Predictors of length of stay in the coronary care unit in patient with acute coronary syndrome based on data mining methods

Abbas Rezaianzadeh^a, Maryam Dastoorpoor^b, Majid Sanaei^c, Cirruse Salehnasab^d,
 Mohammad Javad Mohammadi^e, Ali Mousavizadeh^{c,*}

^a Department of Epidemiology, School of Health, Shiraz University of Medical Sciences, Shiraz, Iran

^b Department of Biostatistics and Epidemiology, Social Determinants of Health Research Center, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran

^c Social Determinants of Health Research Center, Yasuj University of Medical Sciences, Yasuj, Iran

^d School of Paramedicine, Yasuj University of Medical Sciences, Yasuj, Iran

^e Department of Environmental Health Engineering, School of Public Health AND Environmental Technologies Research Center (ETRC), Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran

ARTICLE INFO

Keywords:

Data mining
 Coronary care units
 Length of stay
 Acute coronary syndrome

ABSTRACT

Introduction: Assessing the possibility of patient discharge based on data-mining models is one of the common, user-friendly approaches to optimally exploit the limited capacity of hospital beds.

Objective: The aim of this study was to determine the predictors of length of stay (LOS) in cardiologic care wards developed and carried out based on data-mining approaches.

Methods: Data from 136 patient records were evaluated using data-mining analysis approaches including the Multilayer perceptron artificial neural network (MLP-ANN), Quick unbiased and efficient statistical tree (QUEST), support vector machines (SVM), classification and regression tree (CRT), Advanced decision tree (C5.0), Auto Classifier (AC) and Logistic Regression models.

Results: The median and mean LOS was 4 and 4.15 days (95% CI [3.99, 4.30]), respectively. Predictors are associated with increase in LOS (more than 4 days) were: the ST segment elevation myocardial infarction (STEMI) diagnosis at the time of referral, being in the 50–70 years old group, history of smoking, high blood lipids, history of hypertension, hypertension at the time of admission, and high serum troponin levels.

Conclusion: Using classical models to explain the predictors of outcome is inefficient when the number of predictors is high and sample size is low. Therefore, the analysis based on new data-mining approaches is a desirable alternative solution. Behavioral factors, especially smoking, are among the important factors in determining the long-term stay in the heart care ward.

1. Introduction

In many countries, non-communicable diseases have a rising trend for reasons such as increasing longevity, increasing and prolonging exposure to risk factors and changing lifestyles. Also, in relation to the global burden of disease, non-communication diseases ranked first since 1990.¹ Among non-communicable diseases, ischemic heart disease (IHD), in the developed countries, causes the highest rates of death, disability and financial burden compared to other diseases.² As World Health Organization (WHO) reports, at the end of the 2nd millennium, much of the healthcare budget in the developing countries is earmarked for cardiovascular diseases.³ Today, these diseases have become a social problem in Iran, with the resulting deaths accounting for 6.4 deaths per 10,000 person.^{4,5} In this study, ischemic heart disease refers to a series

of cardiovascular diseases whose common characteristic is imbalance in the need for and absorption of oxygen in the heart tissue. Cardiac ischemic diseases include two major groups called chronic coronary artery disease (often showing themselves as stable angina) and acute coronary syndromes (ACSs). The second group consists of patients with ST-segment elevation myocardial infarction (STEMI), unstable angina (UA), and non-ST-segment elevation myocardial infarction (NSTEMI).⁶

One of the important issues in the management of patients with acute coronary syndrome, especially in developing countries, is the management of cardiovascular beds. Considering the high costs of such beds, the optimal use and proper utilization of these beds is indispensable for providing services that are desirable and appropriate to patients' needs. Unnecessary hospitalization, lack of scientific protocols at the time of discharge, non-admission of referrals in the coronary care

* Corresponding author.

E-mail address: mousavizadeh.ali@yums.ac.ir (A. Mousavizadeh).

