



ОТКРЫТЫЙ ДОСТУП

ОПРЕДЕЛЕНИЕ

Сянхун Ма,
Астонский университет, Великобритания

ОБЗОРИ

Саид Нурул Хасан,
Нью-Йоркский университет, США
Нандини Наир,
Университет штата Пенсильвания,
США

*КОРРЕСПОНДЕНЦИЯ

Мария Попцова ✉
mpoptsova@hse.ru

†Эти авторы внесли равный вклад в
данную работу и разделяют первое
авторство

ПРИНЯТО 20 июня 2024 года

ПРИНЯТО 19 августа 2024 года

Опубликован 05 сентября 2024 года

ЦИТАТА

Кирдеев А, Буркин К, Воробьев А, Збировская Е, Лифшиц Г, Николаев К, Зеленская Е, Донников М, Коваленко Л, Урванцева И и Попцова М (2024) Модели машинного обучения для прогнозирования риска MACE у больных инфарктом миокарда с различными генотипами VEGFR2.

Фронт. Мед. 11:1452239.

doi: 10.3389/fmed.2024.1452239

COPYRIGHT

© 2024 Кирдеев, Буркин, Воробьев, Збировская, Лифшиц, Николаев, Зеленская, Донников, Коваленко, Урванцева, Попцова. Это статья с открытым доступом, распространяемая на условиях [лицензии Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).

Использование, воспроизведение или воспроизведение на других форумах разрешено при условии указания автора(ов) и владельца(ов) авторских прав, а также ссылки на оригинальную публикацию в этом журнале в соответствии с принятой академической практикой. Не допускается использование, воспроизведение или воспроизведение, не соответствующее данным условиям.

Модели машинного обучения для прогнозирования риска MACE у пациентов с инфарктом миокарда с различными генотипами VEGFR2

Александр Кирдеев^{1†}, Константин Буркин^{1†}, Антон Воробьев^{2†}, Елена Збировская¹, Галина Лифшиц³, Константин Николаев⁴, Елена Зеленская², Максим Донников², Людмила Коваленко⁵, Ирина Урванцева^{2,6} и Мария Попцова^{*.1}

¹Факультет компьютерных наук, Институт искусственного интеллекта и цифровых наук, Международная лаборатория биоинформатики, Высшая школа экономики, Москва, Россия,² Кафедра кардиологии, Сургутский государственный университет, Сургут, Россия,³ Институт химической биологии и фундаментальной медицины, Новосибирск, Россия,⁴ Федеральный исследовательский центр Институт цитологии и генетики, Новосибирск, Россия,

⁵Кафедра общей патологии и патофизиологии, Сургутский государственный университет, Сургут, Россия,⁶ Югра

Центр диагностики и сердечно-сосудистой хирургии, Сургут, Россия

История вопроса: Разработка прогностических моделей для выявления пациентов с высоким риском развития инфаркта миокарда (МИ) является важнейшим шагом на пути к персонализированной медицине. Известно, что генетические факторы связаны с повышенным риском сердечно-сосудистых заболеваний, однако мало что известно о том, можно ли их использовать для прогнозирования основных неблагоприятных сердечных событий (ОНСС) у пациентов с МИ. Целью данного исследования было построение модели машинного обучения (ML) для прогнозирования MACE у пациентов с МИ на основе клинических, визуализационных, лабораторных и генетических характеристик и оценка влияния генетики на прогностическую силу модели.

Методы: Проанализированы данные 218 пациентов с МИ, поступивших в отделение неотложной помощи Сургутского окружного центра диагностики и сердечно-сосудистой хирургии (Россия). При поступлении у каждого пациента были собраны стандартные клинические показатели и данные визуализации. Кроме того, пациенты были генотипированы на вариацию VEGFR-2 rs2305948 (генотипы C/C, C/T, T/T, T - аллель минорного риска). Исследование включало 9-летний период наблюдения, в течение которого регистрировались основные ишемические события. Мы обучили и оценили различные модели ML, включая Gradient Boosting, Random Forest, Logistic Regression и AutoML. Для анализа важности признаков применялись методы последовательного выбора признаков (SFS) и схема аддитивного объяснения Шэпли (SHAP).

Результаты: Алгоритм CatBoost с признаками, отобранными с помощью метода SFS, показал наилучшую эффективность на тестовой когорте, достигнув ROC AUC 0,813. Анализ важности признаков выявил дозу статинов как наиболее важный фактор, а генотип VEGFR-2 вошел в пятерку лучших. Другими важными признаками являются поражение коронарных артерий (стенозы коронарных артерий $\geq 70\%$), параметры левого желудочка (ЛЖ), такие как боковая стенка ЛЖ и масса ЛЖ, диабет, тип реваскуляризации (КАБГ или ПКА) и возраст. Мы также показали, что эти факторы являются аддитивными и что высокий риск может быть определен совокупным негативным влиянием различных прогностических факторов.

Выводы: Наш подход, основанный на ML, показал, что генотип VEGFR-2 ассоциируется с повышенным риском MACE у пациентов с МИ. Однако этот риск можно значительно снизить с помощью высокодозных статинов и таких положительных факторов, как отсутствие поражения коронарных артерий, отсутствие диабета и более молодой возраст.

КЛЮЧЕВЫЕ СЛОВА

машинное обучение, инфаркт миокарда, ген рецептора 2 фактора роста эндотелия сосудов, статины, прогностическое моделирование, **MACES**, стенка левого желудочка

1 Введение

Несмотря на значительные достижения в области сердечно-сосудистой хирургии, интервенционных процедур и фармакотерапии, внутрибольничные и долгосрочные показатели основных неблагоприятных сердечных событий (MACE) - включая смерть, повторный острый коронарный синдром (ОКС), инсульт и реваскуляризацию миокарда - остаются высокими во всем мире среди пациентов с инфарктом миокарда (МИ) (1, 2).

