

## **Методы машинного обучения в прогнозировании внутригоспитальной летальности у больных ишемической болезнью сердца после коронарного шунтирования**

Гельцер Борис Израйльевич<sup>1</sup>, Шахгельдян Карина Иосифовна<sup>2</sup>, Рублев Владислав Юрьевич<sup>3</sup>, Котельников Владимир Николаевич<sup>1</sup>, Кригер Александра Борисовна<sup>2</sup>, Широбоков Василий Глебович<sup>2</sup>

<sup>1</sup> ФГАОУ ВО “Дальневосточный федеральный университет”, Школа биомедицины 690091, г. Владивосток, ул. Суханова 8. Far Eastern Federal University

<sup>2</sup> ФГБОУ ВО “Владивостокский государственный университет экономики и сервиса”, Институт информационных технологий 690014, г. Владивосток, Гоголя 41. Vladivostok State University of Economics

<sup>3</sup> ГБУЗ “Краевая клиническая больница №1”, 690091 г.Владивосток, ул. Алеутская 57. Regional Clinical Hospital №1

Адрес для корреспонденции.

690091, г. Владивосток, ул. Суханова 8. Дальневосточный федеральный университет”, Школа биомедицины Департамент клинической медицины профессору Котельникову Владимиру Николаевичу.

E-mail [671235@mail.ru](mailto:671235@mail.ru)

Моб.тел. 89024811235

## Сведения об авторах.

1. Гельцер Борис Израилевич, доктор медицинских наук, профессор, член-корреспондент Российской академии наук

ORCID 0000-0002-9250-557X

690091, Владивосток, ул. Суханова, 8. Дальневосточный федеральный университет. Директор департамента клинической медицины Школы биомедицины

Boris Geltser, MD, professor, corresponding member of the Russian Academy of Sciences. 690091, Vladivostok, Suhanova street, 4. Far Eastern federal university. Director of the department of clinical medicine of the School of biomedicine

2. Шахгельдян Карина Иосифовна, доктор технических наук, профессор.

690014, г. Владивосток, ул. Гоголя, 41, Владивостокский государственный университет экономики и сервиса. Директор института информационных технологий

ORCID 0000-0002-4539-685X

e-mail: carinash@vvsu.ru

Karina Shakhgeldyan, doctor of technical sciences, professor.

690014, Vladivostok, st. Gogolya, 41, Director of the Institute of Information Technologies of the Vladivostok State University of Economics and Service

3. Рублев Владислав Юрьевич

690000, Владивосток, ул. Алеутская, 57. Краевая клиническая больница № 1, врач сердечно-сосудистый хирург

ORCID 0000-0001-7620-4454

dr.rublev.v@gmail.com

Vladislav Rublev,

690000, Vladivostok, Aleutskaya street, 57. Regional Clinical Hospital No. 1, doctor cardiovascular surgeon

4. Котельников Владимир Николаевич, доктор медицинских наук,

690091, Владивосток, ул. Суханова, 8. Дальневосточный федеральный университет. Профессор департамента клинической медицины Школы биомедицины

ORCID 0000-0001-5830-1322

Vladimir Kotelnikov, MD. 690091, Vladivostok, Suhanova street, 4. Far Eastern federal university. Professor of the department of clinical medicine of the School of biomedicine

5. Кригер Александра Борисовна, кандидат физ-мат. наук, доцент.

690014, г. Владивосток, ул. Гоголя, 41, Владивостокский государственный университет экономики и сервиса. Доцент института информационных технологий

ORCID 0000-0002-1350-529x

e-mail: alexandra.kriger @vvsu.ru

Alexandra Krieger, candidate of technical sciences, associate professor.

690014, Vladivostok, st. Gogolya, 41, associate professor of the Institute of Information Technologies of the Vladivostok State University of Economics and Service

6. Ширококов Василий Глебович

690014, г. Владивосток, ул. Гоголя, 41, Владивостокский государственный университет экономики и сервиса. Магистрант института информационных технологий

ORCID 0000-0001-5531-9519

e-mail: wpn@inbox.ru :

Vasilij Shirobokov.

690014, Vladivostok, st. Gogolya, 41, Undergraduate of the Institute of Information Technologies of the Vladivostok State University of Economics and Service

## **Резюме**

**Цель.** Сравнительная оценка прогностической точности моделей внутригоспитальной летальности (ВГЛ) у больных ишемической болезнью сердца (ИБС) после коронарного шунтирования (КШ), разработанных на основе современных технологий машинного обучения (МО).

**Материалы и методы.** Проведен ретроспективный анализ 866 электронных историй болезни пациентов (685 мужчин и 181 женщины), оперированных в 2008-2018 гг. по поводу ИБС методом изолированного КШ. Анализировали результаты клинических, лабораторных и инструментальных исследований, полученных до КШ. Больные были разделены на 2 группы: в первую из которых вошли 35 (4%) человек, умерших в течение первых 30 суток после КШ, во вторую - 831 (96%), – с благоприятным исходом операции. Предикторы ВГЛ выделяли путем многоступенчатой процедуры отбора с анализом статистических гипотез и расчетом весовых коэффициентов. Для построения моделей и верификации предикторов использовали методы МО: многофакторную логистическую регрессию (ЛР), случайный лес (СЛ) и искусственные нейронные сети (ИНС). Точность моделей оценивали по трем метрикам: площадь под ROC-кривой (AUC), чувствительность и специфичность. Кросс-валидацию моделей проводили на тестовых выборках.