<https://doi.org/10.1016/j.cegh.2019.09.007>

Received 21 June 2019; Received in revised form 21 August 2019; Accepted 18 September 2019

Available online 25 September 2019

2213-3984/ © 2019 INDIACLEN. Published by Elsevier, a division of RELX India, Pvt. Ltd. All rights reserved.

units, the lack of identical guideline for the optimal length of stay of patients in these wards will lead into longer hospitalization and thus additional charge on patients as well as hospitals. The variety of factors affecting the length of stay (LOS) and the inseparable role of each of these factors has led the researchers to formulate the present study as the first academic approach to answer the above issues. On the other hand, under conditions such as the low sample size and the large number of predictors variables, due to the limitations of the use of classical model, the design of new data-mining approaches are essential, which is somewhat neglected. Although various studies have been conducted on the factors related to the LOS of patients in the coronary care unit (CCU), the intensive care unit (ICU) and other wards of hospitals with classical approaches, we focused on the issue to fill the gap on the lack of a study similar to that of data-mining analysis.^{7–15} Data-mining analysis refers to a variety of different approaches that can be categorized in two areas of classification approaches and clustering approaches. These approaches have been used frequently in various areas such as survival of kidney transplantation, survival of dialysis patients, survival of the heart and lung transplantation, organ failure rate, cancer diagnosis and prognosis, and classification of breast cancer.^{16–30} The approaches used in a recent study are a combination of approaches such as MPL-ANN, logistic, C5.0, CRT, QUIST, and SVM models. This study was designed to determine the factors associated with LOS in the CCU.

2. Methods

The data of this study included information about 136 patients with acute coronary syndrome diagnosed in CCU of Emam Sajjad Hospital in Yasuj, Kohgiluyeh and Boyer-Ahmad province, Southwest of Iran between 2014 and 2015. Due to the limited number of patients, all people were included in the study by census method. The data needed for this study were collected using a researcher-designed questionnaire and collected in two areas related to patient information and information related to the nature of the disease.

Patient information includes two three-level variables: Age group (Code 0 = less than 50 years old, Code 1 = 50–70 years and Code 2 = over 70 years), electrocardiogram (ECG) based diagnosis (Code 0 = Unstable angina; Code 1 = NSTEMI and Code 2 = STEMI) and 10 two-level variables (Code 0 = with attribute, Code 1 = no attribute) including gender, history of smoking, history of diabetes, history of hypertension, history of anti platelet medicines (aspirin, clopidogrel), blood lipid controller, Beta blocker drugs, history of taking Angiotensin converting enzyme (ACE) drugs, history of taking calcium channel blocker (CCB) medications, history of heart attack, history of Percutaneous coronary intervention (PCI) and Coronary artery bypass grafting (CABG) diagnostic and therapeutic interventions. Disease information includes six two-level variables (Code0 = yes; Code1 = none). The time between the start of the pain to referral to the emergency ward, the initial diagnosis based on changes in heartbeat, hypertension at the time of admission, the patient's initial troponin status, Creatine kinase-MB (CK-MB) status, and the need to receive streptokinase.

The outcome variable in this study was the LOS calculated and used in the model as a binary variable (less than 4 days and more than 4 days). The median and mean the LOS of patients in this study and some other studies were about 4 days. Since, there was no definitive cut-off point, this value was chosen as the cut-off point. Here after, conventional names of early or late discharge are used to refer to the levels of outcome variable of the study.

Data analysis was performed using SPSS modeler V.14. The data flow designed in the study consists of 10 process nodes including, the data file node, the type node, the filter node, the Auto Data Preparation node, the SVM node, the C5.0 node, the CRT node, the QUEST node and the Artificial Neural Network (NNT) node. In order to use the analysis and evaluation of the outputs of each node, two evaluation and analysis

nodes were used. We used the type node to determine the characteristics of the variables and prepare to enter the analytic nodes, the Filter node in order to control the input flow of the variables into the analysis nodes, and utilized the Auto Preparation node to manage the outlier data, the missing data in the predictor and target variables.

Artificial neural network (Multi-Layer Perceptron Artificial Neural Network)(ANN): As a classification technique, an artificial neural network is used to study the complex relationships between predictors and dependent variables like nonlinear and multiple functions.^{17,31,32} In this study, multilayer perceptron neural networks, with back propagation architecture of supervised learning type, were used. The neural network used in this study included an input layer (18 neurons), a bias node, an latent layer with 3 neurons and a bias, and the output layer containing a binary variable (0 and 1).

Support Vector Machine (SVM): Another technique for classifying data-mining is a backup vector machine.^{18,19,32} Using this technique has led to good modeling successes, which in some cases covers the neural network defects.

