

ЕРМАК А.Д.,

ООО «К-Скай», г. Петрозаводск, Россия; e-mail: andrewermakwork@gmail.com

ГАВРИЛОВ Д.В.,

ООО «К-Скай», г. Петрозаводск, Россия; e-mail: dgavrilov@webiomed.ru

КУЗНЕЦОВА Т.Ю.,

д.м.н., доцент, ФГБОУ ВО Петрозаводский государственный университет, г. Петрозаводск, Россия; e-mail: eme@sampo.ru

АНДРЕЙЧЕНКО А.Е.,

к.т.н., Университет ИТМО, г. Санкт-Петербург, Россия; e-mail: anna.ev.andreychenko@yandex.ru

МАКАРОВА Е.А.,

к.т.н., ООО «К-Скай», г. Петрозаводск, Россия; e-mail: emakarova@webiomed.ru

НОВИЦКИЙ Р.Э.,

ООО «К-Скай», г. Петрозаводск, Россия; e-mail: roman@webiomed.ru

ГУСЕВ А.В.,

к.т.н., ФГБУ «Центральный научно-исследовательский институт организации и информатизации здравоохранения» Минздрава России, г. Москва, Россия; e-mail: agusev@webiomed.ru

СРАВНЕНИЕ ТОЧНОСТИ ПРОГНОЗИРОВАНИЯ ФАТАЛЬНЫХ ОСЛОЖНЕНИЙ СЕРДЕЧНО-СОСУДИСТЫХ ЗАБОЛЕВАНИЙ С ИСПОЛЬЗОВАНИЕМ ШКАЛЫ SCORE И МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ

DOI: 10. 25881/18110193_2025_4_86

Аннотация. Выявление пациентов высокого риска фатальных осложнений сердечно-сосудистых заболеваний (ССЗ) является одной из важнейших задач для сокращения предотвратимой заболеваемости и смертности от ССЗ. Для этого широко применяются различные алгоритмы оценки риска и шкалы, недостатком которых является использование ограниченного набора предикторов и невысокая точность. Методы машинного обучения (МО) открывают возможности для устранения данных недостатков и персонализации оценки сердечно-сосудистого риска.

Цель исследования: сравнить точность шкалы SCORE и моделей МО в задаче прогнозирования фатальных осложнений ССЗ.

Материалы и методы. Проведено многоцентровое ретроспективное исследование (1999–2018 гг.). Включен 3891 случай лечения 1064 пациентов Российской Федерации 40–69 лет. Для прогнозирования применялись логистическая регрессия, ансамблевые методы МО и Multi-layer Perceptron. Сравнение с SCORE проводилось на независимом валидационном наборе, включавшем 440 записей.

Результаты. Лучшие показатели точности продемонстрировала модель МО CatBoost (AUROC 0,879; чувствительность 0,938; специфичность 0,777). На валидации CatBoost показал сопоставимую со SCORE дискриминацию, но превосходил шкалу по специфичности (0,653 против 0,408) и точности (0,673 против 0,45) при разделении пациентов на группы низкого и умеренного риска. Ключевыми предикторами для модели были пол, возраст, курение, систолическое давление, индекс массы тела, ЧСС и липидный профиль.

Заключение. Модель МО превзошла шкалу SCORE по точности прогнозирования фатальных сердечно-сосудистых событий. Использование МО в прогнозировании сердечно-сосудистого риска позволяет повысить эффективность профилактики ССЗ и способствует персонализированному ведению пациентов.

Ключевые слова: сердечно-сосудистый риск, SCORE, машинное обучение, CatBoost, профилактика.

Для цитирования: Ермак А.Д., Гаврилов Д.В., Кузнецова Т.Ю., Андрейченко А.Е., Макарова Е.А., Новицкий Р.Э., Гусев А.В. Сравнение точности прогнозирования фатальных осложнений сердечно-сосудистых заболеваний с использованием шкалы SCORE и модели машинного обучения. Врач и информационные технологии. 2025; 4: 86-98. DOI: 10.25881/18110193_2025_4_86.

ERMAK A.D.,

K-SkAI LLC, Petrozavodsk, Russia; e-mail: andrewermakwork@gmail.com

GAVRILOV D.V.,

K-SkAI LLC, Petrozavodsk, Russia; e-mail: dgavrilov@webiomed.ru

KUZNETSOVA T.YU.,

DSc, Associate Professor, Petrozavodsk State University, Petrozavodsk, Russia; e-mail: eme@sampo.ru

ANDREICHENKO A.E.,

PhD, ITMO University, Saint Petersburg, Russia; e-mail: anna.ev.andreychenko@yandex.ru

MAKAROVA E.A.,

PhD, K-SkAI LLC, Petrozavodsk, Russia; e-mail: anna.ev.andreychenko@yandex.ru

NOVITSKIY R.E.,

K-SkAI LLC, Petrozavodsk, Russia; e-mail: roman@webiomed.ru

GUSEV A.V.,

PhD, Federal Research Institute for Health Organization and Informatics, Moscow, Russia;
e-mail: agusev@webiomed.ru

COMPARISON OF PROGNOSTIC ACCURACY OF SCORE SCALE AND A MACHINE LEARNING MODEL IN PREDICTING FATAL CARDIOVASCULAR COMPLICATIONS

DOI: 10. 25881/18110193_2025_4_86

Abstract. Identifying patients at high risk for fatal cardiovascular disease (CVD) complications is a critical task in reducing preventable CVD morbidity and mortality. Various risk assessment algorithms and scores are widely used for this purpose, but their limitations include a limited set of predictors and low accuracy. Machine learning methods offer the potential to address these shortcomings and personalize cardiovascular risk assessment.