Программы вторичной профилактики для таких людей основаны на выявлении пациентов с высоким риском с помощью моделей, ориентированных на клинические признаки МИ (отклонение сегмента ST, фракция выброса левого желудочка (ФВЛЖ), класс Киллипа и поражение коронарных артерий) и традиционные факторы риска (гипертония, диабет, хроническая болезнь почек, дислипидемия и курение), которые ухудшают сердечные исходы (3, 4).

Выявление пациентов с высоким риском MACE имеет решающее значение. Существуют различные клинические баллы, такие как GRACE, ACEF, CADILLAC, TIMI, PROCAM, PREDICT, PURSUIT, DAPT и PRECISE-DAPT. Эти баллы, в основном основанные на регрессионных моделях с использованием клинических параметров, оценивают сердечно-сосудистый (СС) риск в течение 1-3 лет после индекса МИ. Тем не менее, современные персонализированные подходы к лечению лиц после МИ уже требуют более точных клинических систем оценки, включающих биомаркеры и генетические данные (5, 6).

Патологической основой ишемической болезни сердца (ИБС) и острого коронарного синдрома (ОКС) является атеросклероз. Такие факторы, как нарушения липидного обмена, повреждение эндотелиальных клеток сосудов, воспаление и иммунная дисфункция, могут способствовать развитию и прогрессированию коронарного атеросклероза, потенциально приводя к САД и АКШ. Семейство факторов роста эндотелия сосудов (VEGF) участвует в ангиогенезе, воспалении, окислительном стрессе и липидном обмене, представляя потенциальную терапевтическую и прогностическую ценность для АКШ и МИ (7, 8).

Прогностическая эффективность сывороточных уровней VEGF после ишемии сердца вскоре после МИ была ранее изучена в когортном исследовании коронарных заболеваний. VEGF в основном связываются с тремя тирозинкиназными рецепторами (VEGFR-1, VEGFR-2 и VEGFR-3) с разной степенью сродства. Однонуклеотидные полиморфизмы (SNPs) VEGFR-2 связаны с развитием сердечно-сосудистых заболеваний (7). Более того, были выявлены три SNP гена VEGFR-2, которые значимо ассоциированы с коронарным атеросклерозом: +1192C > T (rs2305948), -604 T > C (rs2071559) и +1719A > T (rs1870377).

(9-13). Следовательно, аллельные варианты гена VEGFR-2 могут быть оценены как прогностические маркеры при МИ

(14).

В последнее время подходы машинного обучения (МЛО) доказали свою эффективность в прогнозировании исходов пациентов на основе клинических данных, данных визуализации и биомаркеров. Самое крупное исследование, основанное на клинических данных 23 000 пациентов из регистров BleeMACS и RENAMI, показало, что ML-модели могут предсказывать смерть от всех причин, повторные МИ и крупные кровотечения после АКШ (15), используя 25 клинических признаков. Другие подходы, основанные на значительно меньших наборах данных, также доказали эффективность ML-моделей для исследования сердечно-сосудистых исходов (16-20). Один из них

Преимуществом ML-моделирования является возможность проводить анализ важности признаков и ранжировать их в соответствии с их вкладом в производительность модели.

Это позволяет не только выявить наиболее важные признаки, но и выбрать минимальный набор, который может уменьшить перегрузку и повысить точность модели.

Кроме того, ML-модель с небольшим набором признаков может быть легко применена на практике. Так, ML-модели показали, что только сывороточный креатинин и LVEF могут предсказать смерть от всех причин (16). В другом исследовании (21) ML-модель, обученная предсказывать смерть от всех причин, сократила пространство признаков с 430 до 25. Другие модели подтверждают прогностическую силу ML-моделей и представляют наиболее важные признаки. К

ним относятся модели прогнозирования посттромботического синдрома (22), прогнозирования времени до смерти пациентов в отделении коронарной терапии (23), прогнозирования смертности и госпитализации при сердечной недостаточности (CH) у пациентов с сохраненным LVEF (17), прогнозирования долгосрочного риска МИ и сердечной смерти (24), а также внутрибольничной смертности по всем причинам у пациентов с СН (20). Все вышеупомянутые модели ML были обучены с использованием клинических характеристик, регулярно измеряемых при поступлении в больницу. В некоторых исследованиях эти признаки дополнялись демографическими и социальными характеристиками. Все предложенные прогностические ML-модели не обладают высокими показателями эффективности с ROC AUC в интервале 70-80%.

Одним из способов повышения предсказательной способности ML-моделей является включение дополнительной информации о состоянии пациента, при этом генетические маркеры потенциально могут играть ключевую роль. Добавление генетических маркеров в качестве признаков в ML-модель не означает, что клинические признаки не являются прогностическими или что в будущем они должны быть заменены только генетическими маркерами. Однако генетические маркеры могут значительно улучшить оценку риска пациента, если их учитывать вместе с другими клиническими характеристиками. Применение такого рода моделей ограничено из-за дороговизны генетического тестирования, а большинство наборов клинических данных пациентов с МИ не содержат генетической информации.

Здесь мы впервые включили генетическую информацию в ML-модель для прогнозирования MACE у пациентов с МИ.

проведения экстренной коронарной ангиографии (КАГ). Значительные поражения (стенозы) коронарных артерий считались имеющимися при сужении просвета более чем на 70%. Для левой главной коронарной артерии (ЛГКА) значительным считался стеноз при сужении просвета более чем на 70%.