Model C5.0: This node is in the decision tree techniques and can represent a decision tree or correlation rules. The target field should be classified. This model makes it possible to classify more than two groups.^{32,33}

Classification and regression tree model (CRT): This node is a subset of decision tree techniques that, in addition to classifying, provides the possibility of prediction and association rules (if Then). Input and target fields in this model can be numeric or class. In the output tree, predictors are classified and represented as binary groups.^{32,33}

Quick unbiased and efficient statistical tree (QUEST) model: This node is a subset of the decision tree techniques and provides a binary classification for the decision tree. The performance of this model for long-standing trees is better than the CRT model. The input field can be continuous, but the output field must be a class. In the decision tree, it is possible to divide predictors into more than two classes which allows for its presentation.³²

Auto Classifier node (AC): This model is in the category of classifier models and it generates and compares a different number of models for binary outputs and allows the researcher to choose the best approach for analysis.³²

In the present study, this node was used after auto preparation data to extract the best proposed models and used in the process of data flow. The significance level for the tests was less than 0.05. The statistics used to measure the model's goodness-of-fit or the strength and accuracy of the model's predictions include the level below the rock curve and, if necessary, the Gini coefficient. The criterion for determining the best models were conventional cut off above 75% for the surface below the ROC curve.

After determining the predictive percentage of each variable for each model, a approach The adjusted predicted percent was used and Based on the scores obtained, various variables were ranked. By convention, predictive value of 3% and above is defined as acceptable predictors, predictive value of 2–3%, moderate predictor and less than 2% predicted as poor predictor.

Ethical issues (including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, redundancy, etc.) have been completely observed and considered by the authors. The Ethics Committee of Yasuj University of Medical Sciences approved the study protocol (Ethic Code: IR.YUMS.1394.19).

3. Results

The mean and standard deviation of patients' age were 60.47 ± 13.4 years, with a range of 32–90 years. 63.2% of the samples were male and 60% had history of smoking before admission. The median and mean the LOS of patients in the CCU was estimated to be 4

Table 1

The frequency distribution of predictors and outcome in the study samples.

Variable		LOS in the CCU N(%)		Total(%)	Variable		LOS in the CCU N(%)		Total(%)
		≤ 4 day	> 4 Day				≤ 4 day	> 4 Day	
Sex	Male	52(60.5)	34(39.5)	86(100)	History of ACE drugs	No	63(67.3)	33(35.7)	96(100)
	Female	34(68.0)	16(32.0)	50(100)		Yes	23(60.5)	15(39.5)	38(100)
	Total	86(63.2)	50(36.8)	136(100)		Total	86(63.2)	50(36.8)	136(100)
History of Smoking	Yes	47(57.3)	35(42.7)	82(100)	History of β -blockers drugs	Yes	18(64.3)	10(35.7)	28(100)
	No	39(72.2)	15(27.8)	54(100)		No	68(63.0)	40(37.0)	108(100)
	Total	86(63.8)	50(36.8)	136(100)		Total	86(63.2)	50(36.8)	136(100)
History of IHD	Yes	63(61.2)	40(38.8)	103(100)	History of antiplatelet drugs	Yes	20(69.0)	9(31.0)	29(100)
	No	23(69.7)	10(30.3)	33(100)		No	66(61.7)	41(38.3)	107(100)
	Total	86(63.2)	50(36.8)	136(100)		Total	86(63.2)	50(36.8)	136(100)
History of PCI or CABG	No	77(62.6)	46(37.4)	123(100)	Serum Ck-MB at the time of admission	High	59(60.2)	39(39.8)	98(100)
	Yes	9(69.2)	4(30.8)	13(100)		Normal	27(71.1)	11(28.9)	38(100)
	Total	86(63.2)	50(36.8)	136(100)		Total	86(63.2)	50(36.8)	136(100)
Serum troponin at the time of admission	High	43(55.1)	35(44.9)	78(100)	History of hypertension	Yes	39(58.2)	28(41.8)	67(100)
	Normal	42(73.7)	15(26.3)	57(100)		No	47(68.1)	22(31.9)	69(100)
	Total	85(63.0)	50(37.0)	135(100)		Total	86(63.2)	50(36.8)	136(100)
Hypertension at the time of admission	Normal	67(65.7)	35(34.3)	102(100)	Time interval between the onset of pain and the referral to hospital	≤ 30min	77(63.6)	44(36.4)	121(100)
	High	19(55.9)	15(44.1)	34(100)		> 30 min	9(60)	6(40)	15(100)
	Total	86(63.2)	50(36.8)	136(100)		Total	86(63.2)	50(36.8)	136(100)
Use of streptokinase in the CCU	Yes	56(67.5)	27(32.5)	83(100)	EKG Diagnosis at the time of admission	UA	21(95.5)	1(4.5)	22(100)
	No	30(56.6)	23(43.4)	53(100)		NSTEMI	14(56)	11(44)	25(100)
	Total	50(36.8)	86(63.2)	136(100)		STEMI	51(57.3)	38(42.7)	89(100)
History of Diabetes mellitus	Yes	18(62.1)	11(37.9)	29(100)	Age	Total	86(63.2)	50(36.8)	136(100)
	No	68(63.6)	39(36.4)	107(100)		30–49	23(71.9)	9(28.1)	32(100)
	Total	86(63.2)	50(36.8)	136(100)		50–69	34(51.5)	32(48.5)	66(100)
History of anti-hyperlipidemic drug	Yes	26(74.3)	9(25.7)	35(100)		70–90	29(76.3)	9(23.7)	38(100)
	No	60(59.4)	41(40.6)	101(100)		Total	86(63.2)	50(36.8)	136(100)
	Total	86(63.2)	50(36.8)	136(100)					
History of CCB drug	No	72(62.6)	43(37.4)	115(100)					
	Yes	14(66.7)	7(33.3)	21(100)					
	Total	86(63.2)	50(36.8)	136(100)					

