

ШАХГЕЛЬДЯН К.И.,

д.т.н., доцент, Владивостокский государственный университет, Владивосток, Россия,
e-mail: carinashakh@gmail.com

ГЕЛЬЦЕР Б.И.,

член-корр. РАН, д.м.н., профессор, Дальневосточный федеральный университет,
г. Владивосток, Россия, e-mail: boris.geltser@vvsu.ru

КУКСИН Н.С.,

Дальневосточный федеральный университет, г. Владивосток, Россия, e-mail: nikita.kuksin@vvsu.ru

ДОМЖАЛОВ И.Г.,

Дальневосточный федеральный университет, г. Владивосток, Россия, e-mail: igor@domzhalov.ru

МНОГОУРОВНЕВАЯ КАТЕГОРИЗАЦИЯ НЕПРЕРЫВНЫХ ПЕРЕМЕННЫХ В ЗАДАЧАХ ОБЪЯСНЕНИЯ ПРОГНОЗНЫХ ОЦЕНОК МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ В КЛИНИЧЕСКОЙ МЕДИЦИНЕ

DOI: 10.25881/18110193_2023_3_44

Аннотация. Цель. Сравнительная оценка качества прогностических моделей внутригоспитальной летальности (ВГЛ) у больных инфарктом миокарда с подъемом сегмента ST (ИМnST) после чрескожного коронарного вмешательства (ЧКВ), разработанных на основе предикторов в непрерывной, дихотомической и многоуровневой категориальной формах.

Материалы и методы. Проведено одноцентровое ретроспективное исследование, в рамках которого анализировали данные 4677 историй болезни пациентов с ИМnST после ЧКВ, находившихся на лечении в Региональном сосудистом центре г. Владивостока. Было выделено 2 группы больных: первая- 318 (6,8%) человек, умерших в стационаре, вторая — 4359 (93,2%) — с благоприятным исходом лечения. Прогностические модели ВГЛ с непрерывными переменными были разработаны методами многофакторной логистической регрессии, случайного леса и стохастического градиентного бустинга. Дихотомизация предикторов выполнялась методами поиска на сетке оптимальных точек отсечения, расчета центроидов и аддитивного объяснения Шепли (SHAP). Для многоуровневой категоризации предложено использовать объединение пороговых значений, выделенных при дихотомизации, а также ранжирование порогов отсечения с помощью весовых коэффициентов многофакторной логистической регрессии.

Результаты. По результатам многоступенчатого анализа показателей клинко-функционального статуса больных ИМnST были выделены и валидированы новые предикторы ВГЛ, выполнена их категоризация и разработаны прогностические модели с непрерывными, дихотомическими и многоуровневыми категориальными переменными (AUC: 0.885-0.902). Модели, предикторы которых были выделены методом мультиметрической категоризации, не уступали по точности моделям с непрерывными переменными и имели более высокие метрики качества, чем алгоритмы с дихотомическими предикторами. Преимущество моделей с многоуровневой категоризацией предикторов заключалось в возможности объяснения и клинической интерпретации результатов прогнозирования ВГЛ.

Заключение. Многоуровневая категоризация предикторов является перспективным инструментом для объяснения прогнозных оценок в клинической медицине.

Ключевые слова: прогностические модели; многоуровневая категоризация; дихотомизация; инфаркт миокарда с подъемом сегмента ST; внутригоспитальная летальность; метод аддитивного объяснения Шепли.

Для цитирования: Шахгельдян К.И., Гельцер Б.И., Куксин Н.С., Домжалов И.Г. Многоуровневая категоризация непрерывных переменных в задачах объяснения прогнозных оценок моделей машинного обучения в клинической медицине. Врач и информационные технологии. 2023; 3: 44-57. doi: 10.25881/18110193_2023_3_44.

SHAKHGELDYAN K.I.,

DSc, Associate Professor, Vladivostok State University, Vladivostok, Russia,
e-mail: carinashakh@gmail.com

GELTSER B.I.,

Corr. Member of the RAS, DSc, Professor, Far Eastern Federal University,
Vladivostok, Russia, e-mail: boris.geltser@vvsu.ru

KUKSIN N.S.,

Far Eastern Federal University, Vladivostok, Russia, e-mail: nikita.kuksin@vvsu.ru

DOMZHALOV I.G.,

Far Eastern Federal University, Vladivostok, Russia, e-mail: igor@domzhalov.ru

MULTILEVEL CATEGORIZATION OF CONTINUOUS VARIABLES IN THE TASKS OF EXPLAINING PREDICTIVE ESTIMATES OF MACHINE LEARNING MODELS IN CLINICAL MEDICINE

DOI: 10.25881/18110193_2023_3_44

Abstract. Aim: Comparative assessment of the quality of predictive models of in-hospital mortality (IHM) in patients with ST-segment elevation myocardial infarction (STEMI) after percutaneous coronary artery intervention (PCI), developed on the basis of predictors in continuous, dichotomous and multilevel categorical forms.

Materials and methods: This was a single-center retrospective study analyzing data from 4677 medical records of patients with STEMI PCI who were treated at the Regional Vascular Center of Vladivostok. Two groups of patients were identified: the first consisted of 318 (6.8%) patients who died in hospital, the second — 4359 (93.2%) patients with a favorable treatment outcome. Predictive models of IHF with continuous variables were developed using multivariate logistic regression, random forest, and stochastic gradient boosting. Dichotomization of predictors was performed using grid search methods for optimal cutoff points, centroid calculation, and Shapley additive explanation (SHAP). It was proposed for multi-level categorization to use a combination of threshold values identified during dichotomization, as well as ranking cut-off thresholds using multivariate logistic regression weighting coefficients.

Results: Based on the results of a multistage analysis of indicators of the clinical and functional status of STEMI patients, new predictors of IHM were identified and validated, their categorization was performed, and prognostic models with continuous, dichotomous and multilevel categorical variables were developed (AUC: 0.885-0.902). Models whose predictors were identified using the multimetric categorization method were not inferior in accuracy to models with continuous variables and had higher quality metrics than algorithms with dichotomous predictors. The advantage of models with multilevel categorization of predictors was the ability to explain and clinically interpret the results of IHM prediction.

