Computers in Biology and Medicine 170 (2024) 107953

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Interpretable machine learning for in-hospital mortality risk prediction in patients with ST-elevation myocardial infarction after percutaneous coronary interventions

Karina Iosephovna Shakhgeldyan [a,b], Nikita Sergeevich Kuksin [c], Igor Gennadievich Domzhalov [a], Vladislav Yurievich Rublev [a,b,*], Boris Izrajlevich Geltser [a]

[a] Far Eastern Federal University, School of Medicine and Life Science, 10 Ajax Bay, Russky Island, 690922, Vladivostok, Russia
[b] Vladivostok State University, Institute of Information Technology, Gogolya St. 41, 690014, Vladivostok, Russia
[c] Far Eastern Federal University, Institute of Mathematics and Computer Technology, 10 Ajax Bay, Russky Island, 690922, Vladivostok, Russia

ARTICLE INFO

Keywords:
Predictive machine learning models
Risk factors
Categorizing continuous variables
SHapley additive additive exPlanations (SHAP)

ABSTRACT

Background and objective: Despite the constant improvement of coronary heart disease (CHD) diagnostics and treatment methods it remains one of the main causes of death in most countries around the world. And myocardial infarction with ST segment elevation on the electrocardiogram (STEMI) still is one of the most dangerous clinical variants of CHD. This study aims to develop an explainable machine learning model for in-hospital mortality (IHM) risk prediction in STEMI patients after myocardial revascularization by percutaneous coronary intervention (PCI).
Methods: A single-center observational retrospective study was conducted, enrolling 4677 electronic medical records of patients with STEMI after PCI, which were analyzed using statistical analysis and machine learning methods. A pool of potential IHM predictors was identified, and prognostic models were developed and validated based on multivariate logistic regression, random forest, and stochastic gradient boosting methods at two stages of hospital treatment: during the initial physicians examination in the emergency department and immediately after PCI surgery. To explain the IHM prognosis, threshold values of IHM risk factors were determined using 3 grid search methods for optimal cut-off points, calculating centroids and SHapley Additive exPlanations (SHAP).
Results: IHM prognostic models were developed using clinical and functional status data of STEMI patients during two stages of hospital treatment. The IHM prediction accuracy according to the first scenario was AUC = 0.85, and according to the second - AUC = 0.9. Predictors identified and validated in the models were converted into risk factors. Models whose parameters were risk factors demonstrated high forecast accuracy (AUC = 0.87), with the best model formed using the SHAP method.
Conclusions: For the forecast result interpretation risk factors obtained by categorizing continuous variables can be used by assessing the impact of the latter on the end point using the SHAP method.

1. Introduction

Cardiovascular disease continues to be one of the most pressing health problems in most countries around the globe [1]. Coronary heart disease (CHD) holds one of the leading positions in the cardiovascular diseases mortality structure, amounting to around 20 % of all deaths [2]. The most dangerous clinical form of CHD is myocardial infarction (MI) with electrocardiogram ST segment elevation (STEMI). One of the most common and effective treatments for STEMI is myocardial revascularization through percutaneous coronary intervention (PCI), which should be performed as soon as possible from the onset of disease [3]. Despite the PCI technologies improvement, in-hospital mortality (IHM) after it in emergency indications cases remains high and varies from 4 up to 7 %.

Various scales are used to stratify the risk of IHM in clinical practice relying on the clinical data analysis, results of instrumental and laboratory studies. The most well-known risk measurement tools among them are TIMI, GRACE, PAMI and CADILLAC scales [4–9]. Application

* Corresponding author. Far Eastern Federal University, School of Medicine and Life Science, 10 Ajax Bay, Russky Island, 690922, Vladivostok, Russia.
E-mail addresses: carinash@vvsu.ru (K.I. Shakhgeldyan), kuksin.ns@dvfu.ru (N.S. Kuksin), igor@domzhalov.ru (I.G. Domzhalov), dr.rublev.v@gmail.com (V.Y. Rublev), boris.geltser@vvsu.ru (B.I. Geltser).

https://doi.org/10.1016/j.compbiomed.2024.107953
Received 2 November 2023; Received in revised form 22 December 2023; Accepted 1 January 2024
Available online 2 January 2024

of these scales allows you to determine the optimal treatment option and maximize its effectiveness. In real practice physicians most frequently use the GRACE scale (Global Registry of Acute Coronary Events), which was developed by Cox regression methods, and in recent years several studies describe ideas about its improvement [10–13]. Generally, its structure is supplemented with new variables to create fresh predictive models based on the classical predictors kit. At the same time, more complex machine learning (ML) algorithms, including random forest (RF) and stochastic gradient boosting (SGB) are widely used for IHM models, allowing to achieve higher quality parameters [14–17]. However, new predictors inclusion is often not supported by the prognosis results explanation and the degree of their influence assessment on the end point, which limits their effective utilization in clinical practice.

The aim of the study was to test a new approach for IHM risk factors verification in patients with MI after PCI using SHapley Additive exPlanations (SHAP) method and to develop interpretable prognostic models on this basis.

2. Methods

2.1. Study population cohort

We conducted a single-center observational retrospective study, within the framework of which we analyzed case histories data of patients who were treated at the Regional Vascular Center of Primorsky Regional Clinical Hospital No. 1 in Vladivostok from 2015 to 2021. The study was approved by the Ethics Committee of School of Medicine, Far Eastern Federal University. The total number of patients in the cardiac intensive care unit was 12528, among which 7851 cases were excluded. The inclusion criteria for patient enrollment were confirmed STEMI and PCI during the first day of hospital treatment. Exclusion criteria were non-ST elevation myocardial infarction, unconfirmed STEMI or absence of indications for PCI.

Thus, the study included case histories of 4677 patients who underwent invasive coronary angiography followed by transluminal balloon angioplasty with stenting of infarct-related arteries on the first day of hospital treatment. 30-day mortality (IHM) after PCI was fixed in 318 (6.8 %) patients.

2.2. Data collection

On the first day of hospital treatment, patients' clinical and functional status was assessed by 136 factors (Table 1). In addition to the results of physicians objective study, demographic and anamnestic data, clinical blood test indicators were analyzed: levels of erythrocytes (RBC), leukocytes (WBC), lymphocytes (LYM), neutrophils (NEUT), eosinophils (EOS), hemoglobin (HB), platelets (PLT) and thrombocrit (PCT) along with international normalized ratio (INR), thrombin time (TI), prothrombin index (PTI), activated partial thromboplastin time (aPTT), fibrinogen levels (Fb), serum creatinine (Cr) and glucose (Gl) concentrations. Postoperative echocardiographic study included the determination of both transverse (LA1) and longitudinal (LA2) dimensions of the left and right (RA1 and RA2) atria, the end systolic (ESD) and diastolic (EDD) dimensions of the left ventricle (LV), the ejection fraction (EF) of the LV by the Teicholz method and mean pulmonary artery pressure (MPAP). Besides, a few calculated indicators were also evaluated: relative left ventricle myocardial mass index (LVMI), relative thickness index (RTI) of the left ventricle posterior wall (LVPW) and body mass index (BMI).

Study end point was represented in the form of categorical binary attribute ("absence" or "development") by the IHM (from all causes) indicator in STEMI patients after PCI.

Final dataset was formed by extraction of all necessary data from the hospital medical information system. Encountered dataset gaps were not filled with synthetic values and were further excluded from the analysis. All outliers were analyzed for correctness and changes were made according to the primary data.

Table 1
Patients clinical and functional characteristics.

Predictor		Group 1 (n = 318)	Group 2 (n = 4357)	OR (95 %) CI	p-value
Gender: F, abs. (%)	318	142 (44.65)	1332 (30.5)	1.8 [1.5; 2.3]	<0.000001
Age, years	318	71 (63; 78)	62 (55; 69)	–	<0.000001
Height, cm	287	168 (164; 174)	170 (165; 176)	–	0.000001
SBP, mmHg	317	110 (90; 130)	130 (120; 150)	–	<0.000001
HR, bpm	316	86 (72; 100)	72 (65; 80)	–	<0.000001
Cr, μmol/l	265	130 (96; 193.3)	97 (81; 114.8)	–	<0.000001
AHF T.Killip class					
I	318	71 (22.33 %)	2726 (62.6 %)	0.17 [0.13; 0.23]	<0.000001
II	318	58 (18.2 %)	867 (19.9 %)	0.9 [0.67; 1.20]	0.508052
III	318	66 (20.75)	479 (11)	2.1 [1.6; 2.8]	<0.000001
IV	318	123 (38.7)	269 (6.18)	9.6 [7.4; 12.4]	<0.000001
III-IV	318	189 (59.4)	748 (17.2)	7.1 [5.6; 9]	<0.000001
LV EF, %	174	46.5 (38; 54.8)	56 (50; 61)	–	<0.000001
MPAP, mmHg	170	35 (28.25; 46)	28 (25; 30)	–	<0.000001
NEUT, %	175	81.3 (75.75; 86.5)	66.7 (59.1; 74.9)	–	<0.0001
PCT, %	215	0.22 (0.17; 0.28)	0.2 (0.16; 0.24)	–	0.0012
EOS, %	176	0.1 (0.00; 0.3)	0.9 (0.3; 1.9)	–	<0.000001
Glu, mmol/l	189	7.9 (6.3; 10.31)	5.8 (5.1; 7)	–	<0.000001
AF, abs. (%)	318	129 (40.57 %)	772 (17.69 %)	3.2 [2.51; 4.02]	<0.000001
T2DM, abs. (%)	318	99 (31.13 %)	831 (19.05 %)	1.9 [1.50; 2.46]	<0.000001
CKD, abs. (%)	316	83 (26.1 %)	677 (15.5 %)	1.97 [1.5; 2.6]	<0.000001

Abbreviations: SBP - systolic blood pressure, HR - heart rate, Cr - creatinine, LV EF - left ventricular ejection fraction, MPAP- mean pulmonary artery pressure, PCT – thrombocrit, AF - atrial fibrillation, T2DM - type 2 diabetes mellitus, CKD - chronic kidney disease.