Objective: To compare the accuracy of the SCORE scale and machine learning models in predicting fatal cardiovascular complications.

Materials and Methods: A multicenter retrospective study was conducted (1999–2018), including 3,891 treatment cases of 1,064 patients aged 40–69 years in the Russian Federation. Logistic regression, ensemble machine learning (ML) methods, and Multi-Layer Perceptron were used for forecasting. Comparison with SCORE was performed on an independent validation set consisting of 440 records.

Results: The CatBoost ML model demonstrated the best accuracy (AUROC 0.879; sensitivity 0.938; specificity 0.777). During validation, CatBoost demonstrated comparable discrimination to SCORE but outperformed the scale in specificity (0.653 vs. 0.408) and accuracy (0.673 vs. 0.45) when referencing patients to low- and intermediate-risk groups. Key predictors for the model were gender, age, smoking, systolic blood pressure, body mass index, heart rate, and lipid profile.

Conclusion: The machine learning model outperformed the SCORE scale in predicting fatal cardiovascular events. The use of machine learning in predicting cardiovascular risk can improve the effectiveness of CVD prevention and facilitate personalized patient care.

Keywords: cardiovascular risk, SCORE, machine learning, CatBoost, prevention.

For citation: Ermak A.D., Gavrilov D.V., Kuznetsova T.Yu., Andreichenko A.E., Makarova E.A., Novitskiy R.E., Gusev A.V. Comparison of prognostic accuracy of SCORE scale and a machine learning model in predicting fatal cardiovascular complications. Medical doctor and information technology. 2025; 4: 86-98. DOI: 10.25881/18110193_2025_4_86.

ВВЕДЕНИЕ

Сердечно-сосудистые заболевания (ССЗ) являются лидирующей причиной заболеваемости и смертности во всем мире [1]. Одним из эффективных инструментов сокращения бремени ССЗ является оценка сердечно-сосудистого риска (CCP). Как правило, данная задача выполняется с помощью разработанных шкал-рискометров [2]. В литературе представлено достаточно много данных о том, что существующие на сегодня инструменты оценки CCP имеют существенные недостатки: количество шкал велико, каждая отличается по количеству включаемых в нее факторов риска (ФР) и точностью прогноза. Кроме того, использование шкал в реальной клинической практике сопряжено с большим количеством ошибок или в целом отказом от их применения в силу загруженности врачей, необходимости оценки большого количества информации в электронной медицинской карте (ЭМК), большой вероятности пропуска важной информации, что в целом негативно влияет на правильность оценки CCP [3]. Развитие технологий искусственного интеллекта (ИИ) позволяет предложить практическому здравоохранению более эффективные пути решения этой задачи.

Согласно действующим в России клиническим рекомендациям, CCP может быть оценен, в основном, по шкалам SCORE (оценка фатальных сердечно-сосудистых событий – CCC) и SCORE2 (оценка фатальных и нефатальных CCC) [4]. ФР, включенные в эти шкалы: пол, возраст, курение, систолическое артериальное давление и показатель липидного спектра (общий холестерин и холестерин нелВП соответственно по шкалам). Важно отметить, что перечень известных ФР ССЗ гораздо шире, и особого внимания требуют модифицируемые ФР, помимо выше указанных, это и избыточная масса тела, и гипергликемия, коррекция которых вносит значительный вклад в CCP [5].

Целью данного исследования было сравнение точности шкалы SCORE и моделей машинного обучения (МО) в задаче прогнозирования фатальных осложнений ССЗ.

МАТЕРИАЛЫ И МЕТОДЫ

Дизайн исследования

Проведено многоцентровое ретроспективное обсервационное исследование, состоявшее

из сбора и обработки данных, разработки моделей МО, валидации наилучшей модели и сравнения ее метрик качества и дискриминативной способности с аналогичными показателями шкалы SCORE.

Формирование наборов данных для исследования

В качестве источника для исследования использовалась информация о пациентах из одного региона Российской Федерации, полученная из базы данных платформы Webiomed (<https://webiomed.ru/>), где она представлена в виде отдельных записей о случаях оказания медицинской помощи пациентам. Каждому такому случаю лечения в базе данных соответствуют даты его открытия и закрытия, а также информация о клинических параметрах пациента в период этого случая.

В качестве первоначальных критериев включения в исследование использовали факт наличия в ЭМК пациента в период случая лечения извлеченного значения общего холестерина, а также отсутствие на момент открытия этого случая в перечне зарегистрированных окончательных диагнозов в ЭМК кодов МКБ-10, относящихся к сахарному диабету, ишемической болезни сердца (ИБС), геморрагическому или ишемическому инсульту [6]. Всего после применения этих критериев было отобрано 105 283 случаев лечения для 25 805 уникальных пациентов (рис. 1).

В качестве целевого события выступал факт смерти от ИБС и цереброваскулярных заболеваний в течение 10 лет после анализируемого случая лечения. Факт летального исхода фиксировался при наличии в ЭМК свидетельства о смерти с соответствующим кодом причины смерти. В процессе сбора данных случаи лечения, после которых пациенты умирали в течение десяти лет от других причин, а также случаи, после которых отсутствовала какая-либо информация о пациенте после 10 лет наблюдения, исключались из набора данных. Коды МКБ-10, использовавшиеся для определения наличия (класс 1) или отсутствия (класс 0) целевого события, представлены в табл. 1. После исключения случаев лечения, после которых нельзя было достоверно установить дальнейшую судьбу пациентов в течение последующих 10 лет, для проведения исследования было оставлено 3 891