2 Методы

2.1 Сбор клинических данных

В проспективное наблюдательное исследование последовательно включены 218 пациентов с острым МИ, поступивших в Сургутскую окружную кардиологическую клинику Центра диагностики и сердечно-сосудистой хирургии. Данное исследование было одобрено местным этическим комитетом клиники. Во время первичной госпитализации в отделение неотложной помощи медицинские исследователи (кардиологи) объясняли пациентам основные положения протокола исследования и получали их информированное согласие на участие. Затем пациентов транспортировали в операционный блок для

50%. По результатам КАГ выполнялась баллонная ангиопластика и стентирование или шунтирование коронарных артерий (КАБГ). Во время лечения в стационаре также проводилась стандартная трансторакальная эхокардиография.

Все пациенты получали терапию, основанную на рекомендациях, включая блокаторы РААС, бета-блокаторы, статины и двойную антитромбоцитарную терапию. Высокодозная терапия статинами определялась как прием аторвастатина 40-80 мг и розувастатина 20-40 мг в день, а низкодозная - как прием аторвастатина 10-20 мг и розувастатина 5-10 мг в день в течение как минимум 12 месяцев после ИМ.

Критериями включения были: острый инфаркт миокарда с подъемом или без подъема сегмента ST и возраст 30-70 лет. Критерии исключения: противопоказания к применению статинов; беременность и кормление грудью; прогрессирующее опухолевое заболевание; ВИЧ; отказ пациента от участия в исследовании.

На лабораторном этапе из лейкоцитарных колец в собранных образцах крови выделяли ДНК, которую затем замораживали при температуре -80°C для последующего генетического тестирования. Генотипы VEGFR-2 (rs2305948) (C/C, C/T, T/T) определяли методом полимеразной цепной реакции (ПЦР) в реальном времени с помощью прибора Real-Time CFX96 Touch (Bio-Rad Laboratories, США). Т-аллель VEGFR-2 (rs2305948) обычно рассматривается как аллель риска неблагоприятного исхода ишемии после ИМ. Скорость клубочковой фильтрации (СКФ) оценивали с помощью калькулятора CKD-EPI для классификации хронической болезни почек. Основные ишемические события (сердечно-сосудистая смерть, повторное ОКС, инсульт и реваскуляризация миокарда) регистрировались по результатам клинического наблюдения в период от 9 до лет (2015-2024 гг.) после индексированного МИ. Информация была собрана из внутрибольничных медицинских карт пациентов, повторно поступивших в клинику после индекса МИ. Кроме того, с пациентами связывались по телефону, а факт стационарного лечения по поводу ишемических событий в других медицинских организациях подтверждался путем запроса соответствующих медицинских карт. Случаи смерти в постинфарктном периоде регистрировались в базе данных полисов медицинского страхования, а сердечно-сосудистые (ишемические) причины смерти уточнялись службами клинического вскрытия по месту жительства пациента. Базовый статистический анализ проводился в программном пакете SPSS 23.0.

2.2 Предварительная обработка данных

Для логистической регрессии все некатегориальные признаки были нормализованы с помощью Standard Scaler. Для древовидных моделей, не требующих нормализации, преобразование данных не проводилось. Дополнительная предварительная обработка категориальных признаков не проводилась, поскольку все категориальные признаки были бинарными или упорядоченными.

Для вменения данных мы использовали метод Multivariate Imputation by Chained Equations (MICE) (25). В качестве

оценки для MICE использовалась модель линейной регрессии с параметрами по умолчанию со 100 итерациями для числовых данных. Для категориальных и бинарных признаков мы использовали модель логистической регрессии с решателем 'liblinear' в качестве оценки по умолчанию для MICE со 100 итерациями.

Для разработки и проверки моделей набор данных был случайным образом разделен в соотношении 75:25 на обучающий ($n = 163$) и тестовый ($n = 55$) наборы соответственно. На обучающем наборе были проведены интерполяция и масштабирование данных, настройка гиперпараметров, выбор признаков и окончательное обучение модели. Оценка метрик разработанных моделей проводилась на тестовом наборе.

Мы построили модели машинного обучения на основе алгоритмов Gradient Boosting (Catboost и LightGBM), Random Forest и Logistic Regression с настройкой гиперпараметров с помощью 10-кратной кросс-валидации на обучающем наборе данных. Для подхода AutoML мы использовали FLAML (26), пакет Python для автоматического машинного обучения, который автоматизирует процесс выбора модели, настройки гиперпараметров и построения признаков.

Для оптимизации работы модели мы использовали Optuna (27), пакет Python, автоматизирующий настройку гиперпараметров с помощью алгоритмов байесовской оптимизации. В качестве алгоритма поиска мы использовали древовидную структуру Parzen Estimator с 10 стартовыми испытаниями и применяли 10-кратную кросс-валидацию в процессе оптимизации для получения надежной оценки метрики ROC AUC. Для определения оптимального количества обучающих признаков использовался метод последовательного прямого отбора (SFS). Во время цикла отбора признаков мы использовали 5-кратную кросс-валидацию и использовали среднее значение ROC AUC в качестве метрики оценки. Модель с лучшим средним показателем была выбрана в качестве базовой модели для процедуры выбора признаков. Мы добавляли по одному признаку за раз, пока не исследовали все признаки, а затем выбрали подмножество с лучшей метрикой в качестве окончательного подмножества признаков.

Анализ важности признаков проводился с помощью схемы аддитивного объяснения Шэпли (SHAP) (28). SHAP обеспечивает визуализацию важности признаков для выбранной модели и индивидуального вклада признаков в прогностическую силу модели. Кроме того, мы построили графики частичной зависимости (PDP) и индивидуального условного ожидания (ICE) (29).