Due to urgency hospitalization status for all STEMI patients, the significant part of laboratory tests were absent, which served as the basis for division of IHM prognosis models development considering different stages of hospital treatment: immediately initial physician examination upon emergency department patients admission and after PCI, taking into account laboratory and instrumental studies results, obtained directly after the operation.

2.3. Study design

The predictor selection and models development process consisted of 4 stages (Fig. 1). During the first one potential predictors list was formed by exertion of intergroup comparison tests on deceased and surviving patients with STEMI after PCI (Table 1). The indicators were presented as median values (Me) and interquartile intervals (Q1; Q3) for continuous variables and occurrence frequencies for categorical ones. Normal distribution compliance was assessed by the Kolmogorov-Smirnov test. Since the indicator's distribution differs from the normal, nonparametric statistical methods were used, including the Mann-Whitney test for continuous variables, and the $\chi 2$ for categorical ones. Odds ratios (OR) and their 95 % CIs were calculated by Fisher's exact test. Differences were considered statistically significant at p-value <0.01.

During the second stage, we considered two scenarios for IHM forecast evaluation. The first one involved unfavorable outcome

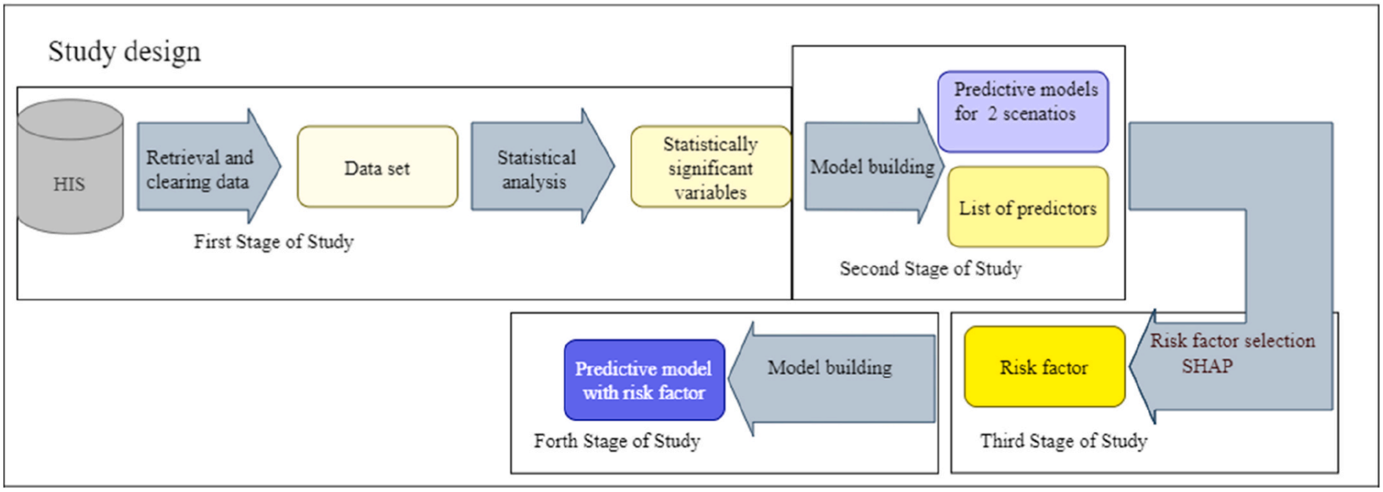

Fig. 1. Study design.

probability determination immediately after patient hospital admission and initial physician examination. This approach's effective implementation was partially limited by the complexity of the urgent fulfillment of the required range of laboratory and instrumental studies, leading to reduction of analyzed indicators set. The second scenario was designed for data collection during the first 24 h of hospitalization, making it possible to use a wider range of diagnostic tests. However, in this scenario, the number of examined patients from the IHM group decreased by 53 people as a result of a lethal outcome during the first hours after PCI. Thus, two variants of IHM prognostic models were developed, differing in the set of laboratory and instrumental data and the number of patients in the first group.

In the first scenario, we used the 4 most accessible GRACE scale predictors for initial physician examination: patients age, heart rate (HR), systolic blood pressure (SBP) and acute heart failure (AHF) class according to T. Killip. In the second IHM prognosis scenario, the basic model included 5 GRACE predictors: patient age, HR, SBP, T. Killip AHF class and Cr concentration. To improve the forecast accuracy, the basic model predictors set was expanded in step-by-step fashion, progressively including new indicators in its structure, followed by model quality assessment. Models were developed utilizing multivariate logistic regression (MLR), RF and SGB methods. Their quality was evaluated by 6 metrics: area under the ROC curve (AUC), sensitivity (Sen), specificity (Spec), F1-score, positive predictive value (PPV), negative predictive value (NPV). Prognostic significance of the predictor was confirmed if its inclusion in the model results in increased AUC. SHapley Additive exPlanations (SHAP) [18] were used to assess the predictors influence degree on the multivariate models end point.

During the third stage, variables included in the best predictive models were transformed into IHM risk factors. As the optimal risk factor thresholds, we considered unique variable values with a step *(max-min)/100*, besides that for their predictive abilities assessment 4 methods were used: OR maximum value, minimum p-value determined by the $\chi 2$ test, the highest AUC value in the univariate LR model and equidistant distance between the centroids of comparison groups [19]. It should be noted that intervals both greater and less than the threshold value were considered as potential IHM risk factors. In addition to the optimal thresholds obtained from isolated variables, we considered estimation of the SHAP influence obtained by multivariate SGB model in the previous step.

During the fourth stage, utilizing obtained risk factors, training, cross-validation and final testing of IHM prediction models was carried out, among which all presented predictors were in a categorical form, making it possible to assess not only the IHM probability, but also to explain the result.

2.4. ML algorithms and models building

Training, cross-validation and final testing of developed models was performed according to the following algorithm (Fig. 2). The complete dataset was divided into 2 samples: for training and cross-validation (80 %) and for final testing (20 %). The training and cross-validation procedure was performed using the stratified kFolders method on 10 samples. The averaged quality metrics AUC, Sen, and Spec were used to select the best model, select predictors, and select optimal hyperparameters. The cutoff threshold for calculating Sen and Spec was determined by looking for a balance between them. For the RF and SGB models, the hyperparameter selection method was used by grid search. For final testing, the best models with optimal parameters and hyperparameters were trained at 80 % and tested on a subgroup for final testing. For quality metrics confidence assessment, the procedure was repeated 100 times, performing the initial division randomly.

Data analysis and model development was carried out in open source Python, version 3.9.16.

3. Results

3.1. Final included cohort

The study included 4675 patients aged 26–93 years with a median of 63 years and 95 % CI [62; 63], of which 318 (6.8 %) people died within 30 days after PCI. The majority (90 %) of deaths occurred within the first 7 days post-surgery, 6 % died within 10–20 days, and 4 % within 20–30 days. Wherein, 53 (16.7 %) patients died in the first hours after the operation, which made it possible to consider them only in the first IHM prognosis scenario.

Intergroup analysis of demographic, clinical, laboratory and instrumental parameters showed that most of them, including all GRACE scale predictors (age, heart rate, SBP, T. Killip AHF class and Cr), have statistically significant differences (Table 1). Among the deceased, older, shorter stature persons, and females prevailed (OR = 1.8). It should be noted that the first group of patients was characterized by the presence of AHF 3–4 T. Killip class (OR = 7.1), lower SBP and LV EF, increased HR, higher levels of MPAP, Cr, NEUT and EOS. Also, they were more likely to have type 2 diabetes mellitus (DM2) (OR = 1.9), atrial fibrillation (AF) (OR = 3.2), and chronic kidney disease (CKD) (OR = 1.97). A more detailed intergroups comparison between surviving and deceased patients described by Table A1 in Appendix A.

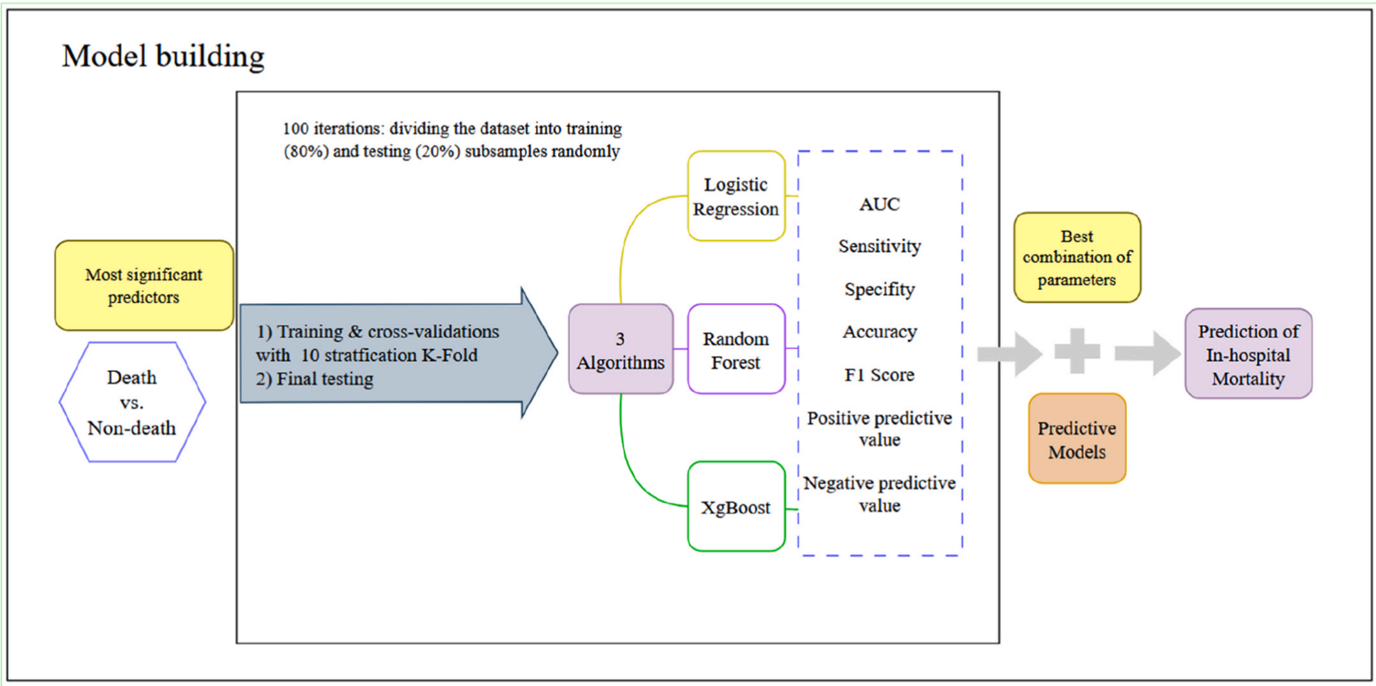

Fig. 2. Diagram of predictive models training and validation.