**Результаты.** Статистический анализ, расчет отношения шансов и весовых коэффициентов позволили выделить факторы риска ВГЛ с наибольшим предиктивным потенциалом. К ним относились 7 факторов из шкалы EuroSCORE II: фракция выброса (ФВ) менее 30%, ФВ 30-50%, возраст больных, недавно перенесенный инфаркт миокарда, поражения периферических артериальных бассейнов, экстренность КШ, хроническая сердечная недостаточность III-IV функциональных классов и 5 дополнительных предикторов (частота сердечных сокращений, систолическое артериальное давление, наличие аортального стеноза, индекс относительной толщины (ИОТ) задней стенки левого желудочка (ЛЖ) и относительный индекс массы миокарда ЛЖ (ОИММЛЖ). Авторские модели на основе методов ЛР, СЛ и ИНС имели

более высокие показатели AUC и чувствительности по сравнению с “классической” шкалой EuroSCORE II. Модель ИНС с включением предикторов ИОТ и ОИММЛЖ продемонстрировала максимальный уровень прогностической точности, что иллюстрировалось значениями метрик качества: AUC - 95%, чувствительность - 92%, специфичность - 98%.

**Заключение:** Использование современных технологий МО позволило разработать новый алгоритм отбора предикторов и высокоточные модели прогнозирования ВГЛ после КШ.

**Ключевые слова:** методы машинного обучения, прогностические модели, коронарное шунтирование, внутригоспитальная летальность.

**Работа выполнена при финансовой поддержке РФФИ (проекты № 18-29-03131 и 19-29-01-077).**

## **Machine learning methods for prediction of intrahospital mortality in patients with coronary heart disease after coronary artery bypass grafting**

Geltser B.I.<sup>1</sup>, Shahgeldyan K.J.<sup>1,2</sup>, Rublev V.Y.<sup>1,3</sup>, Kotelnikov V.N.<sup>1</sup>, Kriger A.B.<sup>2</sup>, Shirobokov V.G.<sup>2</sup>

<sup>1</sup>Far Eastern federal university. School of biomedicine. 690091, Vladivostok, Suhanova street 8, <sup>2</sup>Institute of Information Technologies, Vladivostok State University of Economics and Service, 690014, Vladivostok, Gogolya street 41; <sup>3</sup>Regional Clinical Hospital No. 1, 690091, Vladivostok, Aleutskaya street 57.

**Purpose.** A comparative assessment of the prognostic accuracy of intrahospital mortality (IHM) models in patients with coronary heart disease (CHD) after coronary artery bypass grafting (CABG), developed on the basis of modern machine learning (MO) technologies.

**Materials and methods.** A retrospective analysis of 866 electronic case histories were made. All patients (685 men and 181 women) with CHD were underwent isolated CABG surgery from 2008 through 2018. We analyzed the results of clinical, laboratory and instrumental studies obtained before CABG. The total volume of the dataset consists of 82,000 records. The first group included 35 (4%) patients who died during the first 30 days after CABG, the second group - 831 (96%) with a favorable outcome of the operation. IHM predictors were distinguished by a multi-stage selection procedure with analysis of statistical hypotheses and calculation of weight coefficients. To build models and verify predictors, we used ML methods: multivariate logistic regression (MLR), random forest (RF), and artificial neural networks (ANNs). The accuracy of the models was evaluated by three metrics: the area under the ROC curve (AUC), sensitivity and specificity. Cross-validation of models were performed on test samples.

**Results.** Using the statistical analysis, the calculation of the odds and weight ratios we identify risk factors for IHM with the highest predictive potential. These included 7 factors from the EuroSCORE II scale: ejection fraction (EF) of less than 30%, EF of 30-50%, the age of patients, recent myocardial infarction, lesions of peripheral arterial pools, operation urgency, chronic heart failure III-IV functional classes and 5

additional predictors (heart rate, systolic blood pressure, presence of aortic stenosis, relative thickness index (RTI) of the posterior wall of the left ventricle (LV) and relative LV myocardial mass index (RLVMMI). The authors' models based on the MLR, RF and ANNs methods had higher AUC and sensitivity compared to the “classic” EuroSCORE II scale. The ANNs model with the inclusion of the predictors RTI and RLVMMI demonstrated the maximum level of prognostic accuracy, which was illustrated by the values of quality metrics: AUC - 95%, sensitivity - 92%, specificity - 98%.

**Conclusion.** Using modern ML technologies, it was possible to develop a new predictor selection algorithm and high-precision models for predicting IHL after CABG.

**Key words:** machine learning methods, prognostic models, coronary artery bypass grafting, in-hospital mortality.

**This work was supported by the Russian Foundation for Basic Research (projects № 18-29-03131 and 19-29-01-077).**

Ишемическая болезнь сердца (ИБС) - одна из основных причин инвалидизации и смертности населения в большинстве стран мира. Операции коронарного шунтирования (КШ) относятся к одним из наиболее распространенных видов восстановления коронарного кровотока, в связи с чем постоянно возрастает интерес к изучению роли факторов, влияющих на риск развития неблагоприятных исходов оперативного лечения. По данным Американской ассоциации сердца, внутригоспитальная летальность (ВГЛ) после изолированного КШ составляет 1-3% среди пациентов моложе 70 лет и 5-6% - среди лиц старше 70 лет. При этом уровень летальности при комбинированных кардиохирургических вмешательствах достигает 6% вне зависимости от возраста больных [1]. В последние годы в клинической кардиологии все чаще используются унифицированные инструменты прогнозирования неблагоприятных событий, основанные на результатах крупных проспективных исследований. К “классическим” моделям прогноза ВГЛ или летальности в течение 30 дней после КШ относят EuroSCORE II и STS Score (The Society of Thoracic Surgeons). Первая из них включает 18 предикторов, характеризующих клинико-функциональный статус больных до операции, виды и срочность кардиохирургических вмешательств. Класс убедительности и доказательности этой модели составляет IIa с уровнем достоверности B. В модели стратификации риска общества торакальных хирургов STS Score в модификации 2014 года используются 40 клинических и 2 ангиографических предиктора (класс доказательности IB). В рекомендациях 2018 года рабочей группы по реваскуляризации миокарда Европейского общества кардиологов (ESC) и Европейской ассоциации кардиоторакальных хирургов (EACTS) подтверждается, что шкалы EuroSCORE II и STS Score имеют сопоставимую прогностическую ценность в оценке риска ВГЛ после КШ. Обе прогностические системы демонстрировали приемлемую валидность при использовании в различных когортах больных, которым выполнялось КШ [2]. В ряде работ указывается на более высокую предсказательную способность шкалы EuroSCORE II и завышенную оценку рисков, рассчитанных по модели