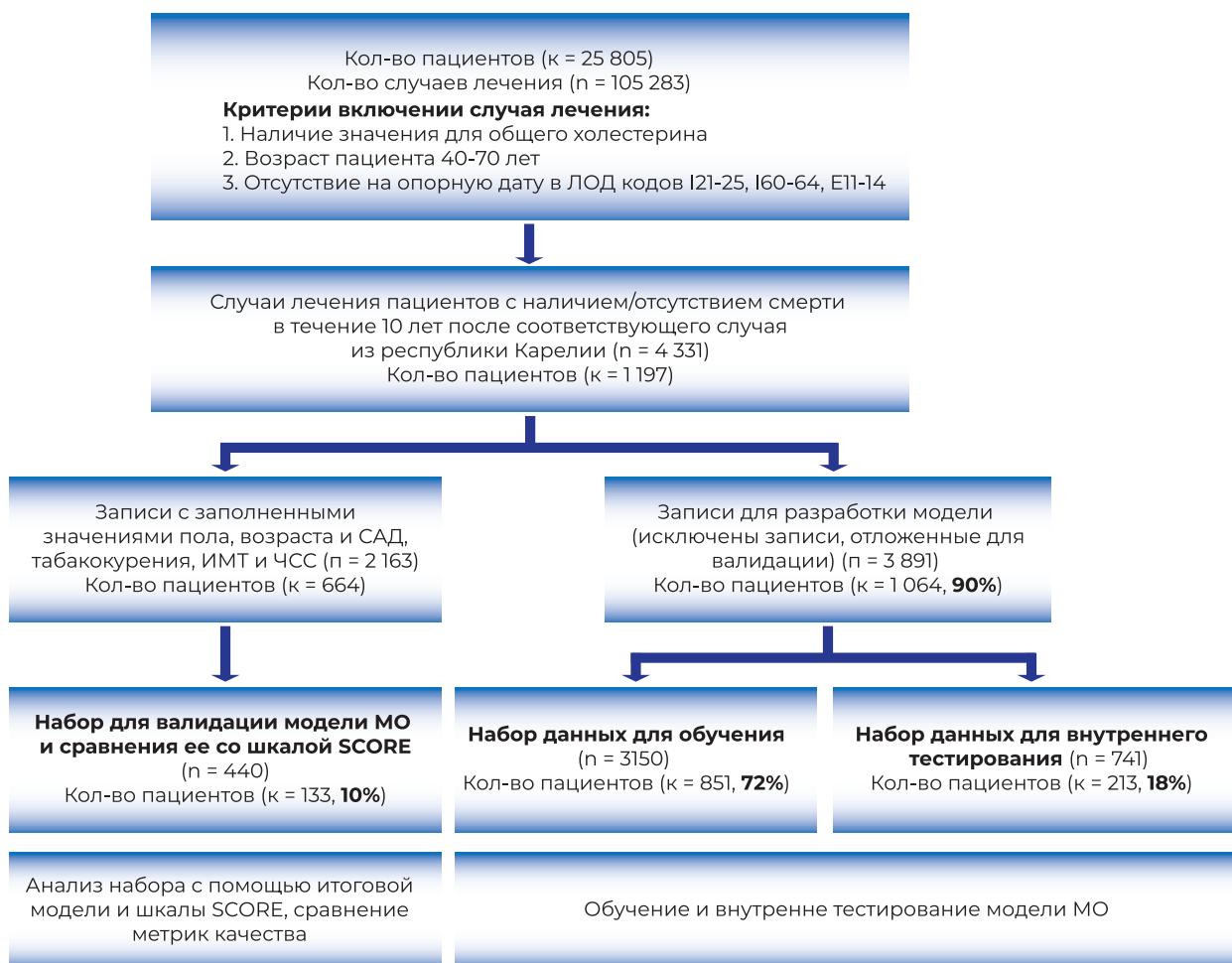


Рисунок 1 — Дизайн исследования. Примечание: n – число записей/случаев лечения у пациента; к – число пациентов.

Таблица 1 — Группы кодов МКБ-10, наличие которых в ЭМК учитывалось на этапе отнесения записи пациента к группе с наличием целевого события или без него

Группа	Коды МКБ
Ишемическая болезнь сердца**	I21*, I22*, I23*, I24*
Геморрагический и ишемический инсульт**	I60*, I61*, I62*, I63*, I64*
Сахарный диабет#	E10*, E11*, E12*, E13*, E14*

Примечание: # – группы заболеваний, использовавшиеся в качестве критериев исключения из исследования (факт отсутствия заболевания на момент открытия случая лечения); * – группы заболеваний, использовавшиеся для отнесения случая к группе с целевым событием; * – включая все подкоды.

случаев лечения от 1 064 уникальных пациентов в возрасте от 40 до 69 лет, полученных в период с 1999 по 2018 гг.

В качестве предикторов фатальных осложнений ССЗ в течение 10 лет был использован ряд признаков, признанных наиболее значимыми

ФР и использующихся в существующих прогнозных шкалах [6, 7]. В каждом случае извлекались значения для отобранных признаков: систолическое артериальное давление (САД), частота сердечных сокращений (ЧСС). Индекс массы тела (ИМТ) рассчитывался на основании роста и

веса. Табакокурение извлекалось по следующему правилу: значение True определялось если факт курения пациента упоминался хотя бы раз за всю историю ЭМК. Возраст определялся именно на момент открытия соответствующего случая лечения, а пол был априори определен у всех пациентов в используемой нами базе данных.