3.2. Training and validation of models

IHM predictive models quality metrics based on 4 continuous form predictors (age, HR, SBP, and AHF T. Killip class) for the first scenario are presented in Table 2.

Analysis results demonstrated good predictive accuracy of models based on a limited set of clinical indicators obtained in the hospital during the first contact between patient and physician. At the same time, the MLR and SGB models differed in higher quality metrics than ML. The predictors influence on the endpoint was assessed on the example of the best SGB model by the SHAP method (Fig. 3). The largest influence on IHM was deployed by the SBP indicator (SHAP-value - 2.92 c.u.). Less significant factors were the patients age (SHAP-value - 1.74 c.u.) and HR (SHAP-value - 1.7 c.u.), with the least effect on IHM by AHF T. Killip class (SHAP-value - 1.4 c.u.).

We have built 3 IHM forecast models according to the second scenario. The first of them included the sum of GRACE scale scores, as the

Table 2
Accuracy estimation of predictive IHM models according to the first scenario.

Metrics	Cross-validation			Final testing		
	MLR	SGB	RF	MLR	SGB	RF
AUC	0.854 [0.842; 0.866]	0.853 [0.84; 0.866]	0.839 [0.826; 0.851]	0.853 [0.805; 0.9]	0.853 [0.807; 0.9]	0.837 [0.789; 0.884]
Sen	0.79 [0.768; 0.781]	0.782 [0.754; 0.81]	0.767 [0.729; 0.805]	0.787 [0.694; 0.88]	0.783 [0.684; 0.882]	0.765 [0.664; 0.866]
Spec	0.781 [0.77; 0.791]	0.785 [0.769; 0.801]	0.764 [0.734; 0.793]	0.781 [0.752; 0.81]	0.785 [0.752; 0.818]	0.765 [0.726; 0.804]
PPV	0.209 [0.198; 0.22]	0.211 [0.197; 0.225]	0.193 [0.177; 0.209]	0.209 [0.184; 0.235]	0.212 [0.184; 0.24]	0.194 [0.161; 0.226]
NPV	0.981 [0.979; 0.983]	0.98 [0.978; 0.983]	0.978 [0.975; 0.981]	0.98 [0.972; 0.989]	0.98 [0.971; 0.989]	0.978 [0.969; 0.987]
F-score	0.345 [0.329; 0.36]	0.347 [0.328; 0.366]	0.322 [0.3; 0.345]	0.345 [0.309; 0.38]	0.349 [0.309; 0.386]	0.324 [0.276; 0.368]

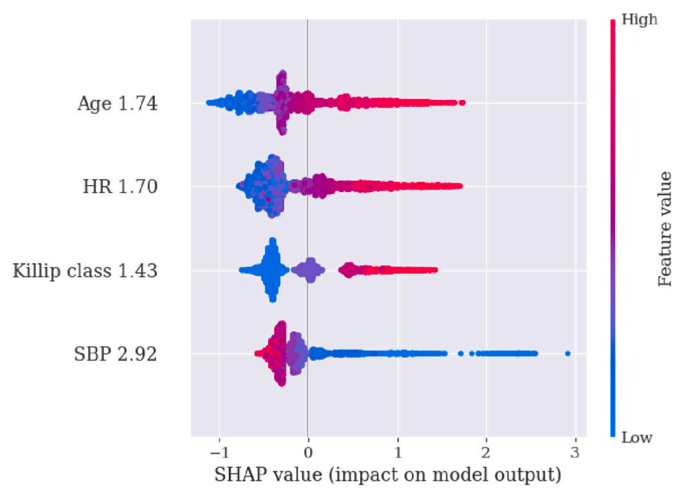

Fig. 3. Predictors impact estimation on IHM in the first scenario forecast.

only predictor in continuous form (Table 3). The second scenario base model used 5 predictors in a continuous form (age, HR, SBP, AHF T. Killip class and Cr), which also corresponded to the set of GRACE scale factors (Table 3). The predictive accuracy of these models was higher compared to the models from the first scenario, which could be explained by the Cr effect on the end point which was confirmed by the high (3.05 c.u.) SHAP value (Fig. 4). The SBP influence on the end point decreased (SHAP-value - 1.8 c.u.), but heart rate, age and AHF according to T. Killip influence did not change significantly (SHAP-value - 1.69 c.u., SHAP-value - 1.67 c.u., SHAP-value - 1.32 c.u., respectively).

The step-by-step inclusion of new factors identified at the first stage of the study into the basic model made it possible to increase forecast accuracy, which amounted to AUC - 0.9 during cross-validation and final testing in the best MLR model (Table 3). The newly identified predictors of this model were LV EF, NEUT, EOS, PCT, and Glu.

Estimation of the extended model predictors impact by the SHAP method showed that the greatest influence on IHM is exerted by the LV

Table 3
Accuracy estimation of predictive IHM models according to the second scenario with predictors in continuous form.

Metrics	Cross-validation			Final testing		
	MLR	SGB	RF	MLR	SGB	RF
	Univariate model based on the GRACE scores sum					
AUC	0.835 [0.821; 0.848]	0.825 [0.809; 0.842]	0.833 [0.819; 0.847]	0.834 [0.781; 0.888]	0.826 [0.767; 0.884]	0.833 [0.779; 0.888]
Sen	0.773 [0.75; 0.796]	0.751 [0.718; 0.785]	0.756 [0.726; 0.787]	0.773 [0.673; 0.873]	0.753 [0.641; 0.866]	0.755 [0.648; 0.862]
Spec	0.759 [0.748; 0.77]	0.772 [0.749; 0.796]	0.771 [0.749; 0.794]	0.759 [0.733; 0.785]	0.773 [0.739; 0.808]	0.772 [0.739; 0.806]
PPV	0.165 [0.155; 0.174]	0.169 [0.155; 0.183]	0.169 [0.155; 0.183]	0.164 [0.143; 0.185]	0.169 [0.143; 0.195]	0.169 [0.144; 0.194]
NPV	0.982 [0.98; 0.984]	0.981 [0.978; 0.983]	0.981 [0.979; 0.983]	0.982 [0.974; 0.99]	0.981 [0.973; 0.989]	0.981 [0.973; 0.989]
F-score	0.283 [0.268; 0.296]	0.288 [0.268; 0.309]	0.288 [0.268; 0.309]	0.281 [0.249; 0.312]	0.288 [0.249; 0.326]	0.288 [0.251; 0.324]
Predictors of the base model: Age, HR, SBP, AHF T. Killip class, Cr						
AUC	0.863 [0.851; 0.876]	0.855 [0,84; 0.869]	0.845 [0.831; 0.858]	0.865 [0.816; 0.914]	0.857 [0.805; 0.91]	0.847 [0.791; 0.903]
Sen	0.78 [0.76; 0.805]	0.773 [0.745; 0.802]	0.776 [0.74; 0.812]	0.783 [0.678; 0.887]	0.779 [0.666; 0.892]	0.781 [0.664; 0.898]
Spec	0.79 [0.777; 0.802]	0.798 [0.815]	0.753 [0.727; 0.779]	0.787 [0.759; 0.815]	0.796 [0.763; 0.829]	0.75 [0.708; 0.792]
PPV	0.186 [0.173; 0.198]	0.191 [0.177; 205]	0.162 [0.15; 0.175]	0.184 [0.159; 0.208]	0.19 [0.163; 0.216]	0.161 [0.131; 0.19]
NPV	0.983 [0.981; 0.985]	0.983 [0.981; 0.985]	0.982 [0.98; 0.985]	0.983 [0.976; 0.991]	0.983 [0.975; 0.992]	0.983 [0.973; 0.992]
F-score	0.313 [0.294; 0.33]	0.32 [0.3; 0.339]	0.278 [0.26; 0.297]	0.31 [0.273; 0.344]	0.318 [0.279; 0.355]	0.277 [0.231; 0.319]
Predictors: Age, HR, SBP, AHF T. Killip class, Cr, LV EF, NEUT, EOS, PCT, Glu						
AUC	0.9 [0.885; 0.916]	0.891 [0.871; 0.911]	0.885 [0.87; 0.9]	0.9 [0.841; 0.959]	0.892 [0.834; 0.951]	0.884 [0.824; 0.944]
Sen	0.843 [0.81; 0.877]	0.825 [0.779; 0.872]	0.796 [0.749; 0.843]	0.843 [0.715; 0.972]	0.824 [0.692; 0.957]	0.798 [0.672; 0.925]
Spec	0.836 [0.824; 0.849]	0.816 [0.797; 0.835]	0.806 [0.784; 0.828]	0.838 [0.807; 0.868]	0.819 [0.783; 0.855]	0.806 [0.766; 0.846]
PPV	0.165 [0.152; 0.178]	0.1146 [0.131; 0.161]	0.136 [0.124; 0.148]	0.168 [0.141; 0.194]	0.15 [0.125; 0.175]	0.138 [0.101; 0.167]
NPV	0.993 [0.992; 0.995]	0.992 [0.99; 0.994]	0.991 [0.989; 0.993]	0.993 [0.987; 0.999]	0.992 [0.986; 0.998]	0.991 [0.985; 0.996]
F-score	0.283 [0.264; 0.302]	0.255 [0.231; 0.277]	0.239 [0.22; 0.258]	0.287 [0.247; 0.325]	0.261 [0.222; 0.298]	0.242 [0.183; 0.286]

EF and Cr (SHAP-value 2.1 c.u. and SHAP-value1.94 c.u. accordingly) (Fig. 5).