Conclusions: Multilevel categorization of predictors is a promising tool for explaining predictive scores in clinical medicine.

Keywords: Predictive Models, Multi-level Categorization, Dichotomization, ST Elevation Myocardial Infarction, In-Hospital Mortality, Shapley Additive Explanation.

For citation: Shakhgeldyan K.I., Geltser B.I., Kuksin N.S., Domzhalov I.G. Multilevel categorization of continuous variables in the tasks of explaining predictive estimates of machine learning models in clinical medicine. Medical doctor and information technology. 2023; 3: 44-57. doi: 10.25881/18110193_2023_3_44.

ВВЕДЕНИЕ

В последние годы в клиническую практику все шире внедряются системы поддержки принятия врачебных решений, одной из задач которых является оценка вероятности развития неблагоприятных событий. Для реализации этой задачи разрабатываются прогностические модели на основе методов машинного обучения (МО), где в качестве предикторов используются показатели клинико-функционального статуса больных, представленные в непрерывной или категориальной формах. Категоризация определяется как преобразование непрерывных переменных с выделением не менее 2-х групп показателей, значения которых выше или ниже порогов отсечения [1]. В клинической медицине использование категориальных переменных имеет ряд преимуществ, так как отклонение анализируемых признаков от пороговых значений позволяет отнести их к факторам риска (ФР) неблагоприятных событий и дать клиническое обоснование предсказанной вероятности [2–3]. Кроме того, на основе комбинации ФР реализуется возможность формирования их фенотипов, характеризующих особенности комплексного воздействия различных признаков на результирующую переменную [4]. Вместе с тем, по мнению ряда авторов, показатели в категориальной и, в частности, дихотомической форме могут приводить к потере информации, искажению результатов анализа [5–9], увеличению доли ложноположительных [10] и ложноотрицательных [11] заключений. Альтернативой дихотомизации является подход, основанный на многоуровневой категоризации медицинских данных, позволяющий более точно классифицировать ФР и принимать оптимальные решения по их устранению [12]. Несмотря на предостережения, связанные с проблемами категоризации данных, рекомендации STROBE (Strengthening the Reporting of Observational studies in Epidemiology) подтверждают целесообразность ее использования при условии указания авторами методов, которыми она проводится [13]. При этом вопрос о необходимости выполнения процедур категоризации до настоящего времени остается в зоне активной дискуссии.

Ишемическая болезнь сердца (ИБС) занимает лидирующую позицию в структуре смертности населения от сердечно-сосудистых заболеваний

[14]. К наиболее опасным клиническим вариантам ИБС относят инфаркт миокарда с подъемом сегмента ST на электрокардиограмме (ИМпST). Одним из эффективных методов лечения ИМпST является реваскуляризация миокарда посредством чрескожного коронарного вмешательства (ЧКВ) [15]. Несмотря на совершенствование технологий ЧКВ внутригоспитальная летальность (ВГЛ) после ее выполнения по экстренным показаниям остается высокой и варьирует от 4 до 7%, что актуализирует необходимость прогнозирования неблагоприятных событий [16]. К наиболее востребованным инструментам риск-метрии относят шкалу GRACE (Global Registry of Acute Coronary Events), совершенствование которой является целью ряда исследований последних лет [17–20]. В большинстве случаев для разработки прогностических моделей базовые факторы данной шкалы дополняют новыми предикторами. Важной задачей при этом является корректное выделение их пороговых значений, что может повысить точность и объяснимость результатов прогнозирования.

Цель исследования состояла в сравнительной оценке качества прогностических моделей ВГЛ у больных ИМпST после ЧКВ, разработанных на основе предикторов в непрерывной, дихотомической и многоуровневой категориальной формах.

МАТЕРИАЛЫ И МЕТОДЫ

Данные

Проведено одноцентровое когортное ретроспективное исследование, в рамках которого анализировали данные истории болезни 4677 пациентов (3203 мужчины и 1574 женщины) с ИМпST в возрасте от 26 до 93 лет с медианой 63 года и 95% доверительным интервалом (ДИ) [62,63], поступивших в Региональный сосудистый центр «Приморская краевая клиническая больница № 1» г. Владивостока в период с 2015 по 2021 гг. [21]. Всем больным выполнено экстренное ЧКВ. Пациенты были разделены на 2 группы, первую из них составили 318 (6,8%) человек, умерших в стационаре, вторую — 4359 (93,2%) — с благоприятным исходом. Исследование соответствовало принципам Хельсинкской декларации и одобрено локальным этическим комитетом ДВФУ (протокол №8 от 8.06.2023 г.).

Критерии включения в исследование: больные с подтвержденным ИМпST и выполнением ЧКВ в первые сутки стационарного лечения. Критерии невключения: нестабильная стенокардия, ИМ без подъема сегмента ST и отсутствие показаний для ЧКВ.

Клинико-функциональный статус больных оценивали в первые сутки стационарного лечения с помощью 136 факторов, наиболее значимые из которых представлены в таблице 1. Данные включали 5 признаков шкалы GRACE: возраст больных, класс острой сердечной недостаточности (ОСН) по Т. Killip, частоту сердечных сокращений (ЧСС), систолическое артериальное давление (САД), концентрацию креатинина (Cr) в сыворотке крови. Анализировали содержание в крови нейтрофилов (NEUT), эозинофилов (EOS), уровень тромбокрит (PCT), глюкозы (Gl), эхокардиографические индикаторы, в том числе фракцию выброса (ФВ) левого желудочка (ЛЖ) и др.

Конечная точка исследования была представлена показателем ВГЛ у больных ИМпST после ЧКВ от всех причин в форме категориального бинарного признака («отсутствие» или «развитие»).