STS Score [3,4]. Отсутствие “идеальных” прогностических технологий инициирует все большее количество исследований, в которых на основе современных методов машинного обучения (МО) реализуются подходы, направленные на совершенствование существующих и разработку новых моделей, стратифицирующих риск ВГЛ после КШ. Помимо логистической регрессии (ЛР), применяемой в “классических” шкалах, для построения новых моделей используются такие методы МО, как: дерево принятия решений, случайный лес (СЛ), машина опорных векторов, наивный байесовский классификатор, регрессия Кокса, искусственные нейронные сети (ИНС) и др. [5]. Повышение качества прогнозирования и управления рисками должно обеспечиваться за счет междисциплинарной кооперации клиницистов и специалистов в области обработки и анализа данных, результатом которой может быть создание информационных систем, поддерживающих принятие обоснованных врачебных решений.

**Цель исследования** состояла в сравнительной оценке точности моделей прогнозирования ВГЛ у больных ИБС после КШ, разработанных на основе современных технологий МО.

### **Материалы и методы**

Проведен ретроспективный анализ базы данных 866 электронных историй болезни (ЭИБ) пациентов (685 мужчин и 181 женщина) в возрасте от 35 до 81 года с медианой (Me) 63 года и 95% доверительным интервалом (ДИ) [63; 64], оперированных в 2008-2018 гг. по поводу ИБС методом изолированного КШ в отделении сердечно-сосудистой хирургии ГБУЗ “Приморская краевая клиническая больница №1” г. Владивостока. Показанием к операции служили стандартные клинические и ангиографические критерии. Все операции КШ были проведены в условиях искусственного кровообращения. Отсутствующие значения отдельных показателей дополнялись информацией, полученной из архива историй болезни на бумажных носителях. Записи, которые не удалось восполнить, исключали из



анализа. Таким образом, размер датасета по отдельным параметрам варьировал от 687 до 866 записей.

Среди обследованной когорты было выделено 2 группы лиц. В первую из них вошли 35 (4%) больных, умерших в течении первых 30 суток после КШ, во вторую - 831 (96%) больных с благоприятным исходом операции. Причиной смерти у 26 больных первой группы был интра- и послеоперационный инфаркт миокарда (ИМ), у 6-ти больных - послеоперационная острая почечная недостаточность. В остальных случаях летальные исходы были связаны с развитием панкреонекроза (1), субарахноидального кровоизлияния (1) и медиастинита (1).

На первом этапе исследования для выделения потенциальных предикторов, линейно связанных с ВГЛ, с помощью тестов Стьюдента, Манна-Уитни,  $\chi^2$  и Фишера был проведен сравнительный анализ 99 факторов, включающих результаты клинических, лабораторных и инструментальных исследований больных первой и второй групп до КШ. Для дальнейшего анализа отбирали показатели со статистически значимыми межгрупповыми различиями. На втором этапе исследования с помощью однофакторных моделей ЛР уточняли предиктивный потенциал выделенных факторов на основе расчета весовых коэффициентов, характеризующих степень влияния отдельных показателей на ВГЛ. Разработка указанных моделей проведена на нормализованных данных. На третьем этапе на исследуемой когорте анализировали прогностическую точность шкалы EuroSCORE II, используя три метрики: чувствительность и специфичность, которые рассчитывали по отношению к риску летальности  $> 5\%$ , а также площадь под ROC-кривой (AUC). В связи с тем, что среди анализируемого контингента отсутствовали пациенты, находящиеся до КШ на гемодиализе, в критическом состоянии, имеющие низкую мобильность, активный эндокардит или выраженную неврологическую дисфункцию, из дальнейшего исследования были исключены соответствующие предикторы шкалы EuroSCORE II. На данном этапе были разработаны авторские модели прогнозирования ВГЛ на основе

многофакторной ЛР с использованием предикторов шкалы EuroSCORE II и дополнительных факторов, выделенных на первом и втором этапах исследования. Процедура кросс-валидации моделей обеспечивалась многократным рандомизированным разделением когорты обследованных на обучающую и тестовую выборки в соотношении 75% и 25%, соответственно. При этом в тестовые выборки включалось одинаковое количество умерших: 8 случайно выбранных пациентов.

На четвертом этапе исследования разрабатывали модели ВГЛ на основе методов СЛ и ИНС. Повышение прогностической точности авторских моделей осуществляли за счет расширения спектра предикторов, не входящих в EuroSCORE II, а также за счет подбора параметров моделей. Основными параметрами разработанных моделей СЛ были 1000 деревьев для “голосования” и от 6 до 8 признаков для расщепления. Архитектура моделей ИНС включала 3 скрытых слоя с 7-10 нейронами в каждом, модели обучали на нормализованных данных. Основным условием включения новых предикторов в модели было улучшение качества последних. Из-за несбалансированности выборок решение об улучшении качества модели принимали на основании одновременного выполнения двух условий: улучшение по крайней мере одной из трех метрик качества и как минимум не снижения значения обобщенной метрики качества, полученной как среднее между чувствительностью и специфичностью. В нашем исследовании обобщенная метрика давала результат аналогичный коэффициенту корреляции Мэтьюса, часто используемому в задачах с несбалансированной выборкой [6].

В процессе статистической обработки данных у всех больных рассчитывали индекс массы тела (ИМТ), индекс коморбидности Чарльсона, а также эхокардиографические индикаторы гипертрофии левого желудочка (ЛЖ): индекс относительной толщины (ИОТ) задней стенки ЛЖ и индекс массы миокарда ЛЖ (ИММЛЖ). Для исключения влияния гендерного фактора ИММЛЖ нормировали на верхнюю границу его референсных значений,

ассоциированных с полом: 115 г/м<sup>2</sup> - для мужчин и 95 г/м<sup>2</sup> - для женщин с вычислением относительного ИММЛЖ (ОИММЛЖ).