To assess the impact of predictor influence on the end point, we considered the maximum SHAP values, since this value is associated with the risk of IHM. By interpreting SHAP value this way, we note a shift in the maximum importance of the SBP factor in the first forecast scenario to an increasing role of LV EF and Cr in the second one. It should also be noted a less noticeable, but equivalent effect on IHM of such indicators as NEUT, HR and blood Glu levels (SHAP-value - 1.35 c. u., SHAP-value - 1.35 c.u., SHAP-value - 1.31 c.u., respectively). In the extended second scenario model, SBP effect on the end point decreased

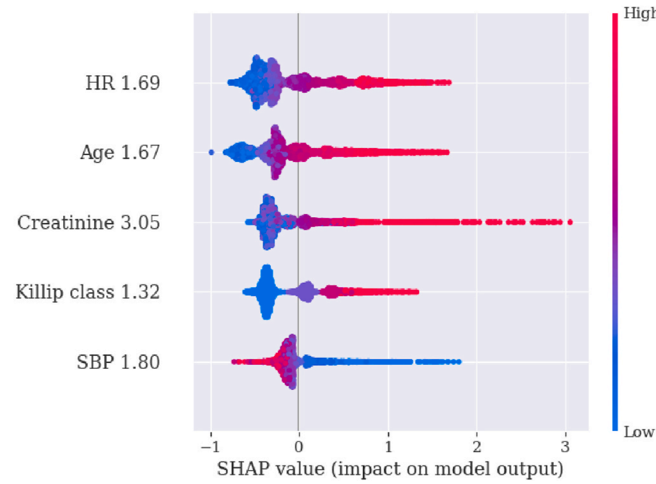

Fig. 4. Predictors impact estimation in base model of the second scenario.

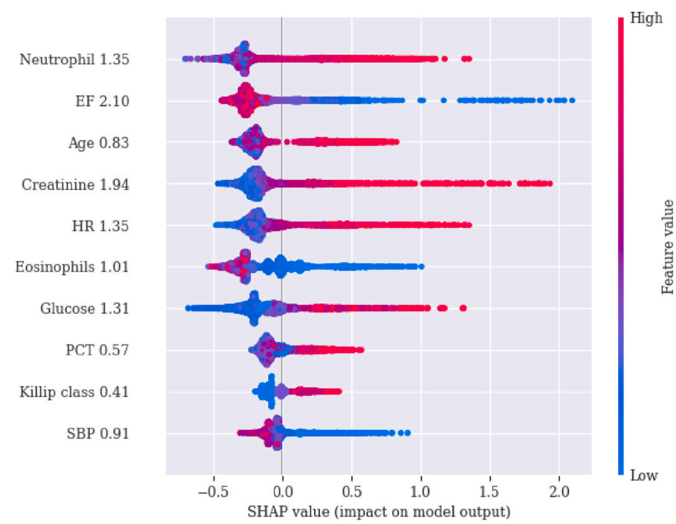

Fig. 5. Predictors impact estimation for expanded model of second scenario.

significantly (SHAP-value - 0.91 c.u.). A similar level of IHM impact was exerted by EOS (SHAP-value - 1.01 c.u.) and age (SHAP-value - 0.83 c. u.). The least influence was demonstrated by PCT (SHAP-value - 0.57 c. u.) and AHF T. Killip class (SHAP-value - 0.41 c.u.).

3.3. Risk factor

Identification of IHM risk factors was performed by searching for the optimal solution on the grid using several objective functions: maximizing OR and AUC values, minimizing the p-value, determined using the $\chi 2$ test, SHapley Additive exPlanations method and calculating the centroids [19]. For such indicators as: age, AHF T. Killip class, HR, Cr, SBP, LV EF, PCT, relative values of NEUT and EOS in the blood threshold values with the highest predictive potential were determined, which were attributed to risk factors (Table 4). The risk factor takes the value "1″ if the predictor value exceeds the threshold with the postfix "+", or does not reach it - with the postfix "-".

Since SHAP methods use multifactorial models, the threshold values of the same indicator estimation may differ depending on the forecast scenario. Thus, in the 1st scenario SBP indicator threshold value was 110 mmHg, but in the second 95 mmHg. The Max (OR) method determines indicators threshold values extremes, which most often associated with the lowest AUC value. All identified with proposed methods threshold values were statistically significant according to the $\chi 2$

Table 4
Determination of IHM risk factors using various methods.

Predictor	Method	Risk Factor	p-value	OR CI (95 %)	AUC
Age	Max (OR)	46+	<0.00001	10.7 [3.4; 33.6]	0.542
	Min (p-value)	70+	<0.00001	4.11 [3.26; 5.19]	0.655
	Max (AUC)	65+	<0.00001	3.73 [2.92; 4.77]	0.658
	Centroid	66.5+	<0.00001	3.65 [2.87; 4.63]	0.650
	SHAP	71+	<0.00001	4.11 [3.26; 5.19]	0.655
SBP	Max (OR)	60-	<0.00001	31.5 [10.9; 91.4]	0.523
	Min (p-value)	92-	<0.00001	11 [8.3; 14.7]	0.613
	Max (AUC)	112-	<0.00001	5.3 [4.2; 6.8]	0.685
	Centroid	120-	<0.00001	4.8 [3.8; 6]	0.676
	SHAP (1st scenario)	110-	<0.00001	7.9 [6.2; 10.1]	0.671
	SHAP (expanded model)	95-	<0.00001	11.6 [8.7; 15.4]	0.638
HR	Max (OR)	150+	<0.00001	41.61 [4.32; 401.16]	0.506
	Min (p-value)	95+	<0.00001	6.17 [4.82; 7.89]	0.65
	Max (AUC)	82+	<0.00001	4.54 [3.6; 5.73]	0.669
	Centroid	79+	<0.00001	3.96 [3.13; 5.02]	0.665
	SHAP	89+	<0.00001	5.5 [4.3; 7]	0.663
AHF T. Killip class	Max (OR)	4	<0.00001	9.6 [7.41; 12.4]	0.662
	Min (p-value)	4	<0.00001	9.6 [7.41; 12.4]	0.662
	Max (AUC)	3+	<0.00001	7.08 [5.59; 8.99]	0.711
	Centroid	3+	<0.00001	7.08 [5.59; 8.99]	0.711
	SHAP	3+	<0.00001	7.08 [5.59; 8.99]	0.711
Cr, mmol/l	Max (OR)	427.0+	<0.00001	30.5 [10.2; 91.7]	0.518
	Min (p-value)	188.6+	<0.00001	13.1 [9.4; 18]	0.625
	Max (AUC)	122.9+	<0.00001	5.8 [4.5; 7.4]	0.701
	Centroid	113.3+	<0.00001	4.6 [3.6; 6]	0.693
	SHAP	166+	<0.00001	10 [7.5; 13.4]	0.645
NEUT, %	Max (OR)	94.2+	<0.00001	23 [4.6; 114.8]	0.513
	Min (p-value)	78.8+	<0.00001	9.1 [6.6; 12.5]	0.730
	Max (AUC)	75+	<0.00001	11.3 [7.8; 16.2]	0.774
	Centroid	74.0+	<0.00001	9.9 [6.9; 14.4]	0.751
	SHAP	77+	<0.00001	11.3 [7.8; 16.2]	0.774
EOS, %	Max (OR)	1.3-	<0.00001	9.8 [5.3; 18.1]	0.666
	Min (p-value)	0.3-	<0.00001	7.9 [5.6; 11]	0.741
	Max (AUC)	0.3-	<0.00001	7.9 [5.6; 11]	0.741
	Centroid	0.5-	<0.00001	7.6 [5.2; 11.1]	0.722
	SHAP	0.2-	<0.00001	6.9 [5; 9.4]	0.712
LV EF, %	Max (OR)	31.0-	<0.00001	19.7 [12.1; 32.1]	0.594
	Min (p-value)	31.0-	<0.00001	19.7 [12.1; 32.1]	0.594
	Max (AUC)	51-	<0.00001	4.9 [3.5; 6.7]	0.69
	Centroid	51-	<0.00001	4.9 [3.5; 6.7]	0.69
	SHAP	45-	<0.00001	5.9 [4.3; 8.1]	0.648
PCT, %	Max (OR)	0.36+	<0.00001	4.5 [2.7; 7.4]	0.524
	Min (p-value)	0.36+	<0.00001	4.5 [2.7; 7.4]	0.524
	Max (AUC)	0.22+	0.00002	1.8 [1.4; 2.4]	0.598
	Centroid	0.21+	0.0009	1.6 [1.2; 2.1]	0.576
	SHAP	0.32+	0.00002	1.8 [1.4; 2.4]	0.598
Glu, mmol/l	Max (OR)	31+	0.002	44.8 [4; 496.5]	0.505
	Min (p-value)	8.4+	<0.00001	5.2 [3.9; 7.1]	0.62
	Max (AUC)	6.5+	<0.00001	4.9 [3.6; 6.8]	0.689
	Centroid	6.9+	<0.00001	5 [3.6; 6.7]	0.688
	SHAP	8.9+	<0.00001	4.7 [3.5; 6.4]	0.631

criterion.

A detailed analysis of the extended model predictors by SHAP made it possible to identify their threshold values, which were determined when the SHAP value was steadily exceeded by 0.1 units: age >70 years, HR > 89 bpm, SBP <95 mmHg, AHF T. Killip class >2, Cr > 166 mmol/l, LV EF < 45 %, NEUT >75 %, EOS <0.22 %, PCT >0.32 % and Glu >8.9 mmol/l (Fig. 6). The predictors threshold values obtained for the first scenario were the same for all indicators except the SBP. In this case, its risk factor value shifted to the level <110 mmHg. (Fig. 7).