Методы статистического анализа и машинного обучения. Распределение данных по тесту Колмогорова-Смирнова отличалось от нормального, в связи с чем использовали непараметрические методы статистики. Показатели были представлены медианой (Me) и межквартильными интервалами (Q1; Q3), для межгрупповых сравнений непрерывных переменных применяли тест Манна-Уитни, а для категориальных — χ^2 . Отношение шансов (ОШ) и их 95% ДИ рассчитывали точным тестом Фишера. Статистически значимыми считали различия при $p < 0,01$.

Прогностические модели ВГЛ с предикторами в непрерывной форме разрабатывали методами многофакторной логистической регрессии (МЛР), случайного леса (СЛ) и стохастического градиентного бустинга (СГБ). Модели, в которых использовались категориальные предикторы, разрабатывали только на основе МЛР и СГБ. Их качество оценивали по 6 метрикам: площадь под ROC-кривой (AUC), чувствительность (Sen), специфичность (Spec), F1-score, positive predictive value (PPV), negative predictive value (NPV).

Для дихотомизации потенциальных предикторов использовали методы оптимизации на сетке с шагом $\Delta = (\max - \min) / 100$: минимизацию

p-value — Min(p-value), максимизацию AUC — Max(AUC), квартили, метод центроидов и аддитивного объяснения Шепли (SHAP) [22, 23]. Последний применяли также для оценки степени влияния предикторов модели на конечную точку. Для выполнения многоуровневой категоризации нами предложено 2 новых подхода. Первый из них предполагал использование пороговых значений, полученных по результатам анализа SHAP-value (рис. 1) Второй, мультиметрический, объединял все пороговые значения, выделенные при дихотомизации показателей различными методами, в том числе многоуровневой категоризации по SHAP. При этом близкие пороговые значения объединились в одно путем усреднения. Кроме того, в работе применяли 2 процедуры категоризации, предложенные ранее. В первой использовали медианы анализируемых факторов в группах сравнения и равноудаленные от них значения (центроиды), с помощью которых выделяли 4 категории для каждого показателя [22]. Вторая процедура была связана с выделением 4-х категорий для каждой переменной по результатам оценки их медиан, 2 и 3 квартили [24].

Датасет был разделен на 2 выборки: для обучения и кросс-валидации (80%) и для итогового тестирования (20%). Процедура обучения и кросс-валидации выполнялась методом стратифицированного k-Folders на 10 выборках. Усредненные метрики качества AUC, Sen и Spec использовали для выбора лучшей модели, отбора и валидации предикторов и подбора оптимальных гиперпараметров путем перебора по сетке допустимых значений. Для заключительного тестирования лучшие модели МЛР, СЛ и СГБ с оптимальными параметрами и гиперпараметрами обучали на 80% датасета, а проверяли на подгруппе для итогового тестирования. Для достоверительной оценки метрик качества процедуру повторяли 500 раз, выполняя первоначальное разделение случайным образом методом Монте-Карло.

Анализ данных и построение моделей проводились в Python с открытым исходным кодом, версии 3.9.16.

ДИЗАЙН ИССЛЕДОВАНИЯ

Дизайн исследования включал 5 этапов. На первом из них с помощью тестов межгрупповых