and 4.15 days, respectively. Sixteen percent of the patients were hospitalized with U.D. diagnosis, 18% with NSTEMI and 65.5% of the patients with STEMI diagnosis.

LOS longer than four days for UA was 4.5% and 44% NSTEMI and 42.7% for STEMI patients. Eleven percent of the patients referred to the hospital between half and 3 h after the onset of pain and the rest referred less than half an hour after the pain. Nine point five percent of the patients had a history of PCI or CABG in their medical record. Fifty seven point three percent of the patients had serum troponin levels higher than normal. Thirty eight point nine percent of the patients needed streptokinase during hospitalization. Table 1 shows the distribution of the frequency of the LOS in the ward classified according to the primary variables.

Based on the area below the curve, and according to cut off determined as 75% and above, the best models were SVM, AC, Logistic, and C5.0 (Fig. 1).

Table 2 shows the accuracy indicators for the AUC and ROC for different models, along with the predictive percentage of each model and their variables. Of the 19 input variables of the model, 7 variables include: having a diagnosis of STEMI at the time of referral, the age group of 50–70 years, history of smoking, history of antihyperlipidemic, history of hypertension, hypertension at the time of admission, and high levels of serum troponin which were identified as strong predictors. Five variables, history of beta blocker intake, calcium channel blockers, need for streptokinase at the beginning of treatment, history of hyperlipidemia, and history of Type-2 diabetes were also determined as mid-range predictors of LOS in wards. The first seven variables (strong predictors) had 55.5% predictive power, five variables with moderate strengths had 11%, and seven weak variables had only 8.9% predictive power. All 19 input variables had 75.5% predictive power.

Data from Table 3 shows that behavioral indicators (smoking and seeking help) attributed 9.5% of predictive power, background (age and sex) 15%, medical and laboratory history 25%, and diagnostic-

therapeutic interventions during the hospitalization 26.92% of predictive power for long-term stay in the CCU.

4. Discussion

The results of the study indicate the predictive role of 7 variables among the 19 input variables in the model: Having a diagnosis of STEMI at the time of referral, the age group of 50–70 years, history of smoking, history of hyperlipidemia, history of hypertension, hypertension at the time of admission, and high serum troponin levels identified as potent predictors. The studies conducted by researchers has revealed that limited studies are carried out with a similar approach and the target population.

In a study with a data-mining approach conducted by Rezaei et al., with the purpose of identifying the role of three data-mining models in predicting the LOS of heart disease patients in the CCU, which was carried out on 4948 patients, the researchers used three approaches: C5.0, SVM and ANN. In this study, the outcome variable of the hospital stay were three-part (less than 5 days, 6–9 days and more than 10 days), and 36 variables were used as predictors. The accuracy of SVM, ANN, C5.0 models was 83.5, 53.9, and 96.4, respectively. According to the SVM model, researchers introduced 16 predictors on the patients' LOS in the hospitals: History of the use of anticoagulants, history of use of nitrate compounds, diagnosis at admission, diastolic blood pressure, cardiac file in echocardiography, disabilities, marital status, chest pain, gender, high-density lipoprotein (HDL), hemoglobin levels, smoking, insurance type, cholesterol, age, and changes in the ST part of the ECG were introduced as predictors.³⁴ EKG-based diagnosis and smoking history are common predictors for the LOS in the CCU in the present study and the other study. Of course, the difference in sample size of the two studies is one of the points that can be considered in justifying the differences between the two studies.