Чтобы оценить прогностическую эффективность разработанных моделей, мы оценили образцы данных из тестового набора. В частности, мы сгенерировали 1000 бутстреп-образцов исходной тестовой выборки, чтобы получить распределение PR AUC и ROC AUC. Усреднив показатели по всем итерациям, мы получили средние значения PR AUC и ROC AUC в качестве итоговых оценочных показателей.

3 Результаты

3.1 Характеристика пациентов с МИ

В набор данных, использованных в данном исследовании, вошли 218 пациентов, поступивших в отделение неотложной помощи Сургутского окружного центра диагностики и сердечно-сосудистой хирургии (Россия) с МИ. Для каждого пациента были отобраны стандартные клинические показатели, включая лабораторные анализы и показатели визуализации, для включения в модель (табл. 1). Кроме того, каждый пациент был генотипирован на вариацию VEGFR-2, rs2305948, которая включает генотипы C/C, C/T и T/T.

3.2 Модели машинного обучения

Данные подверглись интерполяции 339/предварительной обработке и были разделены на обучающие и тестовые наборы в соотношении 75-25% для обучения и проверки. Чтобы выбрать лучшую модель, мы обучили и протестировали следующие алгоритмы ML: Gradient Boosting (CatBoost и LightGBM), Random Forest, Logistic Regression и подход AutoML (подробнее см. раздел "Методы"). Для оценки производительности моделей мы брали

ТАБЛИЦА 1 Клинические и визуализационные характеристики, лабораторные и генетические исследования пациентов с МИ (n = 218).

Характеристики	N
Клинические и визуализационные характеристики	
Пол, n (%)	
Мужчина	169 (78)
Женщина	49 (22)
Возраст, лет, М ± SD	58 ± 10
ИМТ, кг/м ² , Ме [25%; 75%]	29 [26; 33]
Отклонение сегмента ST на ЭКГ, n (%)	
Возвышение	74 (34)
Неэлеваторный	144 (66)
Инфаркты стенок ЛЖ, n (%)	
Передняя стенка	95 (44)
Задняя стенка	97 (45)
Боковая стенка	79 (36)
Класс Killipa, n (%)	
I-II	213 (98)
III-IV	5 (2)
LV EF	
< 40%	16 (7)
≥ 40%	202 (93)
Гипертония, n (%)	179 (82)
Сахарный диабет 2 типа, n (%)	52 (24)
Хроническая болезнь почек ≥ C2, n (%)	74 (34)
Семейный анамнез ССЗ	35 (16)
Стенотические поражения коронарных артерий, n (%)	
Левая основная коронарная артерия	14 (7)
Одно судно	79 (36)
Два судна	63 (29)
≥ 3 судна	72 (33)
Продолжительность стенокардии от начала острого МИ, n (%)	
1-3 h	167 (77)
3-12 h	9 (4)
> 12 h	42 (19)
Реваскуляризация миокарда, n (%)	
PCI	207 (95)
CABG	11 (5)
Статины (аторвастатин, розувастатин)	
Высокие дозы	117 (54)
Низкие дозы	101 (46)
Лабораторные и генетические исследования	
Высокочувствительный тропонин Т, нг/л, Ме [25%; 75%]	110 [45; 408]
Холестерин ЛПНП, ммоль/л, М ± SD	3 ± 1
Гемоглобин, г/л, Ме [25%; 75%]	139 [128; 149]
Глюкоза, ммоль/л, Ме [25%; 75%]	6 [5; 8]
Креатинин, мкмоль/л, Ме [25%; 75%]	83 [72; 95]

(Продолжение)

ТАБЛИЦА 1 (продолжение)

Характеристики	N
Генотипы VEGFR-2 (rs2305948), n (%)	
C/C - дикий тип, гомозиготный по основному аллелю	142 (65)
C/T-гетерозигота (мажорный аллель + минорный аллель)	75 (34)
T/T - гомозиготны по минорному аллелю	1 (1)

C - цитозин-мажорный аллель (дикий); T - тимин-минорный аллель с атеротромботическим риском.

среднее значение ROC AUC, рассчитанное по бутстреп-выборкам из тестового набора. Сравнение производительности моделей представлено на [рисунке 1А](#), и наилучшая метрика была получена для модели CatBoost (ROC AUC 0,787), которая была значительно лучше второго по эффективности Random Forest ($p < 1e-03$, t -test) ([рисунок 1А](#)). ROC-кривые CatBoost для обучающего и тестового наборов приведены на [Дополнительном рисунке S1](#). Мы выбрали модель CatBoost для отбора признаков и анализа их важности.