Анализ данных и разработку моделей выполняли на языке R в среде R-studio ver. 1.0.153 и на языке Python ver.3.7.4. Для построения ИНС использовали библиотеки Keras и Tensorflow.

## Результаты

Статистический анализ факторов, характеризующих клиническо-функциональный статус больных ИБС первой и второй групп до КШ, показал, что большинство из них не имеют достоверных различий между собой (табл. 1). Так, из 18 непрерывных и категориальных переменных шкалы EuroSCORE II статистически значимые межгрупповые различия фиксировались только по 7 параметрам, к которым относились возраст, фракция выброса (ФВ) ЛЖ < 30%, ФВ 30%-50%, ранее перенесенный ИМ, атеросклеротическое поражение периферических артериальных бассейнов, экстренность КШ, хроническая сердечная недостаточность (ХСН) III-IV функциональных классов (ФК). В связи с этим, для построения прогностических моделей факторы риска ВГЛ, ранее апробированные в шкале EuroSCORE II, были дополнены новыми показателями, которые по результатам предварительной обработки данных, обладали наибольшим предиктивным потенциалом. К этим факторам относились: систолическое АД (САД), пульсовое АД (ПАД), индекс коморбидности Чарльсона и наличие аортального стеноза. Наибольшие значения ОШ, характеризующее степень влияния отдельных факторов на вероятность ВГЛ, имели место при ФВ < 30% (ОШ=11,8, 95% ДИ [1,4; 66,5]); ФВ 30%-50% (ОШ=3; 95% ДИ [1,5; 6,9]); недавно перенесенном ИМ (ОШ=5,6; 95% ДИ [2,8; 11,4]); экстренной КШ (ОШ=5,4, 95%ДИ [1,7; 14,3]); ХСН III-IV ФК (ОШ=5,1; 95% ДИ [2,5; 10,3]), наличии аортального стеноза (ОШ=4,5; 95%ДИ [1,44; 11,73]). Уровень ОШ других категориальных показателей был существенно ниже и статистически незначим. Например, коморбидность ИБС с хронической обструктивной болезнью легких (ХОБЛ) или с сахарным диабетом (СД) 2-го типа не увеличивала риск ВГЛ. В отличие от коморбидности ИБС с

аортальным стенозом, повышающей вероятность неблагоприятного исхода, ее сочетание с недостаточностью митрального, аортального или трикуспидального клапанов не влияло на ближайшие результаты КШ. На этом фоне индекс Чарльсона, указывающий на тяжесть сочетанной патологии, был значимо выше среди больных первой группы, что ассоциировалось с их более старшим возрастом, значения которого учитываются при расчете данного показателя. Статистический анализ показал также отсутствие межгрупповых различий по признакам ИМТ, статуса курения, ЧСС, диастолического АД (ДАД), гипертрофии миокарда ЛЖ, концентрации креатинина в сыворотке крови и его клиренса. При этом уровень САД и ПАД у лиц первой группы был достоверно ниже, что может объясняться большей распространенностью среди этих больных аортального стеноза (14,3% vs 3,6%) и меньшей долей АГ (80% vs 92,4%) по сравнению со второй группой.

На втором этапе исследования для верификации влияния отдельных факторов риска на ВГЛ нами были построены однофакторные модели ЛР с расчетом весовых коэффициентов, характеризующих предиктивную ценность анализируемых показателей (табл. 2). Данный подход существенно расширяет возможности для обработки и анализа информации за счет более детальной оценки степени влияния потенциальных предикторов на результирующую переменную.

В результате анализа было установлено, что максимальное значение весового коэффициента (2,44) ассоциируется с ФВ <30%. Сопоставимыми с ним по величине и уровню достоверности были такие переменные, как: недавно перенесенный ИМ, экстренность операции, ХСН III-IV ФК, ФВ 30-50%, наличие аортального стеноза. Менее заметное, но достоверное влияние на ближайшие результаты КШ оказывали: поражения периферических артериальных бассейнов, возраст больных, ЧСС, САД, ПАД, индекс коморбидности.

На третьем этапе исследования авторами были разработаны 2 варианта многофакторных моделей ЛР (табл. 3). В первой из них были использованы 7

предикторов шкалы EuroSCORE II, а во второй - они были дополнены 2 новыми факторами (ЧСС и аортальный стеноз), высокий прогностический потенциал которых был подтвержден в процессе валидации моделей. При построении последних было установлено, что предикторы в форме непрерывных переменных (возраст, креатинин сыворотки крови и ФВ) по отношению к их категориальным аналогам, которые используются в шкале EuroSCORE II, обладали лучшими прогностическими свойствами, что подтверждалось метриками качества. В то же время, такие предикторы EuroSCORE II, как: гендерная принадлежность, ХСН II ФК, стабильная стенокардия IV ФК, ХОБЛ, СД 2-го типа, СрДЛА в авторских многофакторных моделях были статистически незначимы. Напротив, фактор недавно перенесенного ИМ, предсказательная ценность которого в шкале EuroSCORE II была недостаточной ( $p > 0,05$ ), в авторской модели приобретала существенное прогностическое значение ( $p < 0,0001$ ). Это может объясняться большей долей больных с перенесенным ранее ИМ среди лиц первой группы по сравнению со второй (55,9% vs 17,4%). Необходимо также отметить, что в шкале EuroSCORE II использовалась комбинация линейно зависимых показателей возраста и клиренса креатинина, что создавало проблему мультиколлинеарности, ограничивающей ее прогностическую эффективность. Для решения данной проблемы в авторской модели ЛР применяли линейно независимые предикторы (возраст и концентрацию креатинина в сыворотке крови), что позволило повысить точность прогноза. Аналогичные действия, исключающие наличие в одной модели ЛР 2-х коррелирующих признаков, были выполнены и в отношении показателей САД и ПАД, которые были прямо взаимосвязаны с уровнем ЧСС, а также линейно-зависимых факторов возраста и индекса коморбидности. В разработанных моделях все весовые коэффициенты предикторов, за исключением ФВ, имели положительное значение, что указывало на увеличение риска ВГЛ при наличии этих признаков или повышении их уровня (табл.3). Так, например, увеличение возраста, концентрации креатинин в сыворотке крови или ЧСС ассоциируется с

возрастающим риском ВГЛ. Напротив, отрицательное значение весового коэффициента ФВ указывает на снижение вероятности неблагоприятного исхода КШ при увеличении данного фактора.