Unlike previously existing methods for risk factors determination, the SHAP method takes into account the multifactorial impact on the end point, which corresponds to real clinical practice and allows you to explain the modeling results. That is why the risk factors threshold values identified by the SHAP method in most cases differ from the results obtained by other methods.

During the 4th stage of the study IHM models were developed, based on the MLR, the predictors of which were risk factors obtained using 5 methods (Table 5).

The best forecast accuracy based on the AUC, Sen, Spec and F-score metrics was demonstrated by the model obtained using the SHAP method for both forecast scenarios.

4. Discussion

In this work, data mining was carried out on a dataset of STEMI patients after PCI in order to search and validate new IHM predictors, verify fatal outcome risk factors among them, develop prognostic models exceeding accuracy of the well-known GRACE model and represent explanation elements. The allotted potential predictors pool includes 4 GRACE scale factors in the first prognosis scenario and 5 in the second, as well as a number of new indicators, some of which were not previously used for IHM after PCI prognostic models. MLR, SGB and ML, cross-validation and testing procedures confirmed predictive value of those indicators in IHM relation. The models developed on the basis of these methods demonstrated higher quality metrics compared to GRACE and second scenario base model. Wherein, the best prediction accuracy was achieved in the MLR model. High influence degree on IHM was demonstrated by applying SHapley Additive exPlanations for newly isolated predictors: LV EF, NEUT, EOS, PCT and Glu levels (Fig. 5). Utilizing this method, the predictors threshold values were identified,

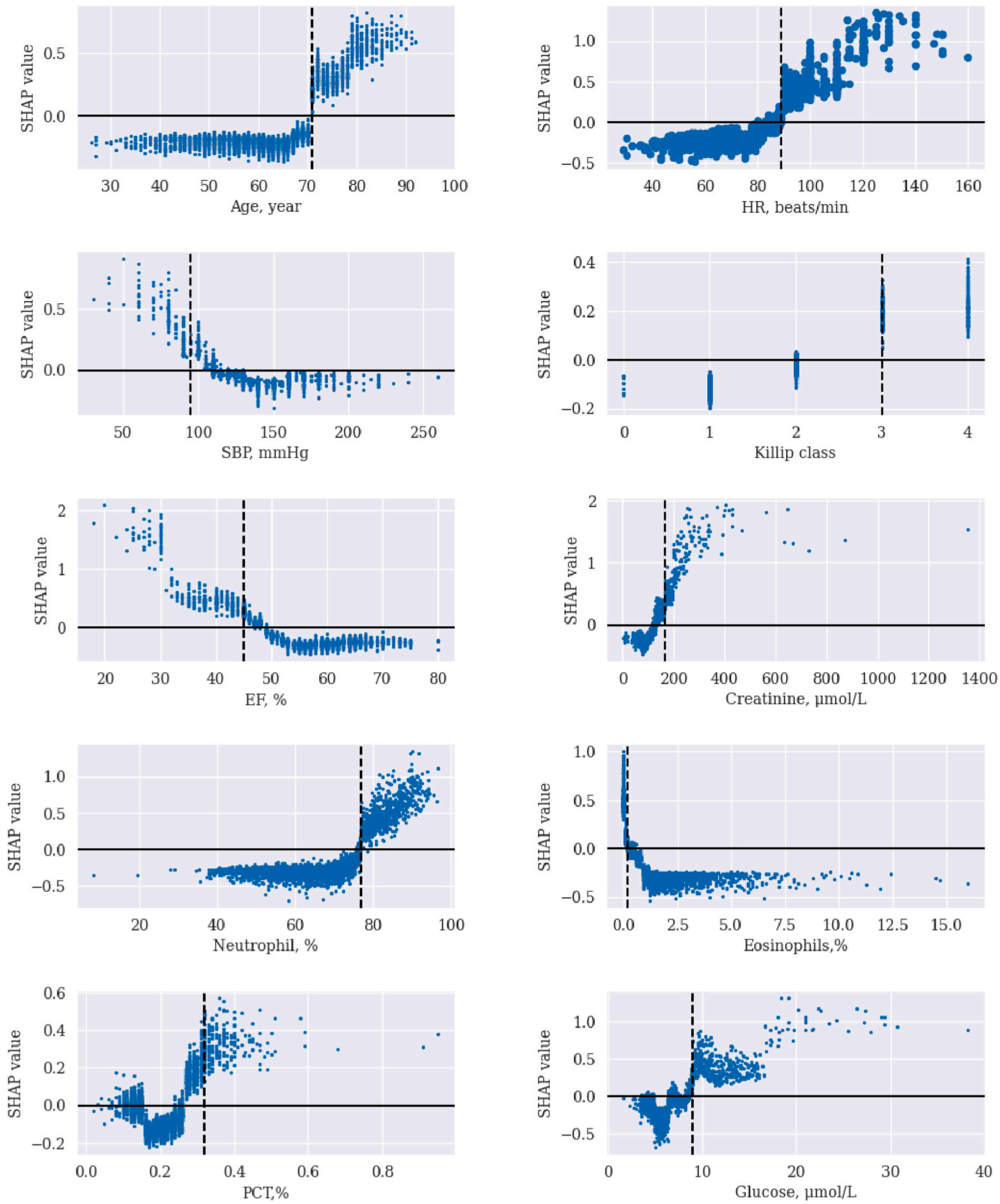

Fig. 6. Evaluation of the second scenario extended model indicators impact based on SHAP.

the deviation from which was attributed to IHM risk factors. This approach allows physicians to assess the likelihood of IHM in real time and carry out personalized correction of physiological functions.

For risk factors determination, we used methods of variables categorization (dichotomization). The data categorization problem is a subject for discussion, and according to a number of authors, predictors usage in a continuous form is more preferable from the position of information saving and taking it into account for the forecasting process [20,21]. It is also noted that dichotomization, in addition to a significant information loss that could be useful for assessing variable impact on the end point [22], can lead to distortion of modeling results [23]. At the same time, the data categorization and, as a special case, its dichotomization, has a number of advantages. First of all, categorized variables allow you to identify risk factors and explain the predicted probability of event occurrence [24]. In addition, based on a risk factors combination, the possibility of phenotype formation is realized, characterizing the complex effect of various traits on the resulting variable, which increases the modeling results explainability [25].

According to the relevant literature sources, researchers use several approaches for risk factors searching. The simplest and most frequently used methods among those were threshold value determination obtained using descriptive statistics: medians or quartiles [22,26]. Most of the categorization criticisms are applied to this approach, which is primarily due to the dependence of such threshold values on a specific

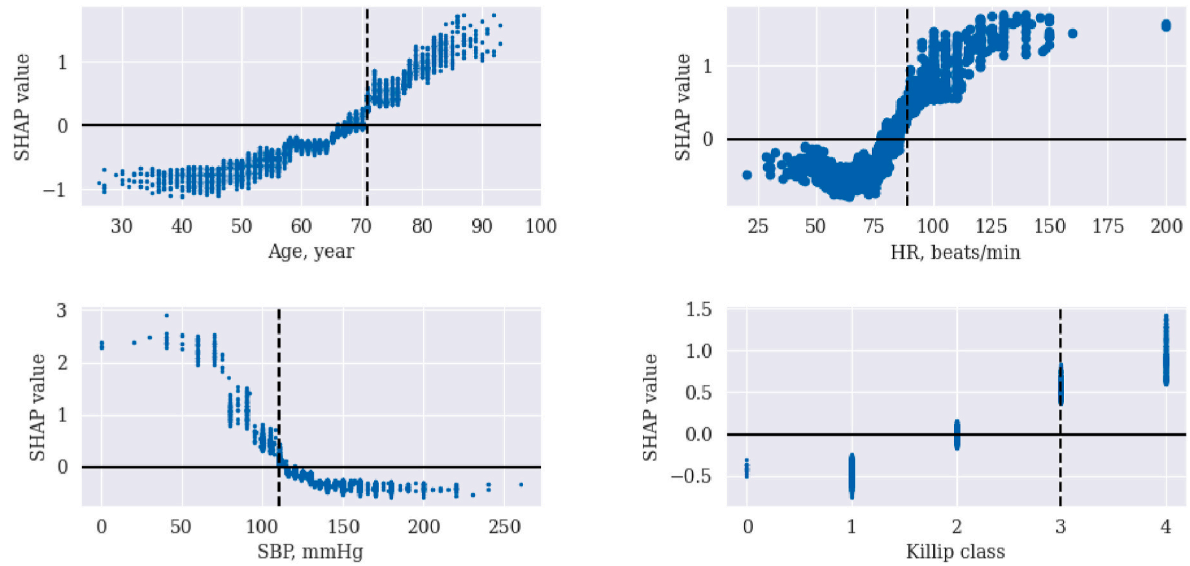

Fig. 7. Evaluation of the first scenario model indicators impact based on SHAP.

Table 5
Final testing assessment of IHM predictive models accuracy based on the risk factors.