3.3 Выбор характеристик для поиска наилучшей модели

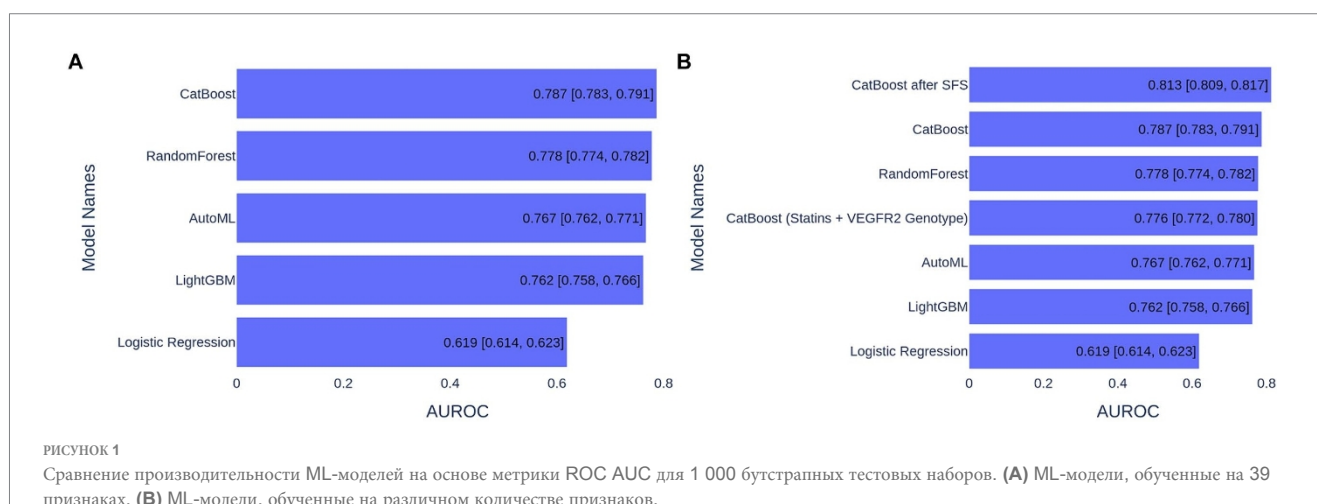
Первоначальные ML-модели были обучены на 39 признаках ([табл. 2](#)). Мы исследовали вклад каждого признака в эффективность модели, используя метод прямого последовательного отбора признаков (SFS) (см. [рис. 2](#)). Мы выбрали метод SFS из-за его баланса между вычислительной эффективностью и надежностью. SFS напрямую оценивает вклад каждого признака в эффективность модели с помощью перекрестной валидации на каждом шаге, обеспечивая устойчивый и точный процесс отбора признаков. Это позволяет избежать потенциальных неточностей, связанных с эвристическими методами, такими как важность признаков в древовидных методах или вменение признаков, которые могут вводить в заблуждение, особенно при высокой кардинальности признаков. Таким образом, SFS - это простой и понятный метод для выбора минимального набора признаков, обеспечивающих максимальную точность ML-модели.

Поскольку CatBoost показал наилучшую производительность с 39 признаками, он был использован в качестве базовой модели для SFS. Производительность модели постепенно увеличивалась при добавлении первых девяти признаков, но затем начала снижаться по мере включения дополнительных признаков. Эти первые девять признаков (представленные на [рисунке 3](#)) были отобраны для переобучения модели CatBoost ([дополнительная таблица S1](#)). Модель CatBoost, обученная на девяти отобранных признаках, продемонстрировала статистически значимое улучшение производительности по сравнению с моделью CatBoost, обученной на 39 признаках (ROC AUC 0,813 против 0,787, $p = 5e-19$, t -test) ([рис. 1В](#) и [дополнительный рисунок S2](#)). Таким образом, для дальнейшего анализа пациентов с МИ была выбрана модель CatBoost, обученная на девяти признаках.

3.4 Анализ важности признаков

Для анализа важности признаков мы выбрали SHAP, поскольку он оценивает вклад выбранных признаков в эффективность модели и учитывает совместное влияние других признаков. Значения SHAP можно суммировать, чтобы показать совместный вклад нескольких признаков в итоговое предсказание ML-модели. Это помогает понять совместное влияние нескольких признаков на оценку риска.

10.3389/fmed.2024.1452239



Мы применили метод SHAP, используя лучшую модель CatBoost с девятью признаками, отобранными методом SFS (рис. 3). Девять признаков ранжированы от наиболее важных (вверху) до наименее важных (внизу). Более высокое значение SHAP для признака указывает на больший риск MACE. Значения характеристик выделены цветом: красным отмечены более высокие значения, синим - более низкие. Из графика SHAP видно, что наиболее важной характеристикой является доза статинов, причем низкие значения дозы соответствуют высокому риску MACE. Интересно, что генотип VEGFR-2 оказался четвертым по значимости признаком, причем рискованный Т-аллель соответствует более высокому риску MACE. Другими важными характеристиками являются индекс коморбидности, поражение коронарных артерий, параметры ЛЖ, такие как боковая стенка ЛЖ, перенесенные КАБГ или ПКА, индекс массы тела (ИМТ) и пол (см. рис. 3).

Кроме того, мы построили графики частичной зависимости (PDP) и индивидуального условного ожидания (ICE) для ключевых признаков, чтобы обеспечить более интуитивное понимание их влияния на предсказания (Дополнительный рисунок S3). Графики PDP показывают среднее влияние признака на прогнозируемый результат, в то время как графики ICE иллюстрируют влияние признака на уровне отдельных наблюдений, подчеркивая изменчивость в разных точках данных. Представленные кривые PDP и ICE дают подробную визуализацию того, как отдельные признаки влияют на прогнозируемый риск MACE у пациентов с МИ. Графики PDP отображают предельный эффект каждого признака, в то время как кривые ICE иллюстрируют изменчивость прогнозов для разных пациентов.

Для непрерывных переменных графики выявили следующие тенденции: повышенный ИМТ и более высокий индекс коморбидности связаны с увеличением риска, а более высокая доза статинов - с уменьшением риска. Аналогично, большая тяжесть поражения коронарных артерий и поражения левой главной коронарной артерии приводит к увеличению риска. Средний эффект представлен оранжевой пунктирной линией, а индивидуальные траектории пациентов показаны синим цветом (Дополнительный рисунок S3).

Для категориальных признаков гистограммы подчеркивают четкие различия в уровнях риска: мужской пол, проведение PCI вместо CABG, поражение боковой стенки ЛЖ при МИ и наличие рискованного Т-аллеля в генотипе VEGFR2 - все это соответствует

большему риску. Эти визуализации подчеркивают важность учета как средних тенденций, так и индивидуальных вариаций при оценке предсказательной силы ML-модели.

