Summary

Introduction. Risk stratification is nowadays crucial when estimating the patient’s prognosis in terms of treatment outcome and it also helps in clinical decision making. Several risk assessment models have been developed to predict short-term outcomes in patients with acute coronary syndrome. This study was aimed at developing an outcome prediction model for patients with acute coronary syndrome submitted to percutaneous coronary intervention using data mining approach. Material and Methods. A total of 2030 patients hospitalized for acute coronary syndrome and treated with percutaneous coronary intervention from December 2008 to December 2011 were assigned to a derivation cohort. Demographic and anamnestic data, clinical characteristics on admission, biochemical analysis of blood parameters on admission, and left ventricular ejection fraction formed the basis of the study. A number of machine learning algorithms available within Waikato Environment for Knowledge Discovery had been evaluated and the most successful was chosen. The predictive model was subsequently validated in a different population of 931 patients (validation cohort), hospitalized during 2012. Results. The best prediction results were achieved using Alternating Decision Tree classifier, which was able to predict in-hospital mortality with 89% accuracy, and preserved good performance on validation cohort with 87% accuracy. Alternating Decision Tree classifier identified a subset of 6 attributes most relevant to mortality prediction: systolic and diastolic blood pressure, heart rate, left ventricular ejection fraction, age, and troponin value. Conclusion. Data mining approach enabled the authors to develop a model capable of predicting the in-hospital outcome following percutaneous coronary intervention. The model showed excellent sensitivity and specificity during internal validation.

Key words: Data Mining; Treatment Outcome; Acute Coronary Syndrome; Risk Assessment; Mortality

Original study
Originalni naučni rad
UDK 616.12-08-036.8
DOI: 10.2298/MPNS1506157S

DATA MINING APPROACH FOR IN-HOSPITAL TREATMENT OUTCOME IN PATIENTS WITH ACUTE CORONARY SYNDROME

PROCENA INTRAHOSPITALNOG ISHODA LEČENJA PACIJENATA SA AKUTNIM KORONARNIM SINDROMOM METODOM ISTRAŽIVANJA PODATAKA

Miroslava SLADOJEVIĆ1, Milenko ČANKOVIĆ2, Snežana ČEMERLIĆ1, Bojan MIHAJLOVIĆ1, Filip ADIĆ1 and Milana JARAKOVIĆ1

Sažetak


Ključne reči: Istraživanje podataka; Ishod lečenja; Akutni koronarni sindrom; Procena rizika; Mortalitet

Corresponding Author: Dr Miroslava Sladojević, Institut za kardiovaskularne bolesti Vojvodine, 21204 Sremska Kamenica, Put dr Goldmana 4, E-mail: miroslava.sladojevic@gmail.com
Introduction

Acute coronary syndrome (ACS) is defined as a spectrum of life-threatening conditions, representing a substantial proportion of all acute hospitalizations [1]. Mortality rates have declined during recent years [2] as a result of therapeutic options which have been expanded largely due to the optimization of timely reperfusion and innovations in pharmacological therapy.

Risk assessment is needed to guide triage and key management decisions in contemporary practice. Several risk models have been developed to predict short-term outcomes in patients with ACS. Global Registry of Acute Coronary Events (GRACE) and Thrombolysis in Myocardial Infarction (TIMI) risk scores are the most popular and valuable ACS prediction models, recommended by current guidelines [3,4].

Knowledge discovery and data mining have achieved numerous successful applications in the domain of medicine over the last 30 years. However, this practice has not been widely adopted in some fields, such as cardiology, regardless of the potential benefits. This study was aimed at developing an in-hospital outcome prediction model based on data mining approach for ACS patients submitted to percutaneous coronary intervention (PCI).

Material and Methods

Data were collected retrospectively. A set of patient-related data was obtained from the hospital information system of the Institute for Cardiovascular Diseases of Vojvodina, situated in Sremska Kamena, Serbia. The Institutional Review Board approved the study and waived the need for informed consent.

A total of 2030 patients hospitalized for ACS and submitted to PCI between December 2008 and December 2011 were assigned to a derivation cohort. Validation cohort contained 931 ACS patients hospitalized during 2012 also submitted to PCI. All patients were examined by an experienced cardiologist immediately after admission.

ACS is defined as “any group of clinical symptoms compatible with acute myocardial ischemia”, which includes unstable angina (UA) and myocardial infarction, with or without ST-segment elevation according to the American Heart Association [5].

All patients underwent an invasive strategy {primary PCI for STEMI (ST elevation myocardial infarction)/urgent PCI for NSTEMI (non-ST elevation myocardial infarction) and UA}, within two hours upon admission to hospital. Coronary stenting directly, or followed by balloon angioplasty, was performed where eligible. After the procedure, the patients were followed in the intensive coronary unit until stabilization.