Метрики	SHAP	Min (p-value)	Max (AUC)	Max (OR)	Centroid
	Age, SBP, HR, AHF T.Killip class				
AUC	0.823 [0.809; 0.838]	0.796 [0.744; 0.847]	0.819 [0.768; 0.87]	0.693 [0.644; 0.743]	0.822 [0.772; 0.872]
Sen	0.755 [0.619; 0.891]	0.72 [0.5; 0.951]	0.774 [0.685; 0.863]	0.556 [0; 1]	0.748 [0.567; 0.929]
Spec	0.725 [0.623; 0.827]	0.723 [0.545; 0.902]	0.769 [0.744; 0.795]	0.683 [0; 1]	0.764 [0.686; 0.842]
PPV	0.172 [0.126; 0.217]	0.171 [0.1; 0.241]	0.198 [0.171; 0.225]	0.238 [0; 0.456]	0.191 [0.161; 0.22]
NPV	0.976 [0.966; 0.987]	0.973 [0.958; 0.989]	0.979 [0.971; 0.987]	0.966 [0.926; 1]	0.977 [0.963; 0.991]
F-score	0.292 [0.223; 0.356]	0.291 [0.181; 0.388]	0.329 [0.291; 0.366]	0.382 [0; 0.626]	0.32 [0.276; 0.36]
Age, SBP, HR, AHF T.Killip class, Cr, NEUT, EOS, PCT, Glu					
AUC	0.896 [0.837; 0.955]	0.893 [0.834; 0.951]	0.894 [0.841; 0.984]	0.802 [0.728; 0.876]	0.882 [0.824; 0.94]
Sen	0.824 [0.687; 0.96]	0.812 [0.677; 0.946]	0.836 [0.703; 0.968]	0.647 [0.135; 1]	0.81 [0.682; 0.939]
Spec	0.827 [0.794; 0.859]	0.817 [0.775; 0.859]	0.813 [0.778; 0.848]	0.721 [0.246; 1]	0.822 [0.812; 0.833]
PPV	0.155 [0.13; 0.181]	0.147 [0.116; 0.178]	0.147 [0.122; 0.172]	0.134 [0; 0.265]	0.15 [0.123; 0.177]
NPV	0.992 [0.986; 0.998]	0.991 [0.985; 0.997]	0.992 [0.986; 0.998]	0.985 [0.966; 1]	0.991 [0.985; 0.997]
F-score	0.268 [0.23; 0.306]	0.249 [0.202; 0.295]	0.256 [0.217; 0.293]	0.236 [0; 0.419]	0.261 [0.219; 0.301]

sample, the lack of relationship with the clinical context, ignoring possible non-linear relationships, etc. The second approach is based on the allocation of a threshold value, known from practice as going beyond the norm [20]. Body mass index >25 kg/m2, for example, is a sign of being overweight and can be considered as a potential risk factor for a number of diseases and their complication development. An alternative approach is to find optimal threshold values based on minimizing or maximizing objective functions. One of the first to be criticized was the method of threshold values searching based on p-value minimization, which shows cases of false positive results [26]. Finding thresholds by odds ratio maximization can lead to Type II predictive errors [27]. In most studies, the threshold values search is carried out based on univariate analysis methods: p-value minimization, odds ratio maximization, median and quartile values and univariate logistic regression weights maximization [28]. Previously, it was proposed to apply validation for the optimal cut-off threshold searching based on p-value minimization and relative risk maximization by randomly dividing the sample into training and validating ones, which allowed the authors to identify 3 threshold values of indicators that were used for further analysis [29]. At the same time, the cut-off threshold search can be considered taking into account the multifactorial influence on the end point [30]. A number of papers indicate that the choice of cut-off values should be related to the clinical context [31]. Therefore, all threshold values of indicators that have normative ranges must be outside their boundaries.

In our opinion, the variables categorization should be considered only within the framework of solving a specific clinical problem, although the identified risk factors may be field of interest for the implementation in other prognostic studies. Whatever method is used for threshold value determination for factor verification, this can lead, on the one hand, to information loss, and, on the other hand, to the introduction of new knowledge. The loss of information may have an insignificant impact on the predictive model quality (in our case, AUC decreased from 0.9 to 0.896), and new knowledge is associated with both a decrease in entropy and a possible prediction results explanation. We have considered a number of methods for risk factor thresholds search, optimizing several criteria, as well as using SHapley Additive exPlanations method. The latter made it possible to identify the most accurate risk factors threshold values, which was confirmed by a multivariate prognostic model. It should also be taken into account that some predictors have multiple thresholds, deviations from which transform them into risk factors associated with various causes of unfavorable outcome. In these cases, threshold values of the same indicators can be associated both with exceeding the upper limit of the reference range, and with going beyond its lower limit. For example, risk factors for IHM in patients with CHD after surgical myocardial revascularization include a thrombin time less than 13 s, which characterizes excessive blood clotting and an increased likelihood of thrombosis. At the same time, its increase more than 27 s indicates a high risk of bleeding in the postoperative period [25]. That is why it is advisable to search for threshold values using the method of step-by-step indicators cutting from the minimum to the maximum value and vice versa, including when the SHAP method is applied.

The first most important predictor of the extended IHM predictive model (in terms of the maximum level of SHAP value) is LV EF. This indicator was also analyzed by other authors [32–36]. Most studies consider a threshold value for EF < 40 %, while in our study it was shown that the risk of IHM already increases at EF < 45 %. The neutrophils blood level more than 75 % has not been identified by other researchers, but in a number of studies the neutrophils to lymphocyte ratio is considered as a IHM predictor [32,37,38]. We also were first to determine the threshold values EOS, PCT, and Glu predictors: EOS <0.22 %, PCT >0.32 %, and Glu >8.9 mmol/L.

5. Limitations

The limitation of this study is due the fact that in our work we analyzed data from only one medical center, which does not guarantee the reproducibility of prediction results on other data sets. Another limitation is the forced exclusion of patients who died during first hours of hospital stay from second scenario models training process, due to the lack of laboratory test results, which makes it impossible to assess IHM prognosis quality for such patients under the second scenario.

6. Conclusion

In this study, IHM predictors were identified and validated in STEMI patients after myocardial revascularization by PCI. Several prognostic models of IHM, associated with two stages of care were developed utilizing MLR, RF and SGB methods. 1st scenario IHM models determined the unfavorable outcome probability immediately after patients were admitted to the hospital according to data obtained during the initial doctor examination. 2nd scenario models took into account the results of laboratory tests and were developed based on continuous, identified by the authors IHM predictors (age, SBP, HR, AHF T. Killip class, Cr, LVEF, NEUT, EOS, GLU, PCT). All developed models showed higher predictive properties compared to the classical GRACE scale (AUC - 0.853 and 0.9 versus 0.834 for GRACE). The best models for both scenarios were obtained by MLR method.

IHM risk factors threshold values determination was carried out based on several methods, the most effective of which was SHAP. This was confirmed by the higher quality predictive models utilizing dichotomized risk factors derived from the SHAP method. An additional advantage of this method is its ability to adapt thresholds to different forecast scenarios by the multivariate models. Risk factors threshold values increase the explainability of prognostic models without reducing their quality and allows practitioners physicians to timely implement measures for fatal outcomes risk reduction.

Ethics approval and consent to participate

The study was approved by the Ethics Committee of School of Medicine, Far Eastern Federal University. The data was collected by our author and physician involved in the operations (Domzhalov I.G.), checked for lack of personalization and registered as a public dataset: certificate number: RU 2023622740, application number: 2023622516, registration date 28/07/2023. All procedures performed in the present study were in accordance with the Declaration of Helsinki. The datasets of all patients in the database were anonymized and publicly available.

Availability of data and materials

The analyzed data and Python code are available by link (https://github.com/NikitaKuksin/PublicCodAticle).

Funding

This study was financially supported by the Ministry of Science and Higher Education of the Russian Federation (the project FZNS-2023-0010 of the State Assignment of the Far Eastern Federal University (FEFU))

CRediT authorship contribution statement

Karina Iosephovna Shakhgeldyan: Data curation, Formal analysis, Project administration, Software, Writing – original draft, Writing – review & editing, Resources, Supervision, Validation. **Nikita Sergeevich Kuksin:** Data curation, Formal analysis, Software, Writing – original draft, Writing – review & editing. **Igor Gennadievich Domzhalov:** Conceptualization, Data curation, Resources, Writing – original draft, Writing – review & editing. **Vladislav Yurievich Rublev:** Conceptualization, Data curation, Investigation, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing. **Boris Izrajlevich Geltser:** Conceptualization, Data curation, Project administration, Supervision, Writing – original draft, Writing – review & editing, Investigation, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Not applicable.

Appendix A. Detailed comparison of surviving and deceased patients groups

Table A.1
Detailed comparison of surviving and deceased patients groups

Predictor		Group 1 (n = 318)	Group 2 (n = 4357)	OR (95 %) CI	p-value
Gender: F, abs. (%)	318	142 (44.65)	1332 (30.5)	1.8 [1.5; 2.3]	<0.000001
Age, years	318	71 (63; 78)	62 (55; 69)	–	<0.000001
Height, cm	287	168 (164; 174)	170 (165; 176)	–	0.000001
Weight, kg	286	78 (70; 85)	80 (71; 90)	–	0.000012
BMI, kg/m2	285	27.04 (26.1; 27.68)	27.68 (27.46; 27.7)	–	0.082
SBP, mmHg	317	110 (90; 130)	130 (120; 150)	–	<0.000001
DBP, mmHg	317	70 (60; 80)	80 (75; 90)	–	<0.000001

(*continued on next page*)

Table A.1 (*continued*)