Мы построили ML-модель на основе только двух признаков: наиболее важного признака - дозы статинов и генетического признака - генотипа VEGFR2, и эффективность модели достигла 0,80 ROC AUC, что

3.5 Модельная оценка риска для отдельных пациентов

С помощью методов SHAP можно оценить вклад каждого параметра в индивидуальный риск пациента. Оценка риска по модели ML для пациентов с высоким и низким риском с соответствующими значениями SHAP представлена на рисунках 4, 5.

У пациента с высоким риском 92% (рис. 4) пять основных факторов риска: высокий индекс коморбидности (оценка = 4; красная стрелка на рис. 4) - 7,9%, затем критическое поражение левой основной коронарной артерии (>50%) - 7%, низкая доза статинов - 5,5%, Т-аллель риска - 5,2% и многососудистое поражение коронарных артерий - 4%. Кроме того, вовлечение боковой стенки ЛЖ, PCI и ИМТ 32 кг/м² увеличивают риск на 3,9, 1,3 и 1,1% соответственно.

На примере пациента с низким риском MACE среди исследуемых пациентов (риск 15%) видно, что все характеристики вносят положительный вклад (обозначены синими стрелками) в снижение общего риска (рис. 5). Первым параметром, снижающим риск, является КАБГ, за которым следует отсутствие поражения коронарных артерий. Однако поражение левой коронарной артерии увеличивает риск почти на 7 % (указано красной стрелкой на рисунке 5). Низкий индекс коморбидности и высокая дозировка статинов снижают риск на 6,6 и 5,9 % соответственно. Другие факторы, такие как нормальная масса тела, генетика без риска, отсутствие поражения боковых отделов ЛЖ и пол, еще больше снижают риск в совокупности на 8 %.

4 Обсуждение

Несколько исследований показали сильную ассоциацию между VEGFs, сывороточным уровнем VEGFRs и другими факторами роста с клиническими характеристиками и прогнозом при СКВ (30). Ранее было показано, что SNPs VEGFR-2 связаны с развитием таких ССЗ, как САД, СКВ и болезнь Кавасаки (11), а также оценивались как прогностические маркеры при инсульте и СКВ. Таким образом,

в

ТАБЛИЦА 2 Перечень 39 признаков, первоначально включенных в модель.

1.	Возраст
2.	Передняя стенка ЛЖ
3.	Фибрилляция предсердий
4.	BMI
5.	BMS или DES
6.	Индекс коморбидности
7.	Поражение коронарных артерий
8.	Креатинин
9.	Диабет
10.	Продолжительность боли
11.	Секс
12.	Глюкоза
13.	Частота сердечных сокращений
14.	Гемоглобин
15.	Семейный анамнез ССЗ
16.	Гипертония
17.	Артерия, связанная с инфарктом
18.	Класс Киллип
19.	LV EDV
20.	LV EF
21.	LV ESV
22.	Размер инфаркта ЛЖ
23.	Масса ЛЖ
24.	Латеральная стенка ЛЖ
25.	Левое предсердие
26.	Поражение левой основной коронарной артерии
27.	Мультифокальный атеросклероз
28.	PCI или CABG
29.	После МИ
30.	Пост-стент
31.	Постинсультное состояние
32.	Сегмент ST
33.	Дозировка статинов
34.	Систолическое АД
35.	Степень TIMI в артерии, связанной с инфарктом
36.	TnTPO
37.	Общий холестерин
38.	Генотип VEGFR2
39.	ЭГФР EPI

Выделены девять признаков, отобранных методом SFS.

В метаанализе Qui et al. (10) VEGFR-2 rs2305948 и rs1870377 (оба находятся в экзонах рецептора VEGF-2 и приводят к аминокислотным заменам, которые снижают сродство связывания VEGF с рецептором VEGF-2), но не rs2071559, были связаны с повышенным риском инсульта. В другом мета-анализе Wang et al. (31) доказали, что полиморфизмы VEGFR2 (rs1870377, rs2071559 и rs2305948) могут быть использованы для выявления лиц с повышенным риском инсульта.

предрасположенность к атеросклеротическим сердечно-сосудистым заболеваниям. В другом исследовании, основанном на Нанкинском регистре инсультов, rs1870377 мог предсказать 3-месячный исход пациентов с атеросклеротическим инсультом на крупных артериях (32). Marks et al. показали, что ни один из SNP VEGFR-2 (rs1870377, rs2071559 и rs2305948) не связан со смертностью у 2067 пациентов с СКВ, включенных в Новозеландское исследование коронарной болезни (New Zealand Coronary Disease Cohort Study) (14). Мы впервые включили генотип VEGFR-2 в качестве признака в прогностическую модель исхода больных МИ и обнаружили, что генетический фактор входит в пятерку наиболее важных признаков.

Мы также оценили влияние различных факторов - клинических, визуализационных, генетических и связанных с лечением - на прогноз риска. Используя разработанный ML-подход, мы продемонстрировали достаточное негативное влияние различных факторов на комбинированные клинические конечные точки, включая сердечно-сосудистую смерть, повторную АКШ, инсульт и реваскуляризацию миокарда. Ключевыми факторами являются: клинические факторы - мужской пол, высокий индекс массы тела и сложная коморбидность; визуализационные факторы - многососудистое поражение коронарных сосудов и поражение боковой стенки ЛЖ; генетические факторы - малые аллельные варианты VEGFR-2 rs2305948; факторы лечения - ЧКВ только на инфарктсвязанную коронарную артерию вместо КАБГ при многососудистом поражении коронарных артерий и низкоинтенсивная дозировка статинов. В то же время высокие и низкие дозы статинов, а также КАБГ и PCI в исследуемой когорте пациентов с МИ показали значительные преимущества в снижении числа MACE в течение 9 лет наблюдения.