The initial set of features including demographic and anamnestic data (age, gender, history of hypertension, diabetes, hyperlipidemia requiring treatment, smoking habits, alcohol consumption), clinical characteristics on admission (systolic and diastolic blood pressure, and heart rate), biochemistry and biohumoral response on admission (hemoglobin, troponin, urea, creatinine), and left ventricular ejection fraction (LVEF) formed the basis of the study.

The echocardiographic examination was performed by a Vivid 7 (GE Medical Systems, Horten, Norway) with a phased-array 3.5-MHz transducer. Echocardiographic examination was performed after primary PCI in STEMI patients mostly within 24 hours upon admission to hospital. In the patients with NSTEMI and UA, echocardiographic examination was performed before the invasive strategy. Echocardiographic examination was performed in a standard manner [6].

In-hospital treatment outcome was defined as the in-hospital all-cause mortality. We defined all-cause mortality as death from any cause.

Using Waikato Environment for Knowledge Analysis (WEKA), an open source data mining tool [7], various machine learning algorithms have been applied in order to build a predictive model and the most successful was chosen. Several of these algorithms are of interest when it comes to the research presented in this study (listed by their WEKA denominations): ADTree, RandomTree, RandomForest, J48, J48graft, IB1, Naïve Bayes and meta Cost Sensitive Classifier. Ten-fold cross validation was used for model validation. Both the accuracy and area under the receiver operating curve (AUROC) parameters were used for evaluation of algorithm performances [8-10]. After the best performing algorithm had been selected, the key attributes were extracted. Records containing only those attributes comprised the reduced “training set”. Cost sensitive classification was explored as an additional methodology to enhance results. Additional validation of the developed predictive model was performed on the “test set” containing 931 patients (validation cohort), hospitalized during 2012.

Statistical analysis was performed using SPSS version 17. Continuous variables were presented as mean ± SD or median (25th percentile – 75th percentile). Comparisons between groups were analyzed by an unpaired t-test or a Mann-Whitney test. Differences were considered significant at p < 0.05.

Results

A total of 2030 patients (aged 61.29 ± 11.70 years, 66.79% males), diagnosed with ACS and treated with PCI from December 2008 to December 2011 were assigned to a derivation sample. Of these, 1495 (73.64%) were STEMI patients, 474 (23.35%) were NSTEMI patients and 61 (3.01%) were UA patients.

Abbreviations

ACS – acute coronary syndrome
GRACE – Global Registry of Acute Coronary Events
TIMI – Thrombolysis in Myocardial Infarction
PCI – percutaneous coronary intervention
UA – unstable angina
AMI – acute myocardial infarction
STEMI – ST elevation myocardial infarction
NSTEMI – non-ST elevation myocardial infarction
LVEF – left ventricular ejection fraction
WEKA – Waikato Environment for Knowledge Analysis
AUROC – accuracy and area under the receiver operating curve
ADTree – Alternating Decision Tree

Statistical analysis was performed using SPSS version 17. Continuous variables were presented as mean ± SD or median (25th percentile – 75th percentile). Comparisons between groups were analyzed by an unpaired t-test or a Mann-Whitney test. Differences were considered significant at p < 0.05.

Results

A total of 2030 patients (aged 61.29 ± 11.70 years, 66.79% males), diagnosed with ACS and treated with PCI from December 2008 to December 2011 were assigned to a derivation sample. Of these, 1495 (73.64%) were STEMI patients, 474 (23.35%) were NSTEMI patients and 61 (3.01%) were UA patients.
Demographic and anamnestic data are shown in Table 1. The survivors were significantly younger. One-third of patients from derivation cohort were women, and they had frequent lethal outcome (10.7%). The patients with already diagnosed hypertension accounted for more than two-thirds of the sample, but hypertension was not associated with the fatal outcome \((p = 0.767)\). Diabetes mellitus was present in 24.7%, while hyperlipidemia occurred in 32.7% (there is no correlation with fatal outcome) of patients. Smoking habit and alcohol consumption was recorded in 42.9% and 1.2% of patients, respectively.

The patient’s profile based on clinical examination during admission was associated with lethal outcome (Table 1). The patients with lethal outcome were found to have higher heart rate on admission. Biochemical analysis of blood parameters on admission is shown in Table 2.

The patients with lethal outcome tended to have lower hemoglobin values than the survivors, while urea and creatinine were higher in the patients with lethal outcome. Troponin was more frequently positive in the patients with mortality. The survivors had higher LV EF (Table 1).