Predictor		Group 1 (n = 318)	Group 2 (n = 4357)	OR (95 %) CI	p-value
PBP, mmHg	317	40 (40; 42)	50 (50; 50)	–	<0.000001
HR, bpm	316	86 (72; 100)	72 (65; 80)	–	<0.000001
Cr, µmol/l	265	130 (96; 193.3)	97 (81; 114.8)	–	<0.000001
AHF T.Killip class					
I	318	71 (22.33 %)	2726 (62.6 %)	0.17 [0.13; 0.23]	<0.000001
II	318	58 (18.2 %)	867 (19.9 %)	0.9 [0.67; 1.20]	0.508052
III	318	66 (20.75)	479 (11)	2.1 [1.6; 2.8]	<0.000001
IV	318	123 (38.7)	269 (6.18)	9.6 [7.4; 12.4]	<0.000001
III-IV	318	189 (59.4)	748 (17.2)	7.1 [5.6; 9]	<0.000001
LV EF, %	174	46.5 (38; 54.8)	56 (50; 61)	–	<0.000001
LV EDD, cm	172	5 (4.6; 5.5)	5 (4.7; 5.3)	–	0.356921
LV ESD, cm	166	3.7 (3.2; 4)	3.4 (3.1; 3.8)	–	<0.000001
RTI	171	0.408 (0.34; 0.47)	0.417 (0.38; 0.47)	–	0.554
RTI LVMI	153	1.06 (0.84; 1.28)	0.96 (0.8; 1.14)	–	0.0003
MPAP, mm Hg	170	35 (28.25; 46)	28 (25; 30)	–	<0.000001
LA1, cm	169	4.1 (3.8; 4.50)	3.9 (3.6; 4.2)	–	<0.000001
LA2, cm	169	5.2 (4.8; 5.7)	4.9 (4.6; 5.2)	–	<0.000001
RA1, cm	168	3.8 (3.5; 4.2)	3.6 (3.4; 3.9)	–	<0.000001
RA2, cm	168	4.8 (4.5; 5.3)	4.7 (4.4; 5)	–	0.00004
WBC, 10^9/l	201	14.5 (10.9; 19.2)	9.8 (7.9; 12.3)	–	<0.000001
NEUT, %	175	81.3 (75.75; 86.5)	66.7 (59.1; 74.9)	–	<0.0001
LYM, %					
EOS, %	176	0.1 (0.00; 0.3)	0.9 (0.3; 1.9)	–	<0.000001
RBC, 10^{12}/l	244	4.25 (3.8; 4.65)	4.48 (4.1; 4.83)	–	<0.000001
HB, g/l	201	130 (114; 142)	141 (128; 152)	–	<0.000001
PLT, 10^9/l	245	228 (187; 288)	221 (185; 266)	–	0.020702
Glu, mmol/l	189	7.9 (6.3; 10.31)	5.8 (5.1; 7)	–	<0.000001
Urea, µmol/l	212	12.12 (8.7; 17.3)	6.7 (5.24; 8.84)	–	<0.000001
PCT, %	215	0.22 (0.17; 0.28)	0.2 (0.16; 0.24)	–	0.0012
PTI, %	214	75.5 (57.6; 87)	89.3 (79.7; 97)	–	<0.000001
INR, units.	211	1.26 (1.1; 1.65)	1.06 (1; 1.16)	–	<0.000001
TT, s	152	21.9 (19.9; 30.4)	21.4 (19.5; 25.7)	–	0.012
APTT, s	208	39.7 (32.7; 58.2)	36.5 (32.2; 42.7)	–	0.000026
Anterior MI, abs. (%)	318	177 (55.66 %)	2017 (46.27 %)	1.5 [1.16; 1.83]	0.00147
AF, abs. (%)	318	129 (40.57 %)	772 (17.69 %)	3.2 [2.51; 4.02]	<0.000001
T2DM, abs. (%)	318	99 (31.13 %)	831 (19.05 %)	1.9 [1.50; 2.46]	<0.000001
CKD, abs. (%)	316	83 (26.1 %)	677 (15.5 %)	1.97 [1.5; 2.6]	<0.000001
MI with stenosis history	318	16 (5 %)	165 (3.8 %)	1.3 [0.8; 2.28]	0.34
COPD	318	25 (7.9 %)	354 (8.1 %)	0.96 [0.63; 1.47]	0.95

Abbreviations: LV EF - left ventricular ejection fraction, LV EDD - LV end diastolic dimension, LV ESD - LV end systolic dimension, MPAP- mean pulmonary artery pressure, LA2 - longitudinal size of the left atrium, RA1 - transverse size of the right atrium, RA2 - longitudinal right atrial size, RBC - erythrocytes, Hb - hemoglobin, PLT - platelets, Glu - glucose, Urea - urea, PCT - thrombocrit; BMI - body mass index, SBP - systolic blood pressure, DBP - diastolic blood pressure, HR - heart rate, WBC - leukocytes, PTI - prothrombin index; INR - international normalized ratio; TT - thrombin time, APTT - activated partial thromboplastin time, Cr - creatinine, AF - atrial fibrillation, T2DM - type 2 diabetes mellitus, CKD - chronic kidney disease.

References

[1] GBD, Mortality and Causes of Death Collaborators. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015, 2016 Oct 8, Lancet 388 (10053) (2015) 1459–1544, https://doi.org/10.1016/S0140-6736(16)31012-1. Erratum in: Lancet. 2017 Jan 7;389(10064):e1. PMID: 27733281; PMCID: PMC5388903.

[2] B. Ibanez, S. James, S. Agewall, M.J. Antunes, C. Bucciarelli-Ducci, H. Bueno, A.L. P. Caforio, F. Crea, J.A. Goudevenos, S. Halvorsen, G. Hindricks, A. Kastrati, M. J. Lenzen, E. Prescott, M. Roffi, M. Valgimigli, C. Varenhorst, P. Vranckx, P. Widimský, ESC Scientific Document Group, 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. 39 (2) (2017) 119–177, https://doi.org/10.1093/eurheartj/ehx393. PMID:28886621, 2018 Jan 7.

[3] B. Ibanez, S. James, S. Agewall, M.J. Antunes, C. Bucciarelli-Ducci, H. Bueno, A.L. P. Caforio, F. Crea, J.A. Goudevenos, S. Halvorsen, G. Hindricks, A. Kastrati, M. J. Lenzen, E. Prescott, M. Roffi, M. Valgimigli, C. Varenhorst, P. Vranckx, P. Widimský, ESC Scientific Document Group, 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. 39 (2) (2017) 119–177, https://doi.org/10.1093/eurheartj/ehx393. PMID:28886621, 2018 Jan 7.

[4] E.W. Tang, C.K. Wong, P. Herbison, Global Registry of Acute Coronary Events (GRACE) hospital discharge risk score accurately predicts long-term mortality post acute coronary syndrome, Am. Heart J. 153 (1) (2007 Jan) 29–35, https://doi.org/10.1016/j.ahj.2006.10.004. Erratum in: Am Heart J. 2007 Nov;154(5):851. PMID: 17174633.

[5] C.B. Granger, R.J. Goldberg, O. Dabbous, K.S. Pieper, K.A. Eagle, C.P. Cannon, F. Van De Werf, A. Avezum, S.G. Goodman, M.D. Flather, K.A. Fox, Global Registry of Acute Coronary Events Investigators, Predictors of hospital mortality in the global registry of acute coronary events, Arch. Intern. Med. 163 (19) (2003 Oct 27) 2345–2353, https://doi.org/10.1001/archinte.163.19.2345.PMID:14581255.

[6] R.S. Wilson, P. Malamas, B. Dembo, S.K. Lall, N. Zaman, B.R. Peterson, The CADILLAC risk score accurately identifies patients at low risk for in-hospital mortality and adverse cardiovascular events following ST elevation myocardial infarction, BMC Cardiovasc. Disord. 21 (1) (2021 Nov 12) 533, https://doi.org/10.1186/s12872-021-02348-0. PMID: 34772341; PMCID: PMC8588705.

[7] Y.T. Kao, Y.C. Hsieh, C.Y. Hsu, C.Y. Huang, M.H. Hsieh, Y.K. Lin, J.S. Yeh, Comparison of the TIMI, GRACE, PAMI and CADILLAC risk scores for prediction of long-term cardiovascular outcomes in Taiwanese diabetic patients with ST-segment elevation myocardial infarction: from the registry of the Taiwan Society of Cardiology, PLoS One 15 (2) (2020 Feb 13) e0229186, https://doi.org/10.1371/journal.pone.0229186. PMID: 32053694; PMCID: PMC7018102.

[8] X.F. Bai, Y.P. Zhang, J. Zhou, Y. Wu, R.F. Li, L.Z. Sun, Q.Q. Ma, B.W. Lou, B. W. Zhai, M.P. Liu, L.L. Cheng, X.N. Tong, Z.Y. Yuan, Combination of the CYP2C19 metabolizer and the GRACE risk score better predicts the long-term major adverse cardiac events in acute coronary syndrome undergoing percutaneous coronary intervention, Thromb. Res. 170 (2018 Oct) 142–147, https://doi.org/10.1016/j.thromres.2018.08.016. Epub 2018 Aug 24. PMID: 30193195.

[9] L. Wang, Z. Zhang, T. Hu, Effectiveness of LODS, OASIS, and SAPS II to predict in-hospital mortality for intensive care patients with ST elevation myocardial infarction, Sci. Rep. 11 (1) (2021 Dec 13) 23887, https://doi.org/10.1038/s41598-021-03397-3. PMID: 34903820; PMCID: PMC8668882.

[10] J.J. Wang, Y. Fan, Y. Zhu, J.D. Zhang, S.M. Zhang, Z.F. Wan, H.L. Su, N. Jiang, Biomarkers enhance the long-term predictive ability of the KAMIR risk score in Chinese patients with ST-elevation myocardial infarction, Chin. Med. J. 132 (1)

(2019 Jan 5) 30–41, https://doi.org/10.1097/CM9.0000000000000015. PMID: 30628957; PMCID: PMC6629309.

[11] X.J. Liu, Z.F. Wan, N. Zhao, Y.P. Zhang, L. Mi, X.H. Wang, D. Zhou, Y. Wu, Z. Y. Yuan, Adjustment of the GRACE score by HemoglobinA1c enables a more accurate prediction of long-term major adverse cardiac events in acute coronary syndrome without diabetes undergoing percutaneous coronary intervention, Cardiovasc. Diabetol. 14 (2015 Aug 19) 110, https://doi.org/10.1186/s12933-015-0274-4. PMID: 26285575; PMCID: PMC4541750.

[12] X. Chen, M. Shao, T. Zhang, W. Zhang, Y. Meng, H. Zhang, H. Hai, G. Li, Prognostic value of the combination of GRACE risk score and mean platelet volume to lymphocyte count ratio in patients with ST-segment elevation myocardial infarction after percutaneous coronary intervention, Exp. Ther. Med. 19 (6) (2020 Jun) 3664–3674, https://doi.org/10.3892/etm.2020.8626. Epub 2020 Mar 26. PMID: 32346430; PMCID: PMC7185188.

[13] F.A. Wenzl, S. Kraler, G. Ambler, C. Weston, S.A. Herzog, L. Räber, et al., Sex-specific evaluation and redevelopment of the GRACE score in non-ST-segment elevation acute coronary syndromes in populations from the UK and Switzerland: a multinational analysis with external cohort validation, Lancet 400 (2022) 744–756.