Из полученного набора отобрали отдельную группу случаев лечения с обязательно заполненными значениями возраста, пола, САД, табакокурения, ИМТ и ЧСС. Из этой группы случайным образом выбрали записи такого числа пациентов, чтобы их отношение к числу пациентов во всем наборе равнялось 10%. Таким образом, был сформирован набор для валидации модели МО и сравнения ее со шкалой SCORE, в состав которого вошло всего 440 записей 133 уникальных пациентов. Оставшиеся записи случайным образом делили на набор данных для обучения моделей МО (80%) и набор данных для их внутреннего тестирования (20%). Разделение проводили таким образом, чтобы в каждой из полученных выборок сохранялся изначальный баланс классов, а также, чтобы все записи, относящиеся к уникальному пациенту, находились только в одной выборке, что позволяло избежать утечки информации между выборками [8].

Разработка моделей

Статистический анализ и построение моделей МО выполняли на языке программирования Python, версия 3.9. При обработке данных применялось два различных подхода к пропущенным значениям в количественных признаках: использование фиксированного значения «-10000» и отсутствие заполнения пропусков [9]. Заполнение пропущенных значений в бинарных признаках во всех случаях проводилось с помощью нулевых значений (отрицаний наличия признака). На этапе масштабирования признаков также применялось несколько способов: трансформация с помощью гистограммной нормализации или стандартизации, а также использование данных в изначальной размерности [10,11]. Для исследования целесообразности коррекции дисбаланса классов использовались следующие алгоритмы: Random Undersampling, Random Oversampling и Synthetic Minority Oversampling Technique [12].

В качестве алгоритмов МО применяли логистическую регрессию (LR), методы, основанные на деревьях решений с использованием градиентного бустинга (AdaBoost, LightGBM, XGBoost, CatBoost) и бэггинга (RandomForest и ExtraTrees) и алгоритм на основе нейронных сетей (Multi-layer Perceptron).

Определение оптимальных гиперпараметров для всех выбранных архитектур для каждого из вариантов предобработки набора данных проводилось с использованием библиотеки Optuna [13]. Для этого проводили кросс-валидацию на наборе для обучения с использованием пяти блоков данных и стратификацией по целевому событию и AUROC в качестве метрики для оптимизации. На следующем этапе модели обучались на всех записях из обучающего набора при оптимальных гиперпараметрах, после чего их работа исследовалась на записях из набора данных для внутреннего тестирования.

Критериями оценки эффективности и качества моделей служили AUROC, AUPRC, чувствительность, специфичность, точность, прогностическая ценность отрицательного и положительного результатов, а также LR-метрики [14-17]. Доверительные интервалы (95%) значений выбранных метрик оценивались с помощью метода бутстреп путем случайной генерации тысячи псевдовыборок, каждая размером в 400 записей [18]. Относительную значимость предикторов определяли по методу Шепли [19]. Выбор итоговой модели основывался на максимальном значении AUROC при внутреннем тестировании с обязательным условием, что доверительный интервал разницы метрик, полученных при обучении и внутреннем тестировании, включает в себя ноль [20].

Валидация итоговой модели и ее сравнение с шкалой SCORE

Для сравнения итоговой модели со шкалой SCORE [6] по описанной ранее схеме был сформирован набор для валидации модели МО, записи из которого не участвовали ни на одном из этапов разработки. В качестве дополнительного критерия включения записей в этот набор использовали условие обязательного наличия признаков, использующихся в шкале SCORE, а также ИМТ и ЧСС.

В шкале SCORE имеются 4 группы риска на основании 3 порогов: низкий ($<0,01$), умеренный ($\geq0,01$ и $<0,05$), высокий ($\geq0,05$ и $<0,1$) и очень высокий ($\geq0,1$). Для ее сравнения с разработанной моделью МО нами была проведена модификация выделения этих групп, чтобы в зависимости от полученного значения риска относить пациента в одну из следующих категорий: низкий риск – менее 0,01, умеренный (0,01 – менее 0,05) и высокий риск – 0,05 и более.

Пороговые значения одной модели или шкалы нельзя применить к другой, потому что каждый инструмент оценивает вероятность события по-разному, учитывая свои особенности и данные для обучения. Поэтому для сравнения шкалы SCORE с разработанной моделью МО мы подобрали для разработанной модели МО такие пороги активации, чтобы уровни чувствительности модели при их применении минимально отличались от чувствительности медицинской шкалы при определении записей высокого или низкого риска. В результате мы анализировали метрики качества при применении порога 0,223 для модели МО и 0,05 для шкалы (разделение пациентов на группы высокого и умеренного риска в оригинальном исследовании), а также порога 0,031 у модели МО и 0,01 у шкалы (разделение пациентов на группы умеренного и низкого риска). Общий дизайн исследования представлен на рис. 1.

РЕЗУЛЬТАТЫ

Описание наборов данных

Сравнение количественных переменных в группах с целевым событием и без проводили при помощи теста Манна-Уитни,

категориальных – при помощи χ^2 , значение $p<0,05$ считалось статистически значимым.

Заполненность признаков в наборе данных для разработки отражена в приложении 1 (ссылка для доступа онлайн https://disk.360.yandex.ru/i/nEV89dCF8Ae_ZA). Распределение между двумя классами в нем имело статистически значимые различия по величине таких количественных характеристик, как возраст и САД. Также среди пациентов с целевым событием было больше мужчин и указаний на табакокурение. Описание полученных наборов данных с целевым событием и без него представлено в табл. 2.

В набор данных для валидации модели в общей сложности вошло 440 записей пациентов, среди которых было отмечено следующее разделение записей по классам: класс 1 – 31 (7%), класс 0 – 409 (93%). При сравнении предикторов между двумя наборами были обнаружены статистически значимые различия в частоте курящих, а также в распределении возраста, ИМТ, холестерина липопротеидов высокой плотности и ЧСС. Описательная характеристика используемых наборов данных и их сравнение представлены в табл. 2.

