
29. Spruance SL, Reid JE, Grace M, Samore M. Hazard ratio in clinical trials. Antimicrob Agents Chemother 2004; 48: 2787-92.
30. Yin H, Yang X, Peng R. Generalized accelerated failure time frailty model for systems subject to imperfect preventive maintenance. Math Probl Eng 2015; 2015: 908742.
31. Diou N. Multi-parametric copula estimation based on moments method under censoring. PhD thesis. Université de mohamed kheider biskra; Faculty of exact science, Science of nature and life, Departmebt of mathematics 2022.

32. Gorfine M, Zucker DM. Shared frailty methods for complex survival data: A review of recent advances. Annu Rev Stat Appl 2023; 10: 51-73.
33. Montaseri M, Rezaei M, Khayati A, Mostafaei S, Taheri M. Survival parametric modeling for patients with heart failure based on Kernel learning. BMC Med Res Methodol 2025; 25: 7.
34. Clayton D, Cuzick J. Multivariate generalizations of the proportional hazards model. Journal of the Royal Statistical Society: Series A (General) 1985; 148: 82-108.
35. Munda M, Legrand C. A diagnostic plot for guiding the choice of the frailty distribution in clustered survival data. Université catholique de Louvain, Institute of Statistics, Biostatistics; Discuss. Pap. 2013; 37.
36. Van Belle V, Pelckmans K, Van Huffel S, Suykens JA. Support vector methods for survival analysis: a comparison between ranking and regression approaches. Artif Intell Med 2011; 53: 107-18.
37. Adham D, Abbasgholizadeh N, Abazari M. Prognostic factors for survival in patients with gastric cancer using a random survival forest. Asian Pac J Cancer Prev 2017; 18: 129-34.
38. Kim H, Park T, Jang J, Lee S. Comparison of survival prediction models for pancreatic cancer: Cox model versus machine learning models. Genomics Inform 2022; 20: e23.
39. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. Institute of Mathematical Statistics (IMS); Annals of Applied Statistics 2008; 2: 841-60.
40. Ishwaran H, Kogalur UB. Consistency of random survival forests. Stat Probab Lett 2010; 80: 1056-64.
41. Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995; 20: 273-97.
42. Khan FM, Zubek VB. Support vector regression for censored data (SVRc): a novel tool for survival analysis. In2008 Eighth IEEE International Conference on Data Mining 2008; pp: 863-8. IEEE.
43. Herbrich R, Graepel T, Obermayer K. Large margin rank boundaries for ordinal regression. MIT Press, Cambridge, MA. MR1820960; Advances in large margin classifiers 2000; pp: 115-32.
44. Daemen A, De Moor B. Development of a kernel function for clinical data. Annu Int Conf IEEE Eng Med Biol Soc 2009; 2009:5913-7.
45. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996; 15: 361-87.
46. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics 2000; 56: 337-44.
47. Brier GW. Verification of forecasts expressed in terms of probability. Mon Weather Rev 1950; 78: 1-3.
48. Fisher LD, Lin DY. Time-dependent covariates in the Cox proportional-hazards regression model. Annu Rev Public Health 1999; 20: 145-57.
49. Ben-Assuli O, Ramon-Gonen R, Heart T, Jacobi A, Klempfner R. Utilizing shared frailty with the Cox proportional hazards regression: Post discharge survival analysis of CHF patients. J Biomed Inform 2023; 140: 104340.
50. Wienke A. Frailty models in survival analysis. 1st ed. New York; Chapman and Hall/CRC, CRC press 2010; pp: 56-7.
51. Abrahantes JC, Legrand C, Burzykowski T, et al. Comparison of different estimation procedures for proportional hazards model with random effects. Comput Stat Data Anal 2007; 51: 3913-30.
52. Gasperoni F, Ieva F, Paganoni AM, Jackson CH, Sharples L. Non-parametric frailty Cox models for hierarchical time-to-event data. Biostatistics 2020; 21: 531-544.
53. Toenges G, Mütze T, Jahn-Eimermacher A. A comparison of semiparametric approaches to evaluate composite endpoints in heart failure trials. Stat Med 2021; 40: 5702-24.
54. Reese JA, Roman MJ, Deen JF, et al. Subclinical atherosclerosis in adolescents and young adults and the risk of cardiovascular disease: The Strong Heart Family Study (SHFS). Nutr Metab Cardiovasc Dis 2022; 32: 1863-71.
55. Cox DR. Regression models and life-tables. J R Stat Soc, Ser B (Methodological) 1972; 34: 187-202.
56. Dokainish H, Teo K, Zhu J, et al. Global mortality variations in patients with heart failure: results from the International Congestive Heart Failure (INTER-CHF) prospective cohort study. Lancet Glob Health 2017; 5: e665-72.
57. Vazquez R, Bayes-Genis A, Cygankiewicz I, et al. The MUSIC Risk score: a simple method for predicting mortality in ambulatory patients with chronic heart failure. Eur Eur Heart J 2009; 30: 1088-96.
58. Wang H, Tucker WJ, Jonnagaddala J, et al. Using machine learning to predict cardiovascular risk using self-reported questionnaires: Findings from the 45 and Up Study. Int J Cardiol 2023; 386: 149-56.
59. Janjani P, Motevaseli S, Salehi N, et al. Predictors of 1-Year Major Cardiovascular Events after ST-Elevation

Myocardial Infarction in a Specialized Cardiovascular Center in Western Iran. The J Tehran Heart Cent 2022; 7:62-70.

60. Mishra S. A comparative study for time-to-event analysis and survival prediction for heart failure condition using machine learning techniques. J Electro Electromed Eng Med Inform 2022; 4: 115-34.
61. Angraal S, Mortazavi BJ, Gupta A, et al. Machine learning prediction of mortality and hospitalization in heart failure with preserved ejection fraction. JACC Heart Fail 2020; 8: 12-21.
62. Sujatha R, Chatterjee JM, Jhanjhi N, Tabbakh TA, Almusaylim ZA. Heart failure patient Survival analysis with multi kernel support vector machine. Intell Autom Soft Comput 2022; 32: 115-29.