In-hospital mortality was 7.73% and 6.64% in the derivation and the validation cohort, respectively.

Machine Learning Algorithm Evaluation

A number of algorithms available within WEKA were evaluated. The results, in terms of accuracy, AUROC and confusion matrix achieved by different algo-

### Table 1. Demographic and anamnestic data, clinical profile and echocardiographic parameters of patients from Derivation Cohort

<table>
<thead>
<tr>
<th>Demographic and anamnestic data</th>
<th>Total (n=2030, 100%)</th>
<th>Mortality = YES (n=157, 7.74%)</th>
<th>Mortality = NO (n=1873, 92.27%)</th>
<th>(p)-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (year)/Starost (godine)</td>
<td>61 (53-71)</td>
<td>71 (62-77)</td>
<td>60 (53-70)</td>
<td>&lt; 0.0005</td>
</tr>
<tr>
<td>Gender/Pol</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male/Muški</td>
<td>1356 (66.8%)</td>
<td>85 (6.3%)</td>
<td>1271 (93.7%)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Female/Ženski</td>
<td>674 (33.2%)</td>
<td>72 (10.7%)</td>
<td>602 (89.3%)</td>
<td></td>
</tr>
<tr>
<td>History of hypertension/Od ranije podatak o hipertenziji</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No/Ne</td>
<td>651 (32.1%)</td>
<td>45 (6.9%)</td>
<td>606 (93.1%)</td>
<td>0.381</td>
</tr>
<tr>
<td>Yes/Da</td>
<td>1376 (67.9%)</td>
<td>112 (8.1%)</td>
<td>1264 (91.9%)</td>
<td></td>
</tr>
<tr>
<td>Diabetes mellitus/Dijabetes melitus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No/Ne</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dietary modifications/Modifikovana ishrana</td>
<td>6.9%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oral antidiabetics/Oralni antidiabetici</td>
<td>13.2%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insulin/-Insulin</td>
<td>4.4%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combination therapy/Kombinovana terapija</td>
<td>0.1%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hyperlipidemia requiring treatment/Hiperlipidemija koja zahteva lečenje</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No/Ne</td>
<td>1364 (67.3%)</td>
<td>115 (8.4%)</td>
<td>1249 (91.6%)</td>
<td>0.117</td>
</tr>
<tr>
<td>Yes/Da</td>
<td>663 (32.7%)</td>
<td>42 (6.3%)</td>
<td>621 (93.7%)</td>
<td></td>
</tr>
<tr>
<td>Smoking habit/Pušenje</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No/Ne</td>
<td>901 (44.4%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes/Da</td>
<td>869 (42.9%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking history/Ranije pušenje</td>
<td>257 (12.7%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol consumption/Konzumiranje alkohola</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No/Ne</td>
<td>2003 (98.8%)</td>
<td>157 (7.8%)</td>
<td>1846 (92.2%)</td>
<td>0.252</td>
</tr>
<tr>
<td>Yes/Da</td>
<td>24 (1.2%)</td>
<td>24 (100%)</td>
<td>0 (0%)</td>
<td></td>
</tr>
<tr>
<td>Systolic blood pressure/Sistolni krvni pritisak (mmHg)</td>
<td>140 (120-160)</td>
<td>120 (90-140)</td>
<td>140 (120-160)</td>
<td>&lt; 0.0005</td>
</tr>
<tr>
<td>Diastolic blood pressure/Dijastolni krvni pritisak (mmHg)</td>
<td>80 (70-100)</td>
<td>70 (60-85)</td>
<td>85 (70-100)</td>
<td>&lt; 0.0005</td>
</tr>
<tr>
<td>Heart rate (bat/min.)/Srčana frekvencija (otkucaja/min.)</td>
<td>80.15 ± 22.90</td>
<td>109.80 ± 19.63</td>
<td>78.22 ± 21.84</td>
<td>0.002</td>
</tr>
<tr>
<td>LV EF (%)</td>
<td>52.00 (45.00-57.00)</td>
<td>38 (28-47)</td>
<td>52 (45-58)</td>
<td>&lt; 0.0005</td>
</tr>
</tbody>
</table>

Values are reported as mean ± SD or median (25-75) percentile or (%) where indicated

Vrednosti su izražene kao srednja ± SD ili medijana (25-75) procenat ili (%) gde je indikovano
The application of data mining when solving the issue of predicting in-hospital outcome in ACS has not been explored extensively in the literature [11-14]. This study includes ACS patients after an invasive strategy (primary PCI for STEMI/urgent PCI for NSTEMI and UA) and focuses on in-hospital mortality. Our model, based on 6 attributes, is able to estimate in-hospital patient outcomes. Hynek Kruzk et al. designed and verified a predictive model of hospital mortality in STEMI patients based on clinical data, and presented results of an experimental evaluation of different machine learning methods. The best performed classifications were based on logistic regression and on simple Bayesian networks [15]. The best prediction results in this study, as well as in those done by Sladojevic et al. [16] and Vaithiyanathan V. et al. [17], were achieved by decision tree classifiers although the applications of data mining in this domain are not always restricted to classification [18].