Преклонный возраст, женский пол в менопаузе (33) и сопутствующие заболевания, такие как диабет, ожирение и дислипидемия, являются хорошо известными традиционными факторами риска MACE (повторное АКШ, инсульт и смерть) у пациентов с МИ. За последние 25 лет в эпидемиологии сердечно-сосудистых событий произошел резкий сдвиг. Заболеваемость МИ и ишемическим инсультом снизилась в 3-4 раза, что можно объяснить существенными изменениями в популяционных факторах риска, в том числе значительным снижением систолического артериального давления, концентрации холестерина и частоты курения, хотя это было компенсировано увеличением индекса массы тела, распространенности диабета и других сопутствующих заболеваний. Кроме того, положительное влияние оказало более широкое использование после ИМ кардиопротективных препаратов, таких как высокодозные статины, ингибиторы P2Y12-тромбоцитов, бета-блокаторы и новые препараты в реальной клинической практике (34-37). Вовлечение боковой стенки ЛЖ при МИ является негативным прогностическим маркером дезадаптивного ремоделирования сердца, часто требующим более современных методов лечения (например, хирургического восстановления желудочков) для предотвращения развития застойной сердечной недостаточности (38). В последние два десятилетия активно обсуждается роль многососудистого поражения сердца (включая поражение левой основной коронарной артерии) в прогнозе пациентов с МИ (39, 40). По данным регистра

SWEDHEART, КАБГ у пациентов с поражением ЛКА ассоциируется с более низкой смертностью и меньшим количеством MACE по сравнению с PCI (41). Полная реваскуляризация с помощью PCI после МИ снижает смертность от всех причин, сердечно-сосудистую смертность, рецидивы АКШ и повторные симптомные реваскуляризации. Немедленная полная реваскуляризация с помощью PCI или КАБГ при МИ может быть одинаково полезной, но это требует проведения дополнительных исследований "голова к голове" (42).

Сравнение с другими прогностическими моделями ML выявило сходство наших результатов в отношении клинических и визуализационных факторов. В исследовании (15) анализ важности признаков выявил LVEF, возраст, диабет, расчетный GFR, многососудистое заболевание, заболевание периферических артерий, полную реваскуляризацию, уровень гемоглобина, предыдущие кровотечения,

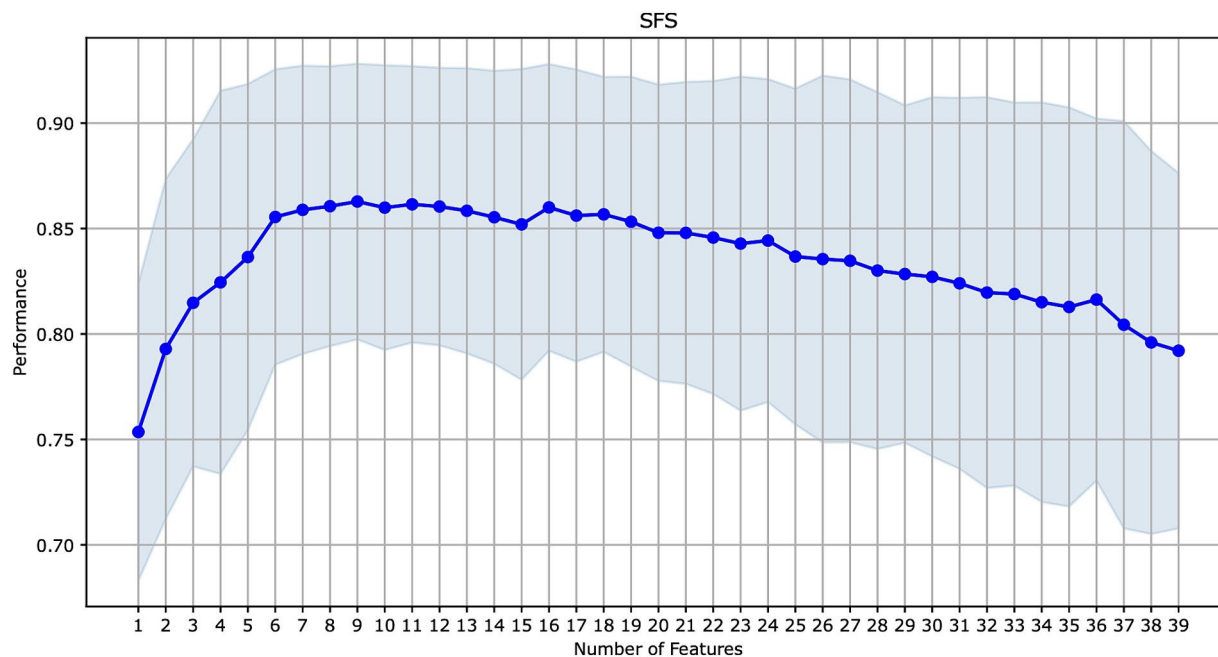


РИСУНОК 2

Последовательный прямой выбор признаков для модели с наилучшей производительностью по всем признакам (CatBoost).