In a study by Heller et al., which evaluated the LOS after heart

Fig. 1. The ROC curves for SVM, AC, Logistic, C5.0, CRT and ANN models.

attack in 438 patients, the results showed that the mean hospitalization time was 13.6 days and 74% of patients stayed in the hospital for more than 10 days. The average LOS in the CCU was 4.5 days. The findings of this study indicated that the increase in CK enzymes, stroke with Q wave, stroke in the anterior vein, and the use of digoxin and nitrates during admission is accompanied with an increase in the LOS in the hospital with a regression model.⁷ Although the mean LOS of the patients in the CCU obtained in this study is similar to our study, but our results are not very consistent with the predictor of longer LOS in a ward. The difference between the predictors of the model and different analytical approaches could be among the most important justifications for this discrepancy. However, EKG-based diagnosis is one of the common predictors in the two studies.

In a study to investigate predictors of LOS in a post-stroke hospital

in Japan on 4113 patients between 1998 and 2003, Kinjo reported a mean LOS of 31.2 days and the effect of clinical variables (patient characteristics, severity of stroke, treatment and intra-hospital complications) predicted a 26-percent increase in LOS, a finding that contradicts the results of the present study that shows a 75% role for the predictive power of the variables entered into the model.³⁵

Laurencet et al. reviewed 370 patients with ACS diagnosis due to delays in the discharge of low risk patients and reported the most important reasons for delay in discharge is due to the need for additional testing, regular decrease in medication and admission during holidays.³⁶ Saczynski et al. studied the effect of reducing the LOS on mortality and reduction of hospitalization in 4184 patients between 1995 and 2005. The results of their study revealed that during the 10 year period of the study, hospitalization was reduced by one third from

Table 2
Power and degree of importance (%) of predictors used in different models.

Category	SVM	C5.0	LOGIT	CRT	ANN	QUEST	Score	Predictive percentage	Strength of predictor
EKG Diagnosis at the time of admission	9	41	13	31	9	10	86.03	14.40	Strong
Background: Age	16	9	12	21	11	32	78.27	13.04	Strong
Behavioral: History of Smoking	13	16	6	9	7	11	48.98	8.10	Strong
History of anti-hyperlipidemic drug	7	20	19	0.01	8	7	47.25	7.88	Strong
Hypertension at the time of admission	13	4	2	8	6	3	29.37	4.89	Strong
History of hypertension	3	0.01	4	4	7	11	22.02	3.67	Strong
Serum troponin at the time of admission	3	0.01	4	4	12	5	21.09	3.50	Strong
History of β -blockers	4	0.01	5	4	6	0.1	14.85	2.48	Moderate
History of CCB drug	0.1	0.01	5	4	8	2	13.94	2.30	Moderate
Use of streptokinase in the CCU	3	0.01	6	4	4	0.01	13.14	2.19	Moderate
History of hyperlipidemia	7	0.01	4	0.01	2	2	12.68	2.11	Moderate
History of Diabetes mellitus	0.001	8	2	4	2	0.1	12.00	2.00	Moderate
Background: Sex	0.1	0.01	3	4	6	3	11.77	1.96	Week
Serum Ck-MB at the time of admission	3	0.01	5	4	3	0.01	11.68	1.94	Week
Time interval between the onset of pain and the referral to hospital	4	0.01	2	0.01	2	3	9.04	1.50	Week
History of antiplatelet drugs	3	0.01	1	0.01	1	4	7.37	1.30	Week
History of PCI or CABG	1	2	3	0.01	2	1	6.93	1.15	Week
History of ACE drugs	3	0.01	2	0.01	2	1	6.58	1.09	Week
History of IHD	1	0.01	0.001	0.01	3	4	6.12	1.02	Week
Accuracy	96.32	76.47	75	72.06	71.32	75			
AUC	0.979	0.796	0.834	0.736	0.733	0.747			
GINI	0.958	0.592	0.668	0.472	0.466	0.494			

7.2 days to 5days. Younger patients, men and non-complicated patients have a shorter admission period.³⁷ Among other studies on the LOS in cardiac patients is the study by Li et al., showed that the median and mean LOS of patients with myocardial infarction was 13 and 14.6 days in 2001, and the same values in 2011 were 11 and 11.9 days, respectively.³⁸ On the other hand, Karabulut et al. attempted to determine the optimal LOS and the predictors of LOS for 267 patients admitted with diagnosis of STEMI. In this study, patients were classified according to the length of one-day, two-day, three-day and four-day and more and compared with respect to other complications. Killip classification, left ventricular ejection, multiple coronary artery disease, and diabetes are the most important predictors of LOS in the study.³⁹ Swaminathan et al., in a study on 33920 patients diagnosed with STEMI, classified and reported delay in hospital discharge as three short, medium and long term categories of 26.9%, 46.3% and 26.8%, respectively, arguing