LVEF, one of the main indicators of left ventricular systolic function, is already known as a key prognostic factor of mortality described in many studies [19,20]. Lilian P. Souza et al. demonstrated that LVEF was the single independent variable, significantly related to in-hospital heart failure in patients with a first STEMI and an LVEF ≤ 0.45 on admission [21]. The result obtained in this study indicates that all patients who died had a low LVEF, with average of 38%.

Similar to the results of our study, approximately two-thirds of patients were men, but this proportion decreased with age [22]. In a study covering more than 2-decade-long experience on more than 8000 patients with acute myocardial infarction (AMI), Robert J Goldberg et al. compared the outcome in the patients under 55 years of age with the outcome in the patients 55 to 64 years old, and concluded that the latter were 2.2 times more likely to die during hospitalization for AMI, whereas patients aged 65 to 74, 75 to 84, and ≥85 years were at 4.2, 7.8, and 10.2 times greater risk of dying, respectively [23].

In GRACE and TIMI risk scores, with the former considered the strongest one in evaluating the risk of adverse outcomes in the patients with initial presentation of ACS, age, heart rate and systolic blood pressure appeared as the most important variables [3, 4]. The same parameters occur in the predictive model developed in the study presented. Eric Boersma et al. analyzed the relation between the baseline characteristics and the 30-day incidence of death and the composite of death or myocardial (re)infarction in almost 9500 patients with ACS without persistent ST-segment elevation. More than 20 significant predictors for mortality and for the composite end point were identified. The most important were age, heart rate, systolic blood pressure, ST-segment depression, signs of heart failure, and elevated cardiac enzymes [24]. In our study, elevated troponin values appeared as one of six attributes most relevant to mortality prediction. Cardiac troponin I levels provide useful prognostic information and enable the early identification of patients with an increased risk of death in ACS, as emphasized by Antman EM et al. [25].

### Discussion

The proposed model might prove to be very helpful in decision-making and optimization of the treatment strategy in selected high-risk patients with acute coronary syndrome submitted to percutaneous coronary intervention. Although the presented study was done on a large number of patients and faithfully illustrates the actual state of patient population of the region where it was developed, the authors believe the model may prove to be useful in the daily work of clinicians worldwide.

Finally, the model for prediction of in-hospital mortality was developed, but fatal outcome cannot be excluded after discharge, so a follow-up of this study is preferable.

### Table 2. Biochemical analysis of blood parameters in the derivation sample

<table>
<thead>
<tr>
<th>Biochemical analysis of blood parameters on admission</th>
<th>Total (n=2030, 100%)</th>
<th>Mortality = YES (n=157, 7.74%)</th>
<th>Mortality = NO (n=1873, 92.27%)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hemoglobin (g/l)/Hemoglobin (g/l)</td>
<td>120.4 (131.0-152.0)</td>
<td>109.0 (106.5-141.5)</td>
<td>139.0 (125.5-150.5)</td>
<td>&lt;0.0005</td>
</tr>
<tr>
<td>Urea (mmol/l)/Urea (mmol/l)</td>
<td>6.40 (5.10-8.20)</td>
<td>14.70 (9.75-20.10)</td>
<td>6.40 (5.00-8.20)</td>
<td>&lt;0.0005</td>
</tr>
<tr>
<td>Creatinine/Creatinin (µmol/l)</td>
<td>93.0 (81.0-108.0)</td>
<td>142.0 (115.0-196.5)</td>
<td>95.0 (79.5-107.5)</td>
<td>&lt;0.0005</td>
</tr>
<tr>
<td>Troponin positive/Pozitivan troponin (%)</td>
<td>51.87 63.69 50.99 &lt;0.0005</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note**: The P-values are calculated using the appropriate statistical tests to compare the biochemical parameters between the groups.
References


6. Lang RM, Bierer M, Devereux RB, Flachskampf FA, Foster E, Pellikka PA, et al. Recommendations for chamber quantification: a report from the American Society of Echocardiography’s Guidelines and Standards Committee and the Chamber Quantification Writing Group, developed in conjunction with the European Association of Echocardiography, a branch of the European Society of Cardiology. J Am Soc Echocardiogr. 2005;18(12):1440-63.