Производительность моделей

Наиболее высокую дискриминативную способность в задаче прогнозирования фатальных осложнений ССЗ в течение 10 лет показала модель CatBoost со значением целевой метрики AUROC 0,879 (95% ДИ 0,837-0,918), точностью 0,784 (95% ДИ 0,748-0,820), чувствительностью – 0,938 (95% ДИ 0,818-1,000) и специфичностью – 0,777 (95% ДИ 0,738-0,814) при внутреннем тестировании с использованием в

Таблица 2 — Описательная характеристика наборов данных. Статистически значимые различия выделены жирным шрифтом ($p<0,05$)

Признак	Класс 0 (N = 3691)	Класс 1 (N = 200)	p	Набор для валидации (N = 440)	Набор для разработки (N = 3891)	p
Возраст	48,0 (44,0-53,0)	50,5 (47,0-54,0)	<0,001	51,0 (45,0-55,0)	48,0 (44,0-53,0)	<0,001
ИМТ	27,3 (24,8-30,3)	27,6 (24,7-29,8)	0,496	27,7 (24,5-32,2)	27,4 (24,8-30,2)	0,029
ОХС	5,4 (4,7-6,1)	5,6 (4,9-6,1)	0,094	5,3 (4,6-6,1)	5,4 (4,7-6,1)	0,141
Мужской пол	1863 (50,5%)	178 (89,0%)	<0,001	212 (48,2%)	2041 (52,5%)	0,099
САД	130,0 (120,0-140,0)	136,0 (125,0-150,0)	<0,001	130,0 (120,0-140,0)	130,0 (120,0-140,0)	0,182
Табакокурение	1257 (34,1%)	131 (65,5%)	<0,001	216 (49,1%)	1388 (35,7%)	<0,001
ХС ЛПВП	1,3 (1,1-1,5)	1,3 (1,1-1,5)	0,588	1,2 (1,0-1,4)	1,3 (1,1-1,5)	0,003
ЧСС	72,0 (68,0-76,0)	74,0 (70,0-76,0)	0,205	72,0 (68,0-76,0)	72,0 (68,0-76,0)	0,006

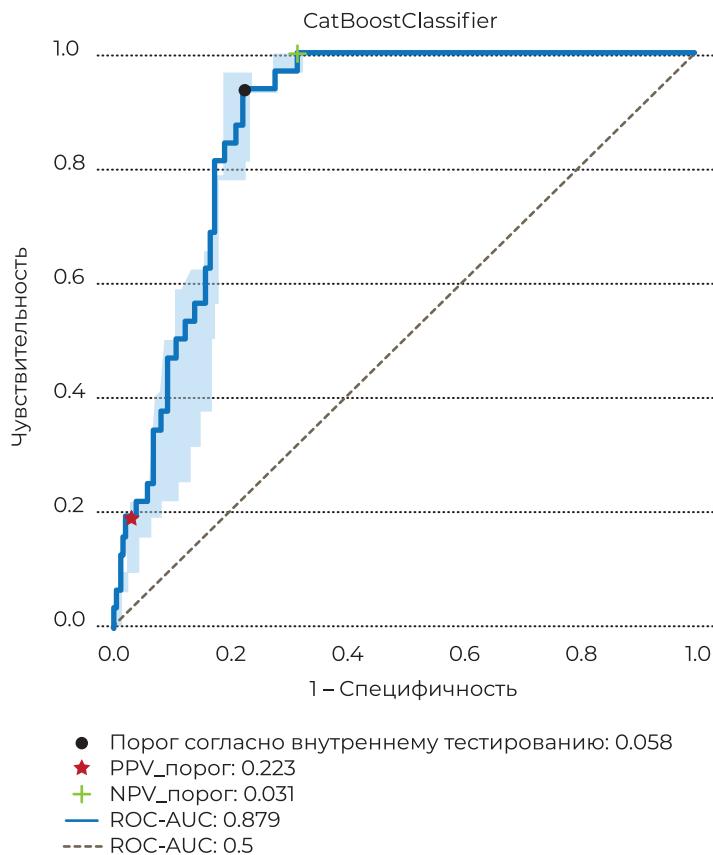


Рисунок 2 — ROC-кривая с 95% ДИ итоговой модели, полученная по результатам прогнозирования на наборе данных для внутреннего тестирования.

качестве порога бинаризации максимума индекса Юдена (0,058). Некоторые метрики представлены в приложении 2 (ссылка для доступа онлайн https://disk.360.yandex.ru/i/nEV89dCF8Ae_ZA). Данные метрики качества были получены при наиболее оптимальной предобработке исходного набора данных, включавшей в себя заполнение пропусков в данных фиксированным значением «-10000», гистограммную нормализацию и применение Random Oversampling. ROC-кривая модели отражена на рис. 2. Средняя абсолютная ошибка калибровочной кривой равнялась 30,34% с систематической переоценкой реального риска развития целевого события.

Сравнение CatBoost с общепринятой медицинской шкалой SCORE

Для данных, не задействованных на этапах разработки и внутреннего тестирования модели, была проведена оценка шкалы SCORE. После

этого на основании подхода с формированием трех групп риска [21–23] нами были рассчитаны два порога бинаризации для выхода модели МО так, чтобы при их использовании на этих данных уровни чувствительности модели и шкалы SCORE для разделения пациентов на группы высокого или низкого риска отличались минимально. Результаты применения полученных порогов показали следующее. Точность нашей модели с порогом классификации 0,031, при котором был достигнут уровень чувствительности, минимально отличающейся от чувствительности шкалы SCORE при выделении групп умеренного и низкого риска, составила 0,673 (95% ДИ 0,630-0,715), специфичность – 0,653 (95% ДИ 0,607-0,699). Метрики качества при использовании второго порога (0,206, при котором чувствительность модели минимально отличалась от шкалы SCORE при выделении записей высокого и умеренного риска): точность – 0,884

(95% ДИ 0,855-0,915), специфичность – 0,917 (95% ДИ 0, 890-0,946).