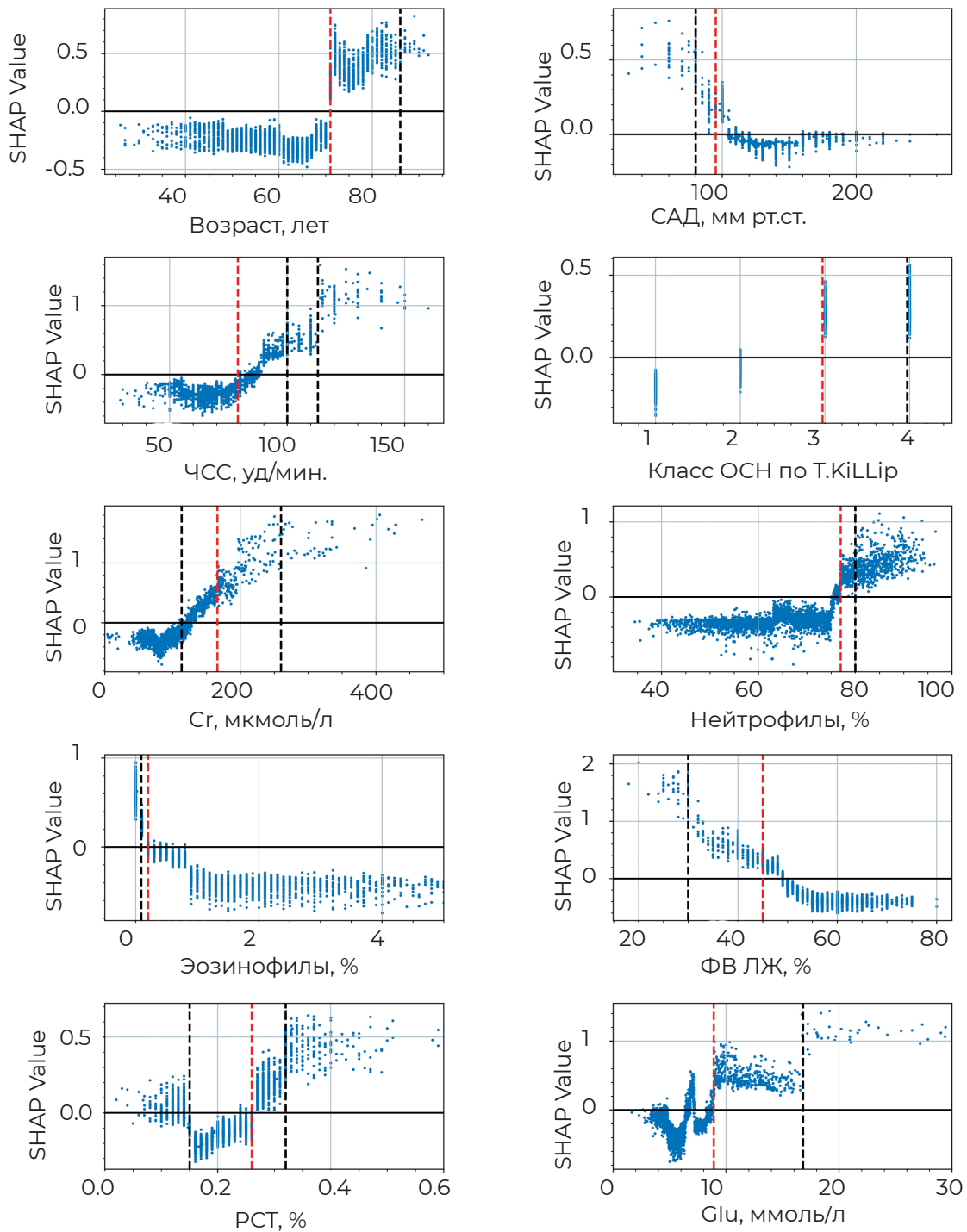


Рисунок 1 — Оценка влияния непрерывных показателей и их пороговых значений на конечную точку по методу SHAP. Примечание: пунктирные линии обозначают пороги отсечения. Сокращения: ФВ ЛЖ — фракция выброса левого желудочка, Glu — глюкоза, РСТ — тромбоциты; САД — систолическое артериальное давление, ЧСС — частота сердечных сокращений, Cr — креатинин.

сравнений формировался пул потенциальных предикторов ВГЛ (табл. 1). На втором этапе исследования с помощью методов МО разрабатывали прогностические модели ВГЛ с предикторами в непрерывной форме, включающие 5 базовых факторов шкалы GRACE. Для повышения точности прогноза структуру моделей пошагово дополняли новыми предикторами,

выделенными на первом этапе исследования, при условии статистически значимого различия на уровне $p\text{-value} < 0.01$. Прогностическая значимость предиктора подтверждалась, если после его включения в модель значение AUC возрастало. На третьем этапе с помощью различных методов выделения пороговых значений осуществляли бинаризацию непрерывных переменных

Таблица 1 — Клинико-функциональная характеристика больных ИМпST

Предиктор	Группа 1 (n = 318)	Группа 2 (n = 4357)	ОШ (95%) ДИ	p-value
Пол: Ж, абс. (%)	142 (44.65)	1332 (30.5)	1.8 [1.5; 2.3]	<0.000001
Возраст, лет	71 (63; 78)	62 (55; 69)	–	<0.000001
Рост, см	168 (164; 174)	170 (165; 176)	–	0.000001
Вес, кг	78 (70; 85)	80 (71; 90)	–	0.000012
САД, мм рт.ст.	110 (90; 130)	130 (120; 150)	–	<0.000001
ДАД, мм рт.ст.	70 (60; 80)	80 (75; 90)	–	<0.000001
ЧСС, уд/мин	86 (72; 100)	72 (65; 80)	–	<0.000001
Сг, мкмоль/л	130 (96; 193.3)	97 (81; 114.8)	–	<0.000001
Класс ОЧН по Т.Killip				
I	71 (22.33%)	2726 (62.6%)	0.17 [0.13; 0.23]	<0.000001
II	58 (18.2%)	867 (19.9%)	0.9 [0.67; 1.20]	0.508052
III	66 (20.75)	479 (11)	2.1 [1.6; 2.8]	<0.000001
IV	123 (38.7)	269(6.18)	9.6 [7.4; 12.4]	<0.000001
III-IV	189 (59.4)	748 (17.2)	7.1 [5.6; 9]	<0.000001
ФВ ЛЖ, %	46.5 (38; 54.8)	56 (50; 61)	–	<0.000001
NEUT, %	81.3 (75.75; 86.5)	66.7 (59.1; 74.9)	–	<0.0001
EOS, %	0.1 (0.00; 0.3)	0.9 (0.3; 1.9)	–	<0.000001
Glu, ммоль/л	7.9 (6.3; 10.31)	5.8 (5.1; 7)	–	<0.000001
PCT, %	0.22 (0.17; 0.28)	0.2 (0.16; 0.24)	–	0.0012
Передний ИМ, абс. (%)	177 (55.66%)	2017 (46.27%)	1.5 [1.16; 1.83]	0.00147
ФП, абс. (%)	129 (40.57%)	772 (17.69%)	3.2 [2.51; 4.02]	<0.000001
СД 2, абс. (%)	99 (31.13%)	831 (19.05%)	1.9 [1.50; 2.46]	<0.000001
ХБП, абс. (%)	83 (26.1%)	677 (15.5%)	1.97 [1.5; 2.6]	<0.000001

Сокращения: ФВ ЛЖ — фракция выброса левого желудочка, Glu — глюкоза, PCT — тромбокрит; САД — систолическое артериальное давление, ДАД — диастолическое артериальное давление, ЧСС — частота сердечных сокращений, Сг — креатинин, ФП — фибрилляция предсердий, СД 2 – сахарный диабет второго типа, ХБП — хроническая болезнь почек.

и на их основе разрабатывали прогностические модели ВГЛ. На четвертом этапе исследования проводили многоуровневую категоризацию переменных. В рамках каждого подхода формировали группы потенциальных предикторов ВГЛ. Для оценки степени их влияния на конечную точку были разработаны модели МЛР, весовые коэффициенты которых использовали для кодирования многоуровневых категориальных предикторов. На пятом этапе, используя СГБ, разработаны 4 новых прогностических модели ВГЛ, предикторы которых были получены разными методами многоуровневой категоризации.