Сравнительный анализ индикаторов прогностической точности показал определенные различия моделей EuroSCORE II и авторских многофакторных моделей ЛР (табл.4). Было установлено, что качество прогноза по метрикам AUC, чувствительности и специфичности для “классической” шкалы EuroSCORE II с 18 предикторами составляло 0,73, 0,25 и 0,92, соответственно. Эти результаты свидетельствуют о ее недостаточной точности при апробации на анализируемой когорте и необходимости совершенствования. Авторская модель ЛР-I, построенная на 7 статистически значимых предикторах EuroSCORE II, обеспечила на тестовой выборке определенный рост показателей чувствительности (0,74) и AUC (0,83) на фоне более низкой специфичности (0,78). Модель ЛР-II с дополнительным параметром ЧСС и признаком аортального стеноза позволила увеличить показатели специфичности до 0,8 и AUC до 0,85. С целью получения более точного прогноза, нами были разработаны модели на основе других методов МО: СЛ и ИНС. Построение последних осуществлялось как с использованием предикторов EuroSCORE II, так и их комбинаций с дополнительными факторами. Модель СЛ-I показала худшую эффективность прогнозирования по отношению к ЛР-I и ЛР-II, но лучшую - по отношению к “классической” EuroSCORE II за счет более высокой чувствительности (0,69 vs 0,25). После включения дополнительных предикторов (ЧСС, САД и аортального стеноза) чувствительность модели СЛ-II возрастала до 0,82. Значительно более высокую точность показали модели ИНС, что иллюстрировалось лучшими метриками качества. При этом максимальный уровень прогноза точности достигалось в модели ИНС-III при включении в ее структуру предикторов ИОТ и ОИММЛЖ. В этих случаях показатель AUC составил 95%, чувствительности - 92%, а специфичности - 98%.

## **Обсуждение**

Методы МО относятся к основным инструментам искусственного интеллекта и все шире используются в различных областях клинической физиологии и медицины, в том числе в прогностических исследованиях [6]. Одним из основных преимуществ современных технологий МО по отношению к “классическим” методам статистики является их способность к обработке и анализу множества переменных, выявлению скрытых или неочевидных закономерностей, а также извлечению новых знаний (Data Mining). За последние годы интерес к этому направлению существенно возрос в связи с общим трендом развития цифровых технологий и их внедрением в различные проекты персонализированной медицины [8]. В настоящей работе прогнозирование вероятности ВГЛ после КШ базировалось на результатах анализа предоперационного статуса больных, что соответствует концепции скрининговой системы EuroSCORE II. Высокое качество предикторов ВГЛ в нашем исследовании обеспечивалось многоступенчатой процедурой отбора, которая помимо традиционных статистических методов, включала оценку их информативности с помощью весовых коэффициентов, полученных при построении однофакторных моделей ЛР. Использование данного алгоритма позволило ранжировать отдельные предикторы по степени влияния на ВГЛ. Из структуры многофакторных моделей ЛР исключались коррелирующие переменные в связи с проблемой мультиколлинеарности, ухудшающей их прогностический потенциал [9]. Выполнение этих условий позволило повысить предсказательную ценность данных моделей, что иллюстрировалось положительной динамикой метрик качества. Вместе с тем, известно, что мультиколлинеарность не является ограничением для моделей на основе СЛ и ИНС [10]. Это позволило при построении последних использовать более широкий спектр дополнительных предикторов и не разобщать их. Так, в построении модели ИНС-III помимо 7 факторов EuroSCORE II, ЧСС, САД, признака аортального стеноза участвовали индикаторы гипертрофии ЛЖ: ИОТ и ОИММЛЖ. Их включение в эту модель позволило максимально повысить уровень ее точности (AUC), специфичности и чувствительности.

Неслучайность комбинации данных предикторов в одной модели может объясняться закономерной патофизиологической взаимосвязью аортального стеноза с индикаторами гипертрофии ЛЖ и более низким уровнем САД, что было характерным для больных с неблагоприятным исходом КШ. В нашем исследовании высокая точность авторских моделей ЛР, СЛ и ИНС подтверждалась также многоступенчатой процедурой валидации на тестовых выборках, что демонстрировалось возрастающими значениями метрик качества. При этом система оценки рисков по шкале EuroSCORE II показала менее точные результаты прогноза по сравнению с авторскими моделями, что было связано прежде всего с ее низкой чувствительностью. По мнению ряда авторов, наилучшую предсказательную способность модели чаще всего демонстрируют в тех популяциях, откуда были получены исходные данные [11]. Меньшая эффективность прогнозирования на других популяциях может объясняться региональными особенностями ресурсного обеспечения здравоохранения и другими факторами. Именно поэтому наличие и постоянная актуализация региональных регистров больных ИБС с применением современных методов МО для обработки и анализа больших данных будет способствовать получению более точной прогностической информации и повышению качества оказания специализированной медицинской помощи.