Среди метрик качества, полученных при применении шкалы SCORE [6] и нашей итоговой модели на основе CatBoost на наборе для валидации, были выявлены отличия в значениях AUROC (приложение 3 (ссылка для доступа онлайн https://disk.360.yandex.ru/i/nEV89dCF8Ae_ZA)), однако не имевшие статистической значимости. При использовании предложенного авторами шкалы порога для разделения пациентов на группы низкого и умеренного риска нами также было отмечено значительное более низкое значение специфичности в сравнении с применением нашей модели с аналогичным порогом. Таким образом, прогнозирование целевого события при использовании модели МО и этого порога с большой надежностью предсказывает летальный исход у анализируемого пациента. Полная характеристика эффективности итоговой модели при внутреннем тестировании и валидации, а также ее сравнение со шкалой SCORE после применения упомянутых порогов и максимума индекса Юдена представлена в приложении 4 (ссылка для доступа онлайн https://disk.360.yandex.ru/i/nEV89dCF8Ae_ZA).

Отдельного внимания заслуживает возможность интерпретации работы модели с клинической точки зрения. Для этого была определена

значимость признаков, рассчитанная согласно векторам Шепли (рис. 3). Первостепенным для прогноза целевого события имели пол, возраст, факт курения, менее значимыми явились значения САД, ИМТ и общего холестерина. В приложении 5А и 5Б (ссылка для доступа онлайн https://disk.360.yandex.ru/i/nEV89dCF8Ae_ZA) представлены интерпретации двух записей (с целевым событием и без него) из набора данных для внутреннего тестирования. На данных схемах отражен конкретный вклад значений отдельных признаков на прогноз и итоговый результат работы модели по прогнозированию целевого события.

Следует отметить, что пол оказался существенным в определении летального исхода. При исследовании значимости пола среди всех записей в тестовом наборе данных нами были выявлены значительные его ассоциации с другими признаками (рис. 4А и 4Б). При сравнении идентичных значений для большинства количественных предикторов между женщинами и мужчинами была выявлена четкая тенденция к более высоким оценкам риска у мужчин. Однако на примере значений возраста обнаружилась более сложная нелинейная закономерность. У женщин прогнозные оценки оставались стабильными вне зависимости от возраста, тогда как у мужчин на уровне 50 лет наблюдалась

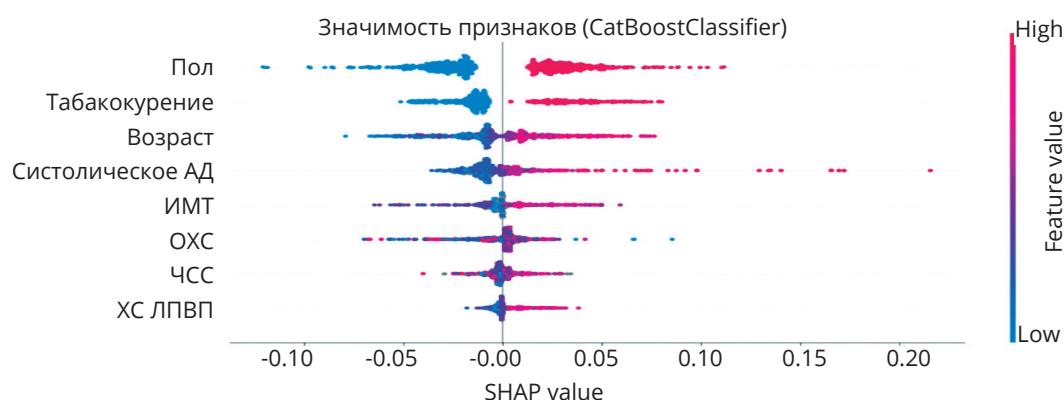


Рисунок 3 — Признаки итоговой модели, показавшие наибольшую значимость в прогнозе целевого события. Примечание: красным цветом обозначено положительное значение (для бинарных признаков) или более высокое значение признака (для количественных), синим цветом обозначено отрицательное значение или более низкое значение признака, соответственно. Серым цветом отмечены пропуски в данных.