РЕЗУЛЬТАТЫ

Межгрупповой анализ клинико-демографических и лабораторных показателей продемонстрировал, что большинство из них, в том числе, все предикторы шкалы GRACE имеют статистически значимые различия (табл. 1). Среди умерших преобладали лица старшего возраста, женского пола и более низкого роста. Для первой группы пациентов было характерным наличие 3 и 4 класса ОЧН по Т.Killip, меньшие значения САД, ДАД, ФВ ЛЖ, EOS и некоторых других показателей, более высокие уровни ЧСС, Cr, NEUT. У них чаще фиксировался передний ИМ, сахарный диабет 2-го типа (СД2), фибрилляция предсердий (ФП) и хроническая болезнь почек (ХБП).

На втором этапе исследования были разработаны прогностические модели ВГЛ, где помимо 5 базовых показателей шкалы GRACE в непрерывной форме использовали новые факторы. Последние были выделены в результате тестирования пула потенциальных предикторов в прогностических моделях. К ним относились показатели ФВ ЛЖ, NEUT, EOS, PCT и Glu, при последовательном включении которых в структуру моделей МЛР, СГБ и СЛ, точность прогноза по метрике AUC возрастала, что свидетельствовало о предиктивной ценности анализируемых факторов. Необходимо также отметить, что сопоставление значений AUC модели, разработанной нами ранее на этой же выборке больных с использованием только предикторов шкалы GRACE [25], и модели, дополненной новыми предикторами, демонстрировало более высокую точность последней (AUC — 0,9 vs 0,836).

Сопоставление предсказательной ценности разработанных моделей МЛР, СГБ и СЛ показало, что они обладают высокой прогностической способностью и имеют близкие значения метрик качества при кросс-валидации и итоговом тестировании (AUC варьировала от 0,884 до 0,9), что свидетельствует об отсутствии их переобучения и хороших обобщающих свойствах (табл. 2).

На третьем этапе исследования выполняли дихотомизацию предикторов ВГЛ в

Таблица 2 — Оценка точности прогностических моделей ВГЛ у больных ИМпСТ после ЧКВ с использованием предикторов в непрерывной форме

Метрики	Кросс-валидация			Итоговое тестирование		
	МЛР	СГБ	СЛ	МЛР	СГБ	СЛ
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	0.275 [0.256; 0.295]	0.248 [0.226; 0.27]	0.231 [0.214; 0.249]	0.278 [0.239; 0.317]	0.253 [0.215; 0.291]	0.235 [0.189; 0.28]

непрерывной форме методами поиска оптимального порога отсека на сетке, методом SHAP и с помощью расчета центраида (табл. 3). Использование пороговых значений, отклонение от которых ассоциируется с увеличением вероятности ВГЛ, позволяет рассматривать бинаризованные данные как ФР неблагоприятных событий. ФР кодируется "1", если значение предиктора превышает порог с постфиксом "+" или не достигает его — с постфиксом "-", в остальных случаях — "0".

Результаты исследования показали, что пороговые значения, полученные различными методами бинаризации, часто отличались между собой. Так, например, порог отсека для показателя ФВ ЛЖ по версии SHAP составлял <45%, в то время как при максимизации AUC точка отсека фиксировалась на уровне 51% (табл. 3). При этом пороговые значения, определенные алгоритмом SHAP, были ближе всего к критериальным границам, верифицированным методом Min(p-value). Необходимо отметить, что метод SHAP позволяет не только определить пороговые границы, но и оценить интенсивность влияния на ВГЛ показателей, значения которых находятся в "зоне риска". К

категориальным факторам, выделенным этим методом, относились следующие признаки: возраст > 70 лет, ЧСС > 89 уд./мин, САД < 95 мм рт.ст., класс ОЧН по Т. Killip > 2, Cr > 166 ммоль/л, ФВ ЛЖ < 45%, NEUT > 77%, EOS < 0.22%, PCT > 0.32%, Glu > 8.9 ммоль/л (рис. 1). На примере диаграммы ФВ ЛЖ видно, что вероятность ВГЛ последовательно увеличивается в диапазоне значений данного показателя 44% — 31% и резко возрастает при его уровне < 30%. Повышение концентрации глюкозы в крови более 8.9 ммоль/л увеличивает риск ВГЛ, но вероятность фатального исхода значительно возрастает при ее уровне более 17 ммоль/л. Оценка динамики изменений SHAP-value, позволяющая объяснить взаимосвязь различных значений предикторов с конечной точкой исследования, явилась основанием для использования этого метода в процедурах многоуровневой категоризации.

На основе МЛР были разработаны прогностические модели ВГЛ с дихотомическими предикторами (табл. 4). Было установлено, что большинство из них обладает высокой предиктивной способностью вне зависимости от метода определения пороговых значений (AUC: 0.885 — 0.896).

Таблица 3 — Дихотомизация непрерывных предикторов ВГЛ с помощью различных методов

Предиктор	Min(p-value)	Max(AUC)	Центроид	SHAP
Возраст	70+	65+	66.5+	71+
САД	92-	112-	120-	95-
ЧСС	95+	82+	79+	89+
Класс ОЧН по Т. Killip	4	3+	3+	3+
Cr, ммоль/л	188.6+	122.9+	113.3+	166+
NEUT, %	78.8+	75+	74.0+	77+
EOS, %	0.3-	0.3-	0.5-	0.2-
ФВ ЛЖ, %	31.0-	51-	51-	45-
PCT, %	0.36+	0.22+	0.21+	0.32+
Glu, ммоль/л	8.4+	6.5+	6.9+	8.9+

Сокращения: ФВ ЛЖ — фракция выброса левого желудочка (ЛЖ), Glu — глюкоза, PCT — тромбокрит; САД — систолическое артериальное давление, ЧСС — частота сердечных сокращений, Cr — креатинин.