### **Заключение**

Разработанный алгоритм обработки и анализа больших объемов данных, характеризующих клинико-функциональный статус больных ИБС до КШ обеспечил высокое качество отбора предикторов ВГЛ. Прогностические модели на основе методов ЛР, СЛ и ИНС показали преимущества в точности расчета риска неблагоприятного исхода КШ по сравнению с “классической” шкалой EuroSCORE II. К универсальным предикторам, определяющим высокую точность прогнозирования, помимо 7 факторов EuroSCORE II, относились показатели ЧСС, САД и наличие аортального стеноза. Модель ИНС-III продемонстрировала наибольшую предсказательную ценность за счет предикторов ИОТ и ОИММЛЖ.



### **Финансирование**

Исследование выполнено при финансовой поддержке РФФИ в рамках научных проектов 18-29-03131 и 19-29-01077.

### **Конфликт интересов**

Конфликт интересов у авторов отсутствует.

Таблица 1.

Клинико-функциональная характеристика больных ИБС до КШ по предикторам EuroSCORE II с дополнениями

| <i>Показатели</i>                                | <i>Объем выборки</i> | <i>1-ая группа, n=35</i> | <i>2-ая группа, n=831</i> | <i>ОШ, (95% ДИ)</i> | <i>p-value</i> |
|--|----------------------|--------------------------|---------------------------|---------------------|----------------|
| Всего пациентов, <i>n</i> (%)                    | 866                  | 35 (4%)                  | 831 (96%)                 | -                   | -              |
| Женщины, <i>n</i> (%)                            | 866                  | 10 (28,6%)               | 171 (20,6%)               | 1,56 [0,7; 3,2]     | 0,35           |
| Мужчины, <i>n</i> (%)                            | 866                  | 25 (71,4%)               | 660 (79,4%)               | 0,64 [0,31; 1,44]   | 0,36           |
| Возраст, лет, <i>M±SD</i>                        | 866                  | 66,9±5,5                 | 62,9±7,5                  | -                   | 0,001          |
| Клиренс креатинина ≤50 мл/мин, <i>n</i> (%)      | 687                  | 7 (20%)                  | 65 (10%)                  | 2,3 [0,88; 5,2]     | 0,109          |
| Клиренс креатинина 50-85 мл/мин, <i>n</i> (%)    | 687                  | 21 (60%)                 | 373 (57,2%)               | 1,11 [0,53; 2,4]    | 0,97           |
| Заболевания периферических артерий, <i>n</i> (%) | 840                  | 18 (51,4%)               | 238 (29,6%)               | 2,6 [1,3; 5,4]      | 0,0066         |
| Операция на сердце в анамнезе, <i>n</i> (%)      | 840                  | 5 (14,3%)                | 75 (9,3%)                 | 1,66[0,54; 4,1]     | 0,49           |
| Недавно перенесенный ИМ, <i>n</i> (%)            | 839                  | 19 (55,9%)               | 140 (17,4%)               | 5,6 [2,8; 11,4]     | <0,0001        |
| ФВ от 30-50%, <i>n</i> (%)                       | 797                  | 16 (45,7%)               | 168 (22%)                 | 3 [1,5; 6,9]        | 0,0023         |
| ФВ < 30%, <i>n</i> (%)                           | 797                  | 2 (5,7%)                 | 4 (0,52%)                 | 11,8 [1,4; 66,5]    | 0,013          |
| ХСН III-IV ФК, <i>n</i> (%)                      | 854                  | 17 (48,6%)               | 130 (15,6%)               | 5,1 [2,5; 10,3]     | <0,0001        |
| Нестабильная стенокардия, <i>n</i> (%)           | 862                  | 16 (45,7%)               | 460 (55,6%)               | 0,67 [0,33; 1,33]   | 0,33           |
| Стабильная стенокардия IV ФК, <i>n</i> (%)       | 862                  | 3 (8,6%)                 | 30 (3,6%)                 | 2,6[0,56; 7,9]      | 0,3            |
| СрДЛА 31-55 мм рт.ст., <i>n</i> (%)              | 778                  | 7 (20%)                  | 111 (14,9%)               | 1,4[0,56; 3,2]      | 0,56           |

|  |     |                  |                  |                   |        |
|--|-----|------------------|------------------|-------------------|--------|
| СрДЛА $\geq$ 55 мм рт.ст., n (%)                 | 778 | 1 (2,8%)         | 13 (1,7%)        | 1,8 [0,07; 9,8]   | 1      |
| Экстренная операция, n (%)                       | 861 | 5 (14,3%)        | 25 (3%)          | 5,4 [1,7; 14,3]   | 0,002  |
| СД, n (%)  | 866 | 7 (20%)          | 28 (23,6%)       | 0,82 [0,32; 1,8]  | 0,77   |
| ХОБЛ, n (%)                                      | 866 | 6 (17,1%)        | 90 (10,8%)       | 1,7 [0,63; 4]     | 0,37   |
| Креатинин сыворотки крови*, мкмоль/л, $M \pm SD$ | 768 | 111,13 $\pm$ 30  | 103,5 $\pm$ 21   | -                 | 0,15   |
| ФВ крови*, %, $M \pm SD$                         | 797 | 50,2 $\pm$ 11,8  | 57,7 $\pm$ 9,6   | -                 | 0,0007 |
| Индекс коморбидности*                            | 864 | 5,5 $\pm$ 1,61   | 4,7 $\pm$ 1,66   | -                 | 0,0049 |
| ИМТ*, $M \pm SD$                                 | 771 | 27,6 $\pm$ 6,2   | 28,8 $\pm$ 5,2   | -                 | 0,28   |
| Статус курения*, n (%)                           | 859 | 12 (34,3%)       | 262 (31,5%)      | 1,1 [0,54; 2,3]   | 0,87   |
| ЧСС*, $M \pm SD$                                 | 866 | 74,5 $\pm$ 16,6  | 70 $\pm$ 10      | -                 | 0,12   |
| САД*, мм рт.ст., $M \pm SD$                      | 866 | 125,7 $\pm$ 17,6 | 133,2 $\pm$ 18,2 | -                 | 0,018  |
| ДАД*, мм рт.ст., $M \pm SD$                      | 866 | 76,7 $\pm$ 9,7   | 79,2 $\pm$ 8,1   | -                 | 0,15   |
| ПАД*, мм рт.ст., $M \pm SD$                      | 866 | 48,9 $\pm$ 10,3  | 54 $\pm$ 13,9    | -                 | 0,0078 |
| Наличие АГ*, n (%)                               | 861 | 28 (80%)         | 768 (92,4%)      | 0,3 [0,13; 0,78]  | 0,011  |
| ИОТ задней стенки ЛЖ*, $M \pm SD$                | 786 | 0,39 $\pm$ 0,1   | 0,42 $\pm$ 0,09  | -                 | 0,123  |
| ОИММЛЖ*, $M \pm SD$                              | 694 | 1,12 $\pm$ 0,346 | 1,1 $\pm$ 0,9    | -                 | 0,8    |
| Аортальный стеноз*, n (%)                        | 866 | 5 (14,3%)        | 30 (3,6%)        | 4,5 [1,44; 11,73] | 0,0069 |