that the long-term predictors associated with long-term discharge include high age, female gender, patients with cardiogenic shock and having to deal with several coronary arteries.⁴⁰ In a 2001 study on 296 CCU patients, Di Chiara et al. reported a percentage of STEMI, NSTEMI and Undefined cases of 65%, 30% and 5% respectively, which is consistent with the findings of the present study. The median LOS was 10 days, which was reported for STEMI, NSTEMI, and Undefined, respectively, as 9,10 and 11 days, respectively.⁴¹ Finally, in a study by Every et al. (1996) on 11,932 patients with myocardial infarction in 19 regional hospitals between 1988 and 1994, showed an LOS mean in the ward decreased from 8.5 days to 6 days in the final year (1994). The demographic and individual characteristics of the patients accounted for 6% of this decrease, and the hospital complications and treatment approaches for 27% of the rest.⁹

In studies conducted in Iran, Vejdani et al., in a study on 3330

Table 3
The association rules for C5.0, CRT and rule association models.

Model	Rule	Antecedent:if	Consequent: then LOS	Instances	Confidence
C5.0	1	diagnosis = 0	0 = less than 4 day	22	0.955
	2	diagnosis = 1, PCI_CABG = 0, bp = 1	0 = less than 4 day	18	0.722
	3	diagnosis = 1, PCI_CABG = 0, bp = 0	1 = more than 4 day	5	0.800
	4	diagnosis = 1, PCI_CABG = 1	1 = more than 4 day	2	1.000
	5	diagnosis = 2, smoke = 1, antilipid = 1, age = 0	0 = less than 4 day	13	0.615
	6	diagnosis = 2, smoke = 1, antilipid = 1, age = 0, dm = 1	1 = more than 4 day	12	0.677
	7	diagnosis = 2, smoke = 1, antilipid = 1, age = 0, dm = 0	1 = more than 4 day	3	1.000
	8	diagnosis = 2, smoke = 1, antilipid = 1, age = 2	1 = more than 4 day	19	0.733
	9	diagnosis = 2, smoke = 1, antilipid = 1	0 = less than 4 day	10	0.800
	10	diagnosis = 2, smoke = 0	0 = less than 4 day	32	0.719
CRT	Rule	Antecedent:if	Consequent: then LOS	Instances	Confidence
	1	Diagnosis = 0	0 = less than 4 day	14	1.000
	2	Diagnosis = 1,2, age = 0,1	0 = less than 4 day	38	0.737
Rule Association	3	Diagnosis = 1,2, age = 2	1 = more than 4 day	41	0.537
	Rule	Prediction = LOS Less than 4 day; if condition is	Condition Support (%)	Lift	confidence
	1	Smoke = 0, SK = 0, Anti lipid = 0, Diagnosis = 2, Age = 2	5.88	2.38	87.50%
	2	Smoke = 0, SK = 0, Diagnosis = 2, Age = 2, Pain duration = 0	5.88	2.38	87.50%
	3	Smoke = 0, SK = 0, Anti lipid = 0, Age = 2	8.82	2.27	83.3%
	4	Sex = 0, smoke = 0, Troponin = 0, Anti lipid = 0, Age = 2	8.82	2.27	83.3%
	5	Smoke = 0, DM = 0, Troponin = 0, Anti lipid = 0, Age = 2	8.82	2.27	83.3%
	6	Smoke = 0, DM, SK = 0, D_Anti_lipid = 0, Age = 2	8.82	2.27	83.3%
	7	Smoke = 0, Troponin = 0, D_BB = 0, D_Anti_lipid = 0, Age = 2	8.82	2.27	83.3%
	8	Smoke = 0, Troponin = 0, D_Anti_lipid = 0, Age = 2, CK_MB = 0	8.82	2.27	83.3%
	9	Smoke = 0, SK, Anti lipid = 0, Age = 2, CCB = 0	8.82	2.27	83.3%
	10	Smoke = 0, SK = 0, Anti lipid = 0, Age = 2, Pain duration = 0	8.82	2.27	83.3%

elderly patients, showed that the mean of patients' LOS in all wards and the CCU ward was 4.8 and 5.1 days, respectively which is consistent with our results. Regarding the factors related to the LOS in the hospital, there was no relationship between sex and LOS, but there was a direct relationship between the age and LOS in the hospital.⁴²

Also, Farajikhayavi, in a study reported an mean LOS in CCU ward, 4.15 and 4.22 days in public and private hospitals, respectively. In this study, a significant relationship was found between the type of stroke, the patient's blood pressure during admission (inverse relation) and age (direct relation).⁴³

Finally, although the low sample size and the large number of input predictors of the models are considered to be limitations of this study, the goodness-of-fit of the models and predictive capability of the data show the efficiency of data-mining models and algorithms at a time when classical approaches face analytical challenges.