Таблица 4 — Оценка точности прогностических моделей ВГЛ на основе предикторов в дихотомической форме

Метрики	SHAP	Min(p-value)	Max(AUC)	Центроиды
AUC	0.885 [0.882; 0.887]	0.876 [0.873; 0.879]	0.896 [0.893; 0.898]	0.888 [0.885;0.890]
Sen	0.815 [0.809; 0.821]	0.815 [0.809; 0.821]	0.815 [0.809; 0.821]	0.815 [0.808;0.821]
Spec	0.825 [0.823; 0.826]	0.826 [0.824; 0.828]	0.821 [0.819; 0.822]	0.823 [0.822;0.825]
PPV	0.149 [0.148; 0.15]	0.147 [0.116; 0.178]	0.148 [0.147; 0.149]	0.15 [0.149; 0.151]
NPV	0.991 [0.991;0.992]	0.991 [0.991;0.992]	0.991 [0.991;0.992]	0.991 [0.991;0.992]
F-score	0.25 [0.248; 0.252]	0.251 [0.25; 0.253]	0.251 [0.25; 0.253]	0.254 [0.252;0.256]

На четвертом этапе исследования на основе использования 4-х методов многоуровневой категоризации были сформированы группы ФР ВГЛ. Первый пул ФР был получен при анализе SHAP-value (рис.). Второй — расширял первый пул ФР дихотомическими пороговыми значениями, полученными на предыдущем этапе. Третья группа включала медианы предикторов в группах сравнения и их центроиды, а четвертая — использовала пороговые значения, соответствующие квартилям предикторов. Для кодирования значений многоуровневых категориальных предикторов использовали весовые коэффициенты (ВК) моделей МЛР, разработанных для каждой группы ФР (табл. 5).

На пятом этапе на основе многоуровневых предикторов, полученных различными методами, были разработаны 4 прогностические модели ВГЛ на основе СГБ (табл. 6). Лучшие прогностические свойства демонстрировала модель с предикторами, выделенными методом мультиметрической категоризации. Последняя была сопоставима по точности с моделью, включающей непрерывные переменные, но в отличие от нее позволяла объяснять прогноз ВГЛ по результатам оценки пороговых значений и ВК ее предикторов (табл. 5). С учетом этих данных было установлено, что у больных ИМпСТ после ЧКВ наибольшая вероятность развития неблагоприятного исхода ассоциируется с показателями ФВ ЛЖ < 31% (ВК — 1.735) и Cr > 189 ммоль/л (ВК — 1.489). К менее значимым факторам риска ВГЛ относятся ЧСС > 95 уд/мин (ВК — 1.151), уровень глюкозы > 17 ммоль/л (ВК — 1.12) и содержание эозинофилов < 0.09% (ВК — 0.935).

Интенсивность воздействия на конечную точку 4-го класса ОЧН по T. Killip, возраста > 71 года, содержания нейтрофилов > 77% и концентрации креатинина в диапазоне 166–189 ммоль/л характеризовалась значениями ВК на уровне 0.75-0.86. Остальные категориальные значения ФР оказывали на ВГЛ существенно меньшее влияние. Сравнительная оценка точности прогностических моделей с предикторами, выделенными методами дихотомизации и многоуровневой категоризации, демонстрировала преимущества последних, что подтверждалось статистически значимыми различиями метрики AUC (p-value<0,00001).

ОБСУЖДЕНИЕ

В последние годы на основе методов МО разрабатываются прогностические модели, применение которых в клинической практике ограничивается сложностью интерпретации результатов прогноза. Перспективными инструментами для решения этой задачи являются алгоритмы объяснимого искусственного интеллекта (ОИИ), к которым относят определение пороговых значений, используемых в моделях ФР, и их ранжирование по интенсивности влияния на реализацию конечной точки [26]. Одним из подходов к определению пороговых значений ФР ВГЛ является их категоризация, которая позволяет детализировать соотношения показателей клиничко-функционального статуса больных ИМпСТ с результирующей переменной. По данным литературы, наиболее доступным методом многоуровневой категоризации является дескриптивная статистика с расчетом медиан,

Таблица 5 — Весовые коэффициенты и пороги предикторов, полученные методами многоуровневой категоризации

Предикторы	SHAP		Мультиметрическая категоризация		Медианы групп и центроид		Квартили	
	Пороги	ВК	Пороги	ВК	Пороги	ВК	Пороги	ВК
Возраст	[71;86] 86+	0.902 0.839	[65;71] 71+	0.112 0.873	[66.5;71] 71+	0.67 0.857	[63;70] 70+	0.018 0.85
САД	80- [80;95]	0.975 0.09	60- [60;92] [92;112] [112; 120]	0.276 0.54 0.352 0.364	110- [110;120]	0.842 0.146	120-	0.367
ЧСС	[79;100] [100;113] 113+	0.422 0.86 1.343	[82;95] 95+	0.538 1.151	86+	1.012	81+	0.976
класс ОСН по Т. Killip	3 4	0.502 0.879	3 4	0.512 0.859	3 4	0.261 0.689	3 4	0.403 1.025
Cr	[113;166] [166;260] 260+	0.431 1.094 1.694	[123;166] [166;189] 189+	0.456 0.774 1.489	[97;113] [113;130] 130+	0.239 0.434 1.062	[98;117] 117+	0.263 1.035
Нейтрофилы	[77;80] 80+	0.847 0.782	[75;77] [77;79] 79+	0.405 0.78 0.75	[74;81.3] 81.3+	0.5 0.758	75.8+	0.832
Эозинофилы	0.09- [0.09;0.2]	0.883 0.484	0.09- [0.09;0.3]	0.935 0.371	0.1- [0.1;0.5] [0.5;0.9]	1.07 0.566 0.695	0.2- [0.2;0.8] [0.8;1.8]	0.913 0.41 0.167
ФВ ЛЖ	30- [30;45]	1.255 0.794	31- [31;45] [45;51]	1.735 0.566 0.311	46.5- [46.5;51]	0.916 0.422	49- [49;56]	0.815 0.039
PCT	0.15- [0.15;0.26] [0.26;0.32] 0.32+	0.304 0 0.374 0.658	[0.31;0.36] 0.36+	0.551 0.582	[0.2;0.21] 0.22+	0.324 0.258	0.16- 0.25+	0.56 0.565
Glu	[8.9;17] 17+	0.856 1.373	[6.5;6.9] [6.9;8.9] [8.9;17] 17+	0.368 0 0.648 1.12	[5.8;6.9] [6.9;7.9] 7.92+	0.077 0.236 0.949	[5.83;7.1] 7.1+	0.33 0.77

Сокращения: ФВ ЛЖ — фракция выброса левого желудочка (ЛЖ), Glu — глюкоза, PCT — тромбокрит; САД — систолическое артериальное давление, ЧСС — частота сердечных сокращений, Cr — креатинин, ВК — весовые коэффициенты.