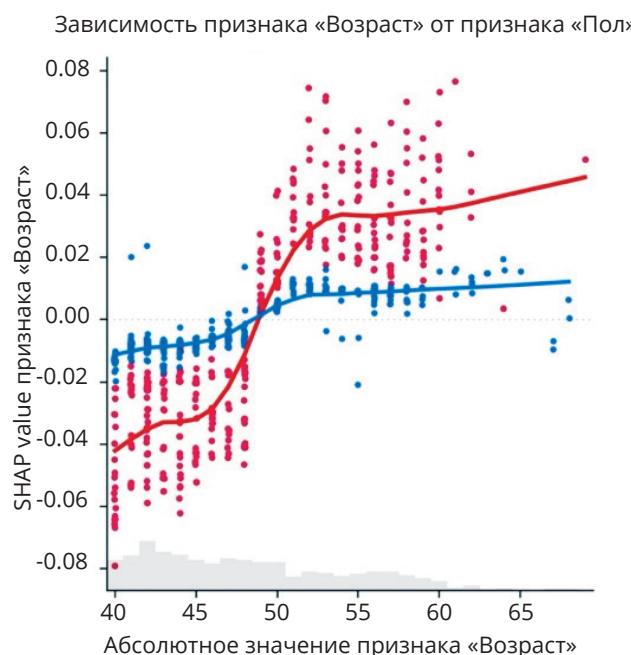
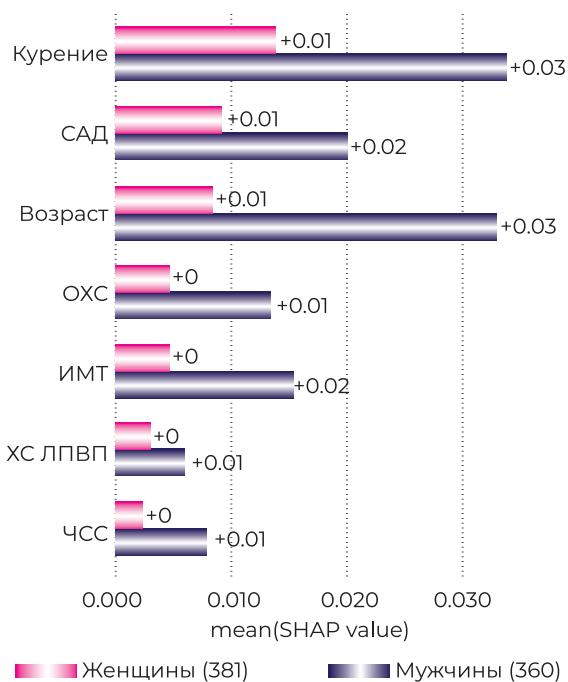

А

Б

Рисунок 4 — Влияние пола пациента на значимость остальных признаков.
Примечание: на рис. 4А красным цветом отмечены наблюдения пациентов мужского пола, синим – женского.

дифференциация. У мужчин младше 50 лет оценки риска были ниже, чем у женщин, тогда как у мужчин старше 50 лет эти оценки значительно превышали женские. Эта выявленная закономерность позволяет нам отчетливо выделить возраст 50 лет как пороговое значение для одного из важнейших признаков для определения группы риска, особенно среди мужчин.

ОБСУЖДЕНИЕ

Широкое внедрение в клиническую практику моделей МО позволяет оптимизировать решение многих рутинных задач в оказании помощи пациенту с сердечно-сосудистой патологией: обоснованное назначение лекарственных препаратов, более объективная обработка результатов инструментальных исследований, диагностика редких заболеваний, предсказание нарушений ритма на основании обработки сигнала ЭКГ и многое другое [24]. В мета-анализе Friedrich S. et al. [25] представлен обзор 215 исследований с применением методов МО и ИИ в кардиологической практике. Примерно поровну распределилось количество используемых

моделей для диагностических и прогностических целей.

Особого внимания требует применение МО в оценке ССР. В опубликованном ранее нами обзоре [26] обсуждалась необходимость и возможность автоматической интерпретации имеющейся в ЭМК информации, позволяющей решить и проблему неточности существующих шкал расчета ССР, и ограниченного количества анализируемых ФР, улучшить возможность врача рассчитывать риск автоматически, не затрачивая на это времени как непосредственно для расчета, так и для поиска необходимых параметров в электронной медицинской системе.

Одно из первых исследований о возможности выявления пациентов высокого риска развития ССЗ с помощью моделей МО – работа Song X. и соавт [27]. В ней авторы обратили внимание на необходимость включать в шкалы оценки риска большее количество входных переменных, что может увеличить их точность. Авторами были созданы две модели, которые анализировали 37 различных признаков и предсказывали смерть пациента от ССЗ в течение ближайших 30 дней

и в течение 1 года, AUC которых составила 0,87 и 0,89 соответственно. Wu J. и соавт. применили несколько методов МО для выявления пациентов высокого риска. Ими была создана модель оценки риска на основе 10 входных признаков и логистической регрессии, обеспечившая AUC 0,77 [28].

Наша модель на основе CatBoost продемонстрировала сопоставимую с SCORE дискриминативную способность (AUROC модели 0,887 против AUROC шкалы SCORE 0,814, без статистически значимых различий), однако обладала преимуществами по ряду других метрик. В частности, при разделении пациентов на группы низкого и умеренного риска специфичность нашей модели была значительно выше, чем у шкалы SCORE, что снижает вероятность ложноположительных классификаций. При разделении пациентов на группы умеренного и высокого риска точность и специфичность также оказались выше у модели МО (0,884 и 0,917, соответственно), что указывает на более надёжное выделение пациентов с высоким риском развития фатального сердечно-сосудистого события. Полученные результаты калибровки, в частности средняя абсолютная ошибка в 30,34%, указывают на направление для дальнейшей оптимизации алгоритма и вероятно будут решаться, в том числе с помощью постобработки результатов. Итоговая модель была протестирована на наборе данных, который статистически значимо отличался по значениям

5 признаков, эта особенность свидетельствует о хорошей способности к обобщению и устойчивости результата модели на новых данных.

Таким образом, предложенная модель превосходит шкалу SCORE по показателям точности и специфичности, что делает её более перспективной для практического применения. Важно отметить, что модели предсказания CCP в нашем исследовании включали большее число параметров, в перечень значимых признаков вошли: пол, возраст, курение, САД, ИМТ, ЧСС, общий холестерин и ХС ЛВП. Кроме того, были выявлены закономерности, связанные с влиянием пола пациента на значимость других признаков в прогнозе (рассчитанную согласно векторам Шепли), что может являться темой для дальнейшего изучения влияния добавления признаков в модели МО.

ЗАКЛЮЧЕНИЕ

Применение методов МО в клинической практике является современным инструментом повышения эффективности профилактики ССЗ и ведения кардиологических пациентов. Созданные нами модели оценки CCP могут быть встроены в медицинскую информационную систему и помогать врачу на рутинном приеме более точно определять риск и назначать конкретные профилактические программы, делая акцент на нескольких, в том числе модифицируемых факторах риска.