[14] R. Li, L. Shen, W. Ma, B. Yan, W. Chen, J. Zhu, L. Li, J. Yuan, C. Pan, Use of machine learning models to predict in-hospital mortality in patients with acute coronary syndrome, Clin. Cardiol. 46 (2) (2023 Feb) 184–194, https://doi.org/10.1002/clc.23957. Epub 2022 Dec 7. PMID: 36479714; PMCID: PMC9933107.

[15] C.J. Zack, C. Senecal, Y. Kinar, Y. Metzger, Y. Bar-Sinai, R.J. Widmer, R. Lennon, M. Singh, M.R. Bell, A. Lerman, R. Gulati, Leveraging machine learning techniques to forecast patient prognosis after percutaneous coronary intervention, JACC Cardiovasc. Interv. 12 (14) (2019 Jul 22) 1304–1311, https://doi.org/10.1016/j.jcin.2019.02.035. Epub 2019 Jun 26. PMID: 31255564.

[16] Xinwei Du, Hao Wang, Shunmin Wang, He Yi, Jinghao Zheng, Haibo Zhang, Zedong Hao, Yiwei Chen, Zhiwei Xu, Zhaohui Lu, Machine learning model for predicting risk of in-hospital mortality after surgery in congenital heart disease patients, Rev. Cardiovasc. Med. 23 (11) (2022 Nov 3) 376, https://doi.org/10.31083/j.rcm2311376.

[17] P. Zhao, C. Liu, C. Zhang, Y. Hou, X. Zhang, J. Zhao, G. Sun, J. Zhou, Using machine learning to predict the in-hospital mortality in women with ST-segment elevation myocardial infarction, Rev. Cardiovasc. Med. 24 (5) (2023) 126, https://doi.org/10.31083/j.rcm2405126.

[18] S.M. Lundberg, S.I. Lee, A unified approach to interpreting model predictions, in: Advances in Neural Information Processing Systems. Proceedings of the 31st Annual Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, 2017, 2017 Dec 04-09.

[19] F. Valente, J. Henriques, S. Paredes, T. Rocha, P. de Carvalho, J. Morais, A new approach for interpretability and reliability in clinical risk prediction: acute coronary syndrome scenario, Artif. Intell. Med. 117 (2021 Jul) 102113, https://doi.org/10.1016/j.artmed.2021.102113. Epub 2021 May 13. PMID: 34127242.

[20] E.L. Turner, J.E. Dobson, S.J. Pocock, Categorisation of continuous risk factors in epidemiological publications: a survey of current practice, Epidemiol. Perspect. Innovat. 7 (2010 Oct 15) 9, https://doi.org/10.1186/1742-5573-7-9. PMID: 20950423; PMCID: PMC2972292.

[21] N.V. Dawson, R. Weiss, Dichotomizing continuous variables in statistical analysis: a practice to avoid, Med. Decis. Making 32 (2) (2012 Mar-Apr) 225–226, https://doi.org/10.1177/0272989X12437605.PMID:22457338.

[22] D.G. Altman, P. Royston, The cost of dichotomising continuous variables, BMJ 332 (7549) (2006 May 6) 1080, https://doi.org/10.1136/bmj.332.7549.1080. PMID: 16675816; PMCID: PMC1458573.

[23] Z. Salis, B. Gallego, A. Sainsbury, Researchers in rheumatology should avoid categorization of continuous predictor variables, BMC Med. Res. Methodol. 23 (1) (2023 Apr 26) 104, https://doi.org/10.1186/s12874-023-01926-4. PMID: 37101144; PMCID: PMC10134601.

[24] R. Gupta, C.N. Day, W.O. Tobin, C.S. Crowson, Understanding the effect of categorization of a continuous predictor with application to neuro-oncology, Neurooncol. Pract. 9 (2) (2021 Aug 4) 87–90, https://doi.org/10.1093/nop/npab049. PMID: 35371519; PMCID: PMC8965047.

[25] B.I. Geltser, K.I. Shakhgeldyan, VYu Rublev, I.G. Domzhalov, M.M. Tsivanyuk, O. I. Shekunova, Phenotyping of risk factors and prediction of inhospital mortality in patients with coronary artery disease after coronary artery bypass grafting based on explainable artificial intelligence methods, Russian J. Cardiol. 28 (4) (2023 Apr 28) 5302, https://doi.org/10.15829/1560-4071-2023-5302 (In Russ.).

[26] O.V. Mabikwa, D.C. Greenwood, P.D. Baxter, S.J. Fleming, Assessing the reporting of categorised quantitative variables in observational epidemiological studies, BMC Health Serv. Res. 17 (1) (2017 Mar 14) 201, https://doi.org/10.1186/s12913-017-2137-z. Erratum in: BMC Health Serv Res. 2017 Apr 26;17 (1):307. PMID: 28288628; PMCID: PMC5348776.

[27] D.G. Altman, B. Lausen, W. Sauerbrei, M. Schumacher, Dangers of using "optimal" cutpoints in the evaluation of prognostic factors, J. Natl. Cancer Inst. 86 (11) (1994 Jun 1) 829–835, https://doi.org/10.1093/jnci/86.11.829.PMID:8182763.

[28] P. Royston, D.G. Altman, W. Sauerbrei, Dichotomizing continuous predictors in multiple regression: a bad idea, Stat. Med. 25 (1) (2006 Jan 15) 127–141, https://doi.org/10.1002/sim.2331.PMID:16217841.

[29] D. Faraggi, R. Simon, A simulation study of cross-validation for selecting an optimal cutpoint in univariate survival analysis, Stat. Med. 15 (20) (1996 Oct 30) 2203–2213, https://doi.org/10.1002/(SICI)1097-0258(19961030)15:20<2203::AID-SIM357>3.0.CO;2-G. PMID: 8910964.

[30] M. Mazumdar, A. Smith, J. Bacik, Methods for categorizing a prognostic variable in a multivariable setting, Stat. Med. 22 (4) (2003 Feb 28) 559–571, https://doi.org/10.1002/sim.1333.PMID:12590414.

[31] T.T. Timbrook, C.D. Garner, K.D. Hueth, G.A. Capraro, L. Zimmer, H.P. Dwivedi, Procalcitonin and risk prediction for diagnosing bacteremia in hospitalized patients: a retrospective, national observational study, Diagnostics 13 (20) (2023 Oct 11) 3174, https://doi.org/10.3390/diagnostics13203174. PMID: 37891995; PMCID: PMC10605738.

[32] Y. Wang, W. Wang, S. Jia, M. Gao, S. Zheng, J. Wang, Y. Dang, Y. Li, X. Qi, Development of a nomogram for the prediction of in-hospital mortality in patients with acute ST-elevation myocardial infarction after primary percutaneous coronary intervention: a multicentre, retrospective, observational study in Hebei province, China, BMJ Open 12 (2) (2022 Feb 2) e056101, https://doi.org/10.1136/bmjopen-2021-056101. PMID: 35110324; PMCID: PMC8811571.

[33] N. Gao, X. Qi, Y. Dang, Y. Li, G. Wang, X. Liu, N. Zhu, J. Fu, Association between total ischemic time and in-hospital mortality after emergency PCI in patients with acute ST-segment elevation myocardial infarction: a retrospective study, BMC Cardiovasc. Disord. 22 (1) (2022 Mar 4) 80, https://doi.org/10.1186/s12872-022-02526-8. PMID: 35246059; PMCID: PMC8896149.

[34] M.F. Yılmaz, A. Karagöz, G. Zeren, Avcı İi, B. Timur, M.A. Sungur, S. Kalkan, B. Şimşek, F. Can, E. Acar, İ.H. Tanboğa, C.Y. Karabay, Relationship between in-hospital mortality and creatinine/albumin in patients with ST-elevation myocardial infarction without standard modifiable risk factors, Biomarkers Med. 16 (14) (2022 Oct) 1043–1053, https://doi.org/10.2217/bmm-2022-0241. Epub 2022 Sep 5. PMID: 36062571.

[35] S.M. Ghorashi, M. Salarifar, H. Poorhosseini, S. Sadeghian, A. Jalali, H. Aghajani, A.M. Haji-Zeinali, N. Omidi, Predictors of in-hospital mortality in diabetic patients with non-ST-elevation myocardial infarction, Egypt Heart J. 74 (1) (2022 Mar 28) 20, https://doi.org/10.1186/s43044-022-00256-y. PMID: 35347479; PMCID: PMC8960521.

[36] T. Kırıs, E. Avcı, A. Çelik, Combined value of left ventricular ejection fraction and the Model for End-Stage Liver Disease (MELD) score for predicting mortality in patients with acute coronary syndrome who were undergoing percutaneous coronary intervention, BMC Cardiovasc. Disord. 18 (1) (2018 Mar 2) 44, https://doi.org/10.1186/s12872-018-0782-8. PMID: 29499644; PMCID: PMC5833061.

[37] F. Tavares, P.I.M. Moraes, J.M. Souza, A.H. Barbosa, E.M. Santos, J.A. Marcondes, M.D. Oliveira, I. Gonçalves Jr., C.V.A. Espírito Santo, H.T. Bianco, F.A.H. Fonseca, V. Moisés, A. Caixeta, Prognostic role of neutrophil-to-lymphocyte ratio in patients with ST-elevation myocardial infarction undergoing to pharmaco-invasive strategy, Cardiovasc. Revasc. Med. 34 (2022 Jan) 99–103, https://doi.org/10.1016/j.carrev.2021.01.027. Epub 2021 Jan 29. PMID: 33736961.

[38] S. Zhang, J. Diao, C. Qi, J. Jin, L. Li, X. Gao, L. Gong, W. Wu, Predictive value of neutrophil to lymphocyte ratio in patients with acute ST segment elevation myocardial infarction after percutaneous coronary intervention: a meta-analysis, BMC Cardiovasc. Disord. 18 (1) (2018 May 2) 75, https://doi.org/10.1186/s12872-018-0812-6. PMID: 29716535; PMCID: PMC5930503.