квартилей или квантилей [1, 9, 27]. Вместе с тем большая часть критических замечаний по категоризации связана именно с этим подходом, что обусловлено, прежде всего, зависимостью таких пороговых значений от конкретной выборки, отсутствием взаимосвязи с клиническим

контекстом, игнорированием возможных нелинейных отношений и др. Другой подход основан на выделении порогового значения известного из практики как выход за пределы нормы, который не связан с клиническим контекстом прогнозируемого неблагоприятного события [27].

Таблица 6 — Оценка точности прогностических моделей ВГЛ на основе предикторов с многоуровневой категоризацией

Метрики	Мультиметрическая категоризация	Категоризация методом SHAP	Медианы групп и центроид	Квартили
AUC	0.902 [0.9; 0.903]	0.897 [0.894; 0.898]	0.894 [0.892; 0.896]	0.887 [0.885; 0.889]
Sen	0.825 [0.819; 0.831]	0.825 [0.82; 0.83]	0.825 [0.819; 0.831]	0.8 [0.795; 0.805]
Spec	0.834 [0.833; 0.836]	0.83 [0.828; 0.831]	0.817 [0.815; 0.818]	0.806 [0.805; 0.808]
PPV	0.157 [0.156; 0.158]	0.153 [0.152; 0.154]	0.146 [0.145; 0.147]	0.14 [0.139; 0.141]
NPV	0.991 [0.991; 0.991]	0.992 [0.992; 0.992]	0.991 [0.991; 0.991]	0.991 [0.991; 0.991]
F-score	0.259 [0.258; 0.261]	0.258 [0.256; 0.259]	0.249 [0.248; 0.25]	0.24 [0.239; 0.242]

Альтернативным методом, учитывающим клинический контекст, является поиск оптимальных пороговых значений на основе минимизации или максимизации целевых функций, например, $\text{Min}(p\text{-value})$ или $\text{Max}(AUC)$. По нашему мнению, категоризация переменных должна рассматриваться только в рамках решения конкретной клинической задачи, хотя выделенные предикторы могут представлять интерес для реализации других прогностических исследований.

В нашей работе на примере базы данных больных ИМПСТ после ЧКВ анализировали эффективность различных методов поиска пороговых значений предикторов, отклонение от которых повышало их предиктивный потенциал и позволяло отнести к ФР ВГЛ. Было установлено, что метод SHAP, который рассматривают как технологию ОИИ, является полезным инструментом категоризации за счет эффективного определения порогов отсечения, в том числе, для многоуровневой категоризации и анализа взаимосвязи предикторов в непрерывной и категориальной формах с конечной точкой исследования. Так, например, с помощью данного метода для показателя ФВ ЛЖ наряду с границей 45% был выделен порог на уровне 31%, ниже которого риск ВГЛ резко возрастает. В то же время было показано, что многоуровневые категориальные предикторы, полученные путем объединения данных SHAP с результатами других методов дихотомизации, обеспечивают более высокую прогностическую точность.

В нашем исследовании потенциальные риски потери информации при

использовании новых методов категоризации были преодолены за счет детализации знаний о взаимосвязи отдельных ФР или их комбинаций с конечной точкой. Это подтверждалось сопоставлением критериев качества прогностических моделей с предикторами в непрерывной и многоуровневой категориальной формах. Так, для лучшей модели с непрерывными предикторами AUC составляла 0.9, а при использовании мультиметрической категоризации — 0.902. Методы категоризации, основанные на дескриптивных статистиках, не компенсировали потерю прогностической точности за счет привнесения новой информации.

Ограничения исследования связаны с его ретроспективным характером, необходимостью валидации моделей на данных из других лечебных учреждений.

ЗАКЛЮЧЕНИЕ

В настоящем исследовании на основе базы данных пациентов с ИМПСТ после ЧКВ апробированы 2 новых подхода к многоуровневой категоризации предикторов ВГЛ: мультиметрическая категоризация и метод SHAP. С помощью разработанных прогностических моделей ВГЛ показано, что предложенные авторами процедуры категоризации обеспечивают высокую точность и прозрачность результатов прогнозирования.

Источник финансирования. Исследование выполнено в рамках проекта Российского научного фонда (РНФ) № 23-21-00250, <https://rscf.ru/project/23-21-00250/>