5. Conclusion

No similar study was found particularly one that uses data-mining approaches. The important point in this matter is the simplicity and importance of exploiting these approaches in the field of medical that have gained considerable prominence over the last few decades. The next important point is attention to the factors that determine the delay in the discharge of patients from the CCU. The top factor here is smoking as a behavioral predictor. Clinical diagnosis along with behavioral tendency to smoking, history of hypertension, and eventually serum troponin marker are the most important predictors. Planning to control smoking and hypertension needs to be emphasized.

Ethics approval and consent to participate

Ethical issues (including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or sub-mission, redundancy, etc.) have been completely observed and considered by the authors. The Ethics Committee of Yasuj University of Medical Sciences approved the study protocol (Ethic Code: IR.YUMS.1394.19).

Funding

This work was funded by Yasuj University of Medical Sciences approved the study protocol (Grant number: 1916).

Declaration of competing interest

The authors declare that they have no competing interests.

Acknowledgements

The authors are grateful to all respondents in this study.

References

- Lopez AD, Murray CC. *The Global Burden of Disease*. 1998; 1998 1990–2020.
- Akhtar S. Ischemic heart disease. *Anesthesiol Clin*. 2006;24:461–485.
- Investigators WMPP. The World Health Organization MONICA Project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. *J Clin Epidemiol*. 1988;41:105–114.
- Shafe Shakori M. *Evaluate the Prevalence of Cardiovascular Disease in Iran* Proceedings Of the 2th Congress of Medical Research Club Young. Iran: Islamic Azad University Ardabil; 2001 [In Persian].
- Adriyani M. *Thrombolytic Therapy in Myocardial Infarction* Proceedings of the 15th Iranian Congress of Cardiology, Tehran, Iran. 2005; 2005.
- Malani PN. Harrison's principles of internal medicine. *J Am Med Assoc*. 2012;308:1813–1814.
- Heller R, Dobson A, Steele P, et al. Length of hospital stay after acute myocardial infarction. *Intern Med J*. 1990;20:558–563.
- Madsen EB, Hougaard P, Gilpin E, Pedersen A. The length of hospitalization after acute myocardial infarction determined by risk calculation. *Circulation*. 1983;68:9–16.
- Every NR, Spertus J, Fihn SD, Hlatky M, Martin JS, Weaver WD. Length of hospital stay after acute myocardial infarction in the Myocardial Infarction Triage and Intervention (MITI) Project registry. *J Am Coll Cardiol*. 1996;28:287–293.
- Eagle KA, Mulley AG, Skates SJ, et al. Length of stay in the intensive care unit: effects of practice guidelines and feedback. *Jama*. 1990;264:992–997.
- Rapoport J, Teres D, Zhao Y, Lemeshow S. Length of stay data as a guide to hospital economic performance for ICU patients. *Med Care*. 2003;41:386–397.
- Lilly CM, Cody S, Zhao H, et al. Hospital mortality, length of stay, and preventable complications among critically ill patients before and after tele-ICU reengineering of critical care processes. *Jama*. 2011;305:2175–2183.
- Wright S, Verouhis D, Gamble G, Swedberg K, Sharpe N, Doughty R. Factors influencing the length of hospital stay of patients with heart failure. *Eur J Heart Fail*. 2003;5:201–209.
- Gruenberg DA, Shelton W, Rose SL, Rutter AE, Socaris S, McGee G. Factors influencing length of stay in the intensive care unit. *Am J Crit Care*. 2006;15:502–509.
- Herring A, Wilper A, Himmelstein DU, et al. Increasing length of stay among adult visits to US emergency departments, 2001–2005. *Acad Emerg Med*. 2009;16:609–616.
- Yahav I, Shmueli G. Predicting potential survival rates of kidney transplant candidates from databases with existing allocation policies. *Proceedings of the 5th INFORMS Workshop on Data Mining and Health Informatics (DM-HI 2010)*, Austin, TX. 2010; 2010.
- Silva Á, Cortez P, Santos MF, Gomes L, Neves J. Rating organ failure via adverse events using data mining in the intensive care unit. *Artif Intell Med*. 2008;43:179–193.