ЛИТЕРАТУРА/REFERENCES

1. Roth GA, Mensah GA, Johnson CO, Addolorato G, et al. Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019: Update From the GBD 2019 Study. *Journal of the American College of Cardiology*. Elsevier Inc. 2020; 76: 2982-3021.
2. Damen JAAG, Hooft L, Schuit E, Debray TPA, et al. Prediction models for cardiovascular disease risk in the general population: Systematic review. *BMJ (Online)*. BMJ Publishing Group. 2016; 353.
3. Gavrilov DV, Gusev AV, Nikulina AV, Kuznetsova TY, Drapkina OM. Correctness of cardiovascular risk assessment in daily clinical practice. *Profilakticheskaya Meditsina*. 2021; 24(4): 69-75.
4. Boytsov SA, Pogosova NV. Cardiovascular prevention 2022. Russian national guidelines: Russian Society of Cardiology, National Society of Preventive Cardiology. *Russian Journal of Cardiology*. 2023; 28(5).
5. Global Effect of Modifiable Risk Factors on Cardiovascular Disease and Mortality. *New England Journal of Medicine [Internet]*. 2023; 389(14): 1273-85. Available from: <http://www.nejm.org/doi/10.1056/NEJMoa2206916>.
6. Conroy RM, Pyörälä K, Fitzgerald AP, Sans S, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: The SCORE project. *Eur Heart J*. 2003; 24(11): 987-1003.
7. Tokgozoglu L, Torp-Pedersen C. Redefining cardiovascular risk prediction: Is the crystal ball clearer now? *European Heart Journal*. Oxford University Press. 2021; 42: 2468-71.

8. Kapoor S, Narayanan A. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*. 2023; 4(9).
9. Com YD, Simonoff JS. An Investigation of Missing Data Methods for Classification Trees Applied to Binary Response Data Yufeng Ding. *Journal of Machine Learning Research*. 2010; 11.
10. Cao XH, Stojkovic I, Obradovic Z. A robust data scaling algorithm to improve classification accuracies in biomedical data. *BMC Bioinformatics*. 2016; 17(1).
11. De Amorim LB V, Cavalcanti GDC, Cruz RMO. The choice of scaling technique matters for classification performance [Internet]. Available from: <https://github.com/amorimlb/scaling>.
12. Weiss GM. Foundations of imbalanced learning. 2012.
13. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A Next-generation Hyperparameter Optimization Framework. 2019 Jul 25. Available from: <http://arxiv.org/abs/1907.10902>.
14. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. *Inf Process Manag*. 2009; 45(4): 427-37.
15. Brodersen KH, Ong CS, Stephan KE, Buhmann JM. The balanced accuracy and its posterior distribution. In: *Proceedings – International Conference on Pattern Recognition*. 2010. P.3121-4.
16. Barandela R, Sá Anchez B, JS, Garcā V, Rangel E. Rapid and Brief Communication Strategies for learning in class imbalance problems [Internet]. *Pattern Recognition*. 2003. Available from: www.elsevier.com/locate/patcog
17. Chicco D, Jurman G. The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification. *BioData Min*. 2023; 16(1).
18. Zoubir AM, Iskandler DR. Bootstrap methods and applications. *IEEE Signal Process Mag*. 2007; 24(4): 10-9.
19. Lundberg SM, Erion G, Chen H, Degrave A, et al. Explainable AI for Trees: From Local Explanations to Global Understanding [Internet]. Available from: <https://github.com/suinleelab/treeexplainer-study>.
20. Wester DB. Comparing treatment means: overlapping standard errors, overlapping confidence intervals, and tests of hypothesis. *Biom Biostat Int J*. 2018; 7(1): 73-85.
21. Fischer BG, Evans AT. SpPin and SnNout Are Not Enough. It's Time to Fully Embrace Likelihood Ratios and Probabilistic Reasoning to Achieve Diagnostic Excellence. *J Gen Intern Med*. 2023; 38(9): 2202-4.
22. Baduashvili A, Guyatt G, Evans AT. ROC Anatomy — Getting the Most Out of Your Diagnostic Test. *J Gen Intern Med*. 2019; 34(9): 1892-8.
23. Thomas G, Kenny LC, Baker PN, Tuytten R. A novel method for interrogating receiver operating characteristic curves for assessing prognostic tests. *Diagn Progn Res*. 2017; 1(1).
24. Gill SK, Karwath A, Uh HW, Cardoso VR, et al. Artificial intelligence to enhance clinical value across the spectrum of cardiovascular healthcare. *European Heart Journal*. Oxford University Press. 2023; 44: 713-25.
25. Friedrich S, Groß S, König IR, Engelhardt S, et al. Applications of artificial intelligence/machine learning approaches in cardiovascular medicine: A systematic review with recommendations. *European Heart Journal — Digital Health*. Oxford University Press. 2021; 2: 424-36.
26. Gusev AV, Gavrilov DV, Novitsky RE, Kuznetsova TY, Boytsov SA. Improvement of cardiovascular risk assessment using machine learning methods. *Russian Journal of Cardiology*. 2021; 26(12): 171-80.
27. Song X MACJRK. Comparison of machine learning techniques with classical statistical models in predicting health outcomes. *Stud Health Technol Inform*. 2004; 107: 736-40.
28. Wu J, Roy J, Stewart WF. Prediction Modeling Using EHR Data Challenges, Strategies, and a Comparison of Machine Learning Approaches [Internet]. 2010. Available from: www.lww-medicalcare.com.