Примечание: В таблицу включены предикторы EuroSCORE II и дополнительные факторы с наибольшим предиктивным потенциалом. ОШ рассчитывали только для категориальных переменных. \* - дополнительные факторы. ФВ - фракция выброса; ХСН - хроническая сердечная недостаточность; ФК - функциональный класс; СрДЛА - среднее давление в легочной артерии; ХОБЛ - хроническая обструктивная болезнь легких; ИМТ - индекс массы тела; ЧСС - частота сердечных сокращений; САД - систолическое артериальное давление; ДАД - диастолическое артериальное давление; ПАД - пульсовое артериальное давление; АГ - артериальная гипертензия; ЛЖ - левый желудочек; ИОТ - индекс относительной толщины задней стенки левого желудочка; ИММЛЖ - индекс массы миокарда левого желудочка

Таблица 2

Весовые коэффициенты однофакторных моделей логистической регрессии  
для оценки риска внутригоспитальной летальности

| Показатели                                   | Коэффициент | Стандартная ошибка | <i>p-value</i> |
|--|-------------|--------------------|----------------|
| Возраст (непрерывный)                        | 0,0788      | 0,0257             | 0,0021         |
| Возраст (непрерывный при > 60 лет)           | 0,0888      | 0,0302             | 0,0033         |
| Женщины                                      | 0,4343      | 0,3839             | 0,25793        |
| Клиренс креатинина ≤50 мл/мин                | 0,8143      | 0,4423             | 0,0656         |
| Клиренс креатинина 50-85 мл/мин              | 0,1151      | 0,5398             | 0,7451         |
| Поражение периферических артерий             | 0,9253      | 0,3469             | 0,0077         |
| Операция на сердце в анамнезе                | 0,4838      | 0,4980             | 0,3313         |
| Ранее перенесенный ИМ                        | 1,7285      | 0,3518             | 0,000009       |
| ФВ от 30-50%                                 | 1,0851      | 0,3503             | 0,0020         |
| ФВ < 30%                                     | 2,4423      | 0,8841             | 0,0057         |
| ХСН III-IV ФК                                | 1,6105      | 0,3515             | 0,000005       |
| Нестабильная стенокардия                     | -0,3977     | 0,3465             | 0,2510         |
| Стабильная стенокардия IV ФК                 | 0,9125      | 0,318              | 0,1486         |
| СрДЛА, мм рт.ст.                             | 0,0063      | 0,0209             | 0,7645         |
| СрДЛА 31-55 мм рт.ст.                        | 0,3546      | 0,4349             | 0,4148         |
| СрДЛА ≥ 55 мм рт.ст.                         | 0,5031      | 1,0525             | 0,6326         |
| Экстренная операция                          | 1,6752      | 0,5240             | 0,0014         |
| Другие операции, исключая операции на сердце | -12,4024    | 840,2741           | 0,9882         |
| Сахарный диабет 2го типа                     | -0,2108     | 0,4304             | 0,6243         |
| ХОБЛ   | 0,5327      | 0,4622             | 0,2491         |
| Креатинин*, ммоль/мл                         | 0,01331     | 0,0065             | 0,0413         |
| Индекс коморбидности*                        | 0,2722      | 0,0948             | 0,0041         |
| ИМТ*   | -0,0456     | 0,0355             | 0,1993         |
| Статус курильщика*                           | 0,0043      | 0,0097             | 0,6603         |
| ЧСС *  | 0,0276      | 0,0116             | 0,0172         |
| САД*   | -0,0294     | 0,0122             | 0,0160         |

|                    |         |        |        |
|--------------------|---------|--------|--------|
| ДАД,*              | -0,0386 | 0,0220 | 0,0796 |
| ПАД*               | -0,0348 | 0,0163 | 0,0331 |
| ИОТ*               | -2,2966 | 2,2428 | 0,3058 |
| ОИММЛЖ*            | 0,0194  | 0,1734 | 0,9109 |
| Аортальный стеноз* | 1,4929  | 0,5176 | 0,0039 |

Примечание: \* - дополнительные факторы; ФВ - фракция выброса; ХСН - хроническая сердечная недостаточность; ФК - функциональный класс; СрДЛА - среднее давление в легочной артерии; ХОБЛ - хроническая обструктивная болезнь легких; ИМТ - индекс массы тела; ЧСС - частота сердечных сокращений; САД - систолическое артериальное давление; ДАД - диастолическое артериальное давление; ПАД - пульсовое артериальное давление; ИММЛЖ - индекс массы миокарда левого желудочка

Таблица 3.