- Palaniappan S, Awang R. Intelligent heart disease prediction system using data mining techniques computer systems and applications. 2008 AICCSA 2008 IEEE/ACS International Conference on. IEEE; 2008:108–115.
- Oztekin A, Delen D, Kong ZJ. Predicting the graft survival for heart-lung transplantation patients: an integrated data mining methodology. *Int J Med Inform*. 2009;78:e84–e96.
- Obenshain MK. Application of data mining techniques to healthcare data. *Infect Control Hosp Epidemiol*. 2004;25:690–695.
- Lisboa PJ, Vellido A, Tagliaferri R, et al. Data mining in cancer research [application notes]. *IEEE Comput Intell Mag*. 2010;5:14–18.
- Li L, Tang H, Wu Z, et al. Data mining techniques for cancer detection using serum proteomic profiling. *Artif Intell Med*. 2004;32:71–83.
- Labib NM, Malek MN. Data mining for cancer management in Egypt case study: childhood acute lymphoblastic leukemia. *World Acad Sci Eng Technol*. 2005;8:309–314.
- Kusiak A, Dixon B, Shah S. Predicting survival time for kidney dialysis patients: a data mining approach. *Comput Biol Med*. 2005;35:311–327.
- Kharya S. *Using Data Mining Techniques for Diagnosis and Prognosis of Cancer Disease*. 2012; 2012 arXiv preprint arXiv:12051923.
- Gupta S, Kumar D, Sharma A. Data mining classification techniques applied for breast cancer diagnosis and prognosis. *Indian J Comput Sci Eng (IJCSE)*. 2011;2:188–195.
- Delen D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. *Artif Intell Med*. 2005;34:113–127.
- Cao X, Maloney KB, Brusie V. Data mining of cancer vaccine trials: a bird's-eye view. *Immunome Res*. 2008;4:7.
- Bellazzi R, Zupan B. Predictive data mining in clinical medicine: current issues and guidelines. *Int J Med Inform*. 2008;77:81–97.
- Antonie M-L, Zaiane OR, Coman A. Application of data mining techniques for medical image classification. *Proceedings of the Second International Conference on Multimedia Data Mining*. Springer-Verlag; 2001:94–101.
- Liu H, Motoda H. *Feature Selection for Knowledge Discovery and Data Mining*. Springer Science & Business Media; 2012.
- Trybula WJ. Data mining and knowledge discovery. *Annu Rev Inf Sci Technol*. 1997;32:197–229.
- Xu G, Zong Y, Yang Z. *Applied Data Mining*. CRC Press; 2013.
- Rezaei Hachesu P, Ahmadi M, Alizadeh S, Sadoughi F. Use of data mining techniques to determine and predict length of stay of cardiac patients. *Healthc Inf Res*. 2013;19:121–129.
- Kinjo K, Sato H, Nakatani D, et al. Predictors of length of hospital stay after acute myocardial infarction in Japan. *Circ J*. 2004;68:809–815.
- Laurent M-E, Girardin F, Rigamonti F, et al. Early discharge in low-risk patients hospitalized for acute coronary syndromes: feasibility, safety and reasons for prolonged length of stay. *PLoS One*. 2016;11:e0161493.
- Saczynski JS, Lessard D, Spencer FA, et al. Declining length of stay for patients hospitalized with AML: impact on mortality and readmissions. *Am J Med*. 2010;123:1007–1015.
- Li Q, Lin Z, Masoudi FA, et al. National trends in hospital length of stay for acute myocardial infarction in China. *BMC Cardiovasc Disord*. 2015;15.
- Karabulut A, Cakmak M, Uzunlar B, Bilici A. What is the optimal length of stay in hospital for ST elevation myocardial infarction treated with primary percutaneous coronary intervention? *Cardiol J*. 2011;18:378–384.
- Swaminathan RV, Rao SV, McCoy LA, et al. Hospital length of stay and clinical outcomes in older STEMI patients after primary PCI: a report from the National Cardiovascular Data Registry. *J Am Coll Cardiol*. 2015;65:1161–1171.
- Di Chiara A, Chiarella F, Savonitto S, et al. Epidemiology of acute myocardial infarction in the Italian CCU network: the BLITZ study. *Eur Heart J*. 2003;24:1616–1629.
- Vejdani M, Saffari SE, Borabadi M, Vejdani M, Nejatizadehgan-Eidgahi Z, Foji S. Assessment the factors affecting on length of hospitalization stay for elderly of Vaseie hospital in Sabzevar (2014) using count regression models. *J Sabzevar Univ Med Sci*. 2016;22:1105–1116.
- Faraji Khayavi F. Assessment of patient satisfactory ,day bed cost and length of stay in ccu of Tehran general hospitals: Persian language. *Health Manag J*. 1998;4:51–58.