ЛИТЕРАТУРА/REFERENCES

1. Mabikwa OV, Greenwood DC, Baxter PD, Fleming SJ. Assessing the reporting of categorised quantitative variables in observational epidemiological studies. *BMC Health Serv Res.* 2017; 17(1): 201. doi:10.1186/s12913-017-2137-z.
2. MacCallum RC, Zhang S, Preacher KJ, Rucker DD. On the practice of dichotomization of quantitative variables. *Psychol Methods.* 2002; 7(1): 19-40. doi:10.1037/1082-989x.7.1.19.
3. Gupta R, Day CN, Tobin WO, Crowson CS. Understanding the effect of categorization of a continuous predictor with application to neuro-oncology. *Neurooncol Pract.* 2021; 9(2): 87-90. doi:10.1093/nop/npab049.
4. Гельцер Б.И., Шахгельдян К.И., Рублев В.Ю. и др. Фенотипирование факторов риска и прогнозирование внутригоспитальной летальности у больных ишемической болезнью сердца после коронарного шунтирования на основе методов объяснимого искусственного интеллекта // Российский кардиологический журнал. — 2023. — Т.28. — №4. — С.85-93. [Geltser BI, Shakhgeldyan KI, Rublev VYu, Domzhalov IG, Tsivanyuk MM, Shekunova OI. 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 Journal of Cardiology.* 2023; 28(4): 5302. (In Russ.) doi:10.15829/1560-4071-2023-5302.
5. Altman DG, Lausen B, Sauerbrei W, Schumacher M. Dangers of using «optimal» cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst.* 1994; 86(11): 829-835. doi:10.1093/jnci/86.11.829.
6. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Med.* 2015; 13: 1. doi:10.1186/s12916-014-0241-z.
7. Dawson NV, Weiss R. Dichotomizing continuous variables in statistical analysis: a practice to avoid. *Med Decis Making.* 2012; 32(2): 225-226. doi:10.1177/0272989X12437605.
8. Salis Z, Gallego B, Sainsbury A. Researchers in rheumatology should avoid categorization of continuous predictor variables. *BMC Med Res Methodol.* 2023; 23(1): 104. doi:10.1186/s12874-023-01926-4.
9. Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ.* 2006; 332(7549): 1080. doi:10.1136/bmj.332.7549.1080.
10. Austin PC, Brunner LJ. Inflation of the type I error rate when a continuous confounding variable is categorized in logistic regression analyses. *Stat Med.* 2004; 23(7): 1159-1178. doi:10.1002/sim.1687.
11. Streiner DL. Breaking up is hard to do: the heartbreak of dichotomizing continuous data. *Can J Psychiatry.* 2002; 47(3): 262-266. doi: 10.1177/070674370204700307.
12. Chen Y, Huang J, He X, et al. A novel approach to determine two optimal cut-points of a continuous predictor with a U-shaped relationship to hazard ratio in survival data: simulation and application. *BMC Med Res Methodol.* 2019; 19(1): 96. Published 2019. doi:10.1186/s12874-019-0738-4.
13. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet.* 2007; 370(9596): 1453-1457. doi:10.1016/S0140-6736(07)61602-X.
14. The World Health Organization, The top 10 causes of death [Internet]. 2019. Available from: <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death> [cited 2023 Nov 30].
15. Ibáñez B, James S, Agewall S, et al. 2017 ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation. *Rev Esp Cardiol (Engl Ed).* 2017; 70(12): 1082. doi:10.1016/j.rec.2017.11.010.

16. Pfuntner A, Wier LM, Stocks C. Most Frequent Procedures Performed in U.S. Hospitals, 2011. In: Healthcare Cost and Utilization Project (HCUP) Statistical Briefs. Rockville (MD): Agency for Healthcare Research and Quality (US); October 2013.
17. Wang JJ, Fan Y, Zhu Y, et al. Biomarkers enhance the long-term predictive ability of the KAMIR risk score in Chinese patients with ST-elevation myocardial infarction. *Chin Med J*. 2019; 132(1): 30-41. doi:10.1097/CM9.000000000000015.
18. Liu XJ, Wan ZF, Zhao N, et al. 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*. 2015; 14: 110. doi: 10.1186/s12933-015-0274-4.
19. Chen X, Shao M, Zhang T, et al. 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*. 2020; 19(6): 3664-3674. doi: 10.3892/etm.2020.8626.
20. Wenzl FA, Kraler S, Ambler G, 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*. 2022; 400(10354): 744-756. doi: 10.1016/S0140-6736(22)01483-0.
21. Гельцер Б.И., Шахгельдян К.И., Домжалов И.Г., и др. Прогностическая оценка клинико-функционального статуса пациентов с инфарктом миокарда с подъемом сегмента ST после чрескожного коронарного вмешательства. Свидетельство о регистрации базы данных 2023622740, 10.08.2023. Заявка № 2023622516 от 28.07.2023. [Geltser BI, Shakhgelyan KI, Domzhalov IG, et al. Prognosticheskaya ocenka kliniko-funkcional'nogo statusa pacientov s infarktomiokarda s pod'emom segmenta ST posle chreskozhnogo koronarnogo vmeshatel'stva. Certificate of registration of the database 2023622740, 10.08.2023. Request № 2023622516. 28.07.2023.]

22. Valente F, Henriques J, Paredes S, et al. A new approach for interpretability and reliability in clinical risk prediction: Acute coronary syndrome scenario. *Artif Intell Med.* 2021; 117: 102113. doi: 10.1016/j.artmed.2021.102113.
23. Lundberg SM, Lee SI. 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; 2017 Dec 04-09; Long Beach, USA.* doi: 10.48550/arXiv.1705.07874.
24. Evenson KR, Wen F, Herring AH. Associations of Accelerometry-Assessed and Self-Reported Physical Activity and Sedentary Behavior With All-Cause and Cardiovascular Mortality Among US Adults. *Am J Epidemiol.* 2016; 184(9): 621-632. doi: 10.1093/aje/kww070.
25. Гельцер Б.И., Шахгельдян К.И., Домжалов И.Г. и др. Прогнозирование внутригоспитальной летальности у больных острым инфарктом миокарда с подъемом сегмента ST после чрескожного коронарного вмешательства // *Российский кардиологический журнал.* — 2023. — Т.28 — №6. — С.31-39. [Geltser BI, Shahgeldyan KI, Domzhalov IG, et al. Prediction of in-hospital mortality in patients with ST-segment elevation acute myocardial infarction after percutaneous coronary intervention. *Russian Journal of Cardiology.* 2023; 28(6): 5414. (In Russ.)] doi: 10.15829/1560-4071-2023-5414.
26. Molnar C. *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable* [Internet]. North Charleston: Independently published. 2023. [cited 2023 Nov 30]. Available from <https://christophm.github.io/interpretable-ml-book>.
27. Turner EL, Dobson JE, Pocock SJ. Categorisation of continuous risk factors in epidemiological publications: a survey of current practice. *Epidemiol Perspect Innov.* 2010; 7: 9. doi:10.1186/1742-5573-7-9.