**Весовые коэффициенты предикторов многофакторных моделей ЛР для прогнозирования внутригоспитальной летальности после КШ**

| Предикторы                         | Авторская модель ЛР с предикторами EuroSCORE II |          | Авторская модель ЛР с предикторами EuroSCORE II и дополнениями |         |
|------------------------------------|---|----------|--|---------|
|                                    | Коэффициент и 95% ДИ                            | p-value  | Коэффициент и 95% ДИ   | p-value |
| Возраст                            | 0,056 [0,0024; 0,11]                            | 0,04     | 0,057  | 0,04    |
| Креатинин сыворотки крови          | 0,015 [0; 0,03]                                 | 0,05     | 0,015  | 0,07    |
| Заболевания периферических артерий | 0,89 [0,12; 1,66]                               | 0,023    | 1  | 0,014   |
| Недавно перенесенный ИМ            | 1,76 [0,98; 2,53]                               | <0,0001  | 1,92   | <0,0001 |
| ФВ                                 | -0,042 [-0,078; -0,0065]                        | 0,02     | -0,014   | 0,026   |
| Экстренная операция                | 1,71 [0,38; 3,04]                               | 0,011    | 1,78   | 0,011   |
| ХСН III-IV ФК                      | 1,69 [0,89; 2,53]                               | <0,0001  | 1,73   | <0,0001 |
| ЧСС*                               | -   | -        | 0,032  | 0,021   |
| Аортальный стеноз*                 | -   | -        | 1,75   | 0,0074  |
| Константа                          | -7,5 [-11,6; -3,31]                             | 0,000045 | -10,17   | 0,00007 |

Примечание: показатели креатинина, ФВ и ЧСС представлены в форме непрерывных факторов; \* - дополнительные факторы; ФВ - фракция выброса; ХСН - хроническая сердечная недостаточность; ФК - функциональный класс; ЧСС - частота сердечных сокращений

Таблица 4

Оценка точности авторских моделей для прогнозирования внутригоспитальной летальности после КШ

| Метод                 | Предикторы  | AUC  | Чувствительность | Специфичность |
|-----------------------|---|------|------------------|---------------|
| ЛР<br>EuroSCORE<br>II | EuroSCORE II (18)   | 0,75 | 0,25             | 0,92          |
| ЛР-I                  | EuroSCORE II (7)  | 0,83 | 0,74             | 0,78          |
| ЛР-II                 | EuroSCORE II (7) + ЧСС<br>+ Аортальный стеноз                       | 0,85 | 0,7              | 0,8           |
| СЛ-I                  | EuroSCORE II (7<br>предикторов)                                     | 0,71 | 0,69             | 0,71          |
| СЛ-II                 | EuroSCORE II (7) + ЧСС<br>+ Аортальный стеноз +<br>САД              | 0,78 | 0,82             | 0,74          |
| ИНС-I                 | EuroSCORE II (7)  | 0,85 | 0,798            | 0,895         |
| ИНС-II                | EuroSCORE II (7) + ЧСС<br>+ Аортальный стеноз +<br>САД              | 0,9  | 0,86             | 0,935         |
| ИНС-III               | EuroSCORE II (7) + ЧСС<br>+ Аортальный стеноз +<br>САД+ИОТ + ОИММЛЖ | 0,95 | 0,92             | 0,98          |

Примечание: ЧСС - частота сердечных сокращений; САД - систолическое артериальное давление; ИОТ - индекс относительной толщины задней стенки левого желудочка; ОИММЛЖ - относительный индекс массы миокарда левого желудочка. В скобках указано количество предикторов шкалы EuroSCORE II.

## Литература

1. Bockeria L.A., Sokolskaya N.O., Kopylova N.S., Alshibaya M.M. Echocardiographic predictors of the severity of the early postoperative period in patients after surgical myocardial revascularization. *Anesteziologiya I reanimatologiya*. 2015; 60 (5): 8-11 (In Russ.)
2. Ad N, Holmes SD, Patel J, et al. Comparison of EuroSCORE II, Original EuroSCORE, and The Society of Thoracic Surgeons Risk Score in Cardiac Surgery Patients. *Ann Thorac Surg*. 2016; 102 (2): 573-79. DOI:10.1016/j.athoracsur.2016.01.105
3. Yamaoka H., Kuwaki K., Inaba H., Yamamoto T., Kato T.S., Dohi S. et al. Comparison of modern risk scores in predicting operative mortality for patients undergoing aortic valve replacement for aortic stenosis. *J. Cardiol*. 2015; PII: S0914-5087(15) 00282-8. DOI: 10.1016/j.jjcc.2015.08.017.
4. Sedaghat A., Sinning J.-M., Vasa-Nicotera M., Ghanem A., Hammerstingl C., Grube E. et al. The revised EuroSCORE II for the prediction of mortality in patients undergoing transcatheter aortic valve implantation. *Clin. Res. Cardiol*. 2013; 102 (11): 821–829.
5. Kurt I, Ture M, Kurum AT. Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease. *Expert Systems with Applications*. 2008; 34(1): 366-374. DOI: 10.1016/j.eswa.2006.09.004
6. Matthews B. Comparison of the predicted and observed secondary structure of T4 phagelysozyme. *Biochimica et biophysica acta*.1975; 405(2):442-451. DOI: 10.1016/0005-2795(75)90109-9
7. Steele AJ, Denaxas SC, Shah AD, Hemingway H, Luscombe NM (2018) Machine learning models in electronic health records can outperform conventional survival models for predicting patient mortality in coronary artery disease. *PLoS ONE*/ 2018; 13(8): e0202344. DOI:10.1371/journal.pone.0202344



8. Al'Aref S.J. и др. Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. *European Heart Journal*. 2018; 40(24): 1975–1986.

9. Ranganathan P., Pramesh C.S., Aggarwal R. Common pitfalls in statistical analysis: Logistic regression. *Perspect Clin Res*. 2017; 8(3): 148–151. DOI: 10.4103/picr.PICR\_87\_17

10. Flach P. *Machine learning: the art and science of algorithms that make sense of data*. Cambridge University Press, 2017.

11. Boytsov S.A. и др. Progress and Problems of Practical Cardiology in Russia at the Present Stage. *Kardiologiya*. 2019; 59 (3): 53–59. (In Russ.) DOI: 10.18087/cardio.2019.3.10242