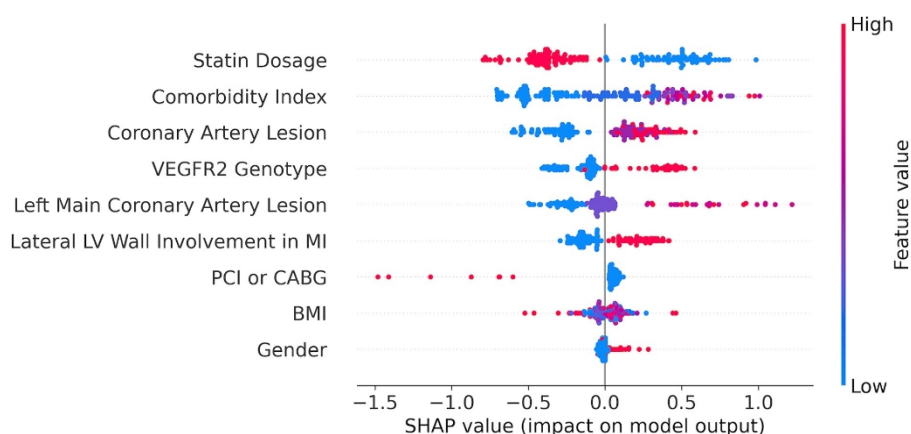


РИСУНОК 3

График важности признаков SHAP для модели Catboost, построенной на основе 9 признаков, отобранных по SFS.

Злокачественные новообразования, ингибиторы АПФ или АРБ при выписке и терапия статинами при выписке - наиболее важные характеристики для прогнозирования смерти от всех причин и повторного МИ. Как и в наших исследованиях, значимыми факторами риска МАСЕ оказались сопутствующие заболевания (возраст, многососудистая патология, диабет, хроническая болезнь почек и злокачественные новообразования), полная реваскуляризация миокарда, а также применение высокоинтенсивных статинов в стационаре и после выписки.

ML-модели для прогнозирования смерти от всех причин в течение 3 лет наблюдения, основанные на клинических данных общенационального проспективного регистра ОИМ в Корее ($n = 13\,104$) (43), выявили следующие 10 ведущих предикторов: использование статинов при выписке, пол, индекс массы тела, использование ингибиторов гликопротеина IIb/IIIa, продолжительность госпитализации,

классификация коронарного поражения, NT-proBNP, общий холестерин, время от двери до баллона, и

Кирдеев и др.

Большинство предикторов оказались значимыми и для российской популяции МИ в нашей прогностической модели для комбинированной конечной точки. В частности, к наиболее значимым характеристикам относятся: поражение коронарных сосудов, прием статинов (в высокой дозе) при выписке, пол и индекс массы тела. Wang et al. (44) представили модель риска ML, включающую 21 переменную, характеризующую пациента, для 6-месячного прогнозирования MACE для 1004 китайских пациентов, перенесших коронарную реваскуляризацию. Модель также выявила, что возраст и CABG являются наиболее важными факторами.

В другом крупном проспективном исследовании (45), основанном на когорте UK Biobank, авторы выявили 10 предикторов для 473 611 участников, не страдающих ССЗ, а именно: возраст, пол, уровень холестерина и артериального давления, соотношение холестерина (общий/липопротеины высокой плотности), систолическое давление.

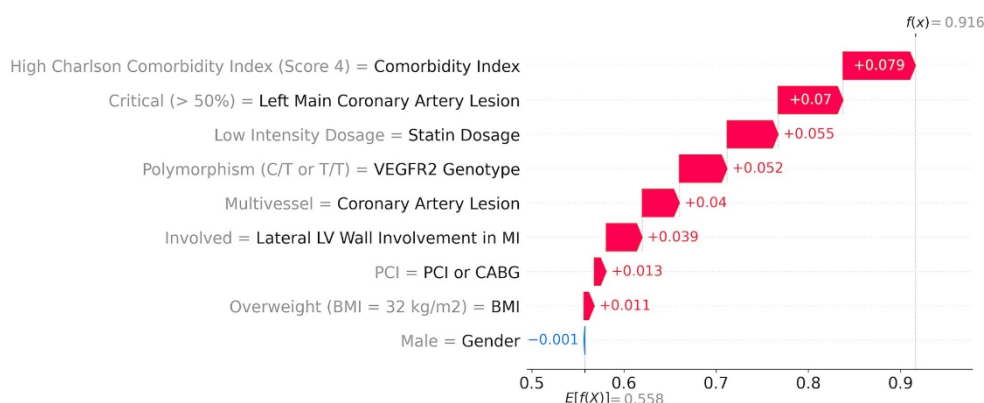


РИСУНОК 4

ML-модель оценки риска для пациента с высоким риском.

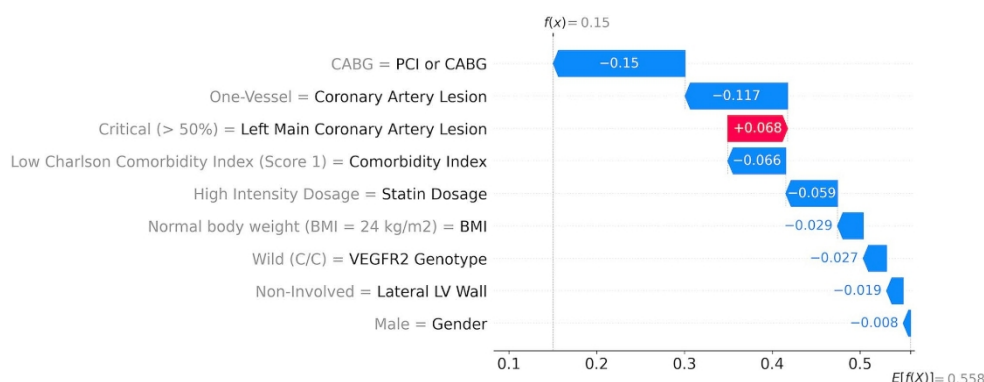


РИСУНОК 5

Оценка риска по модели ML для пациента с низким риском.

артериальное давление, предшествующая стенокардия или болезнь сердца, количество принимаемых лекарств, цистатин С, боль в груди и количество пачек курения в год. Прямое сравнение затруднено из-за различий в когортах (участники без ССЗ или пациенты после ИБС); тем не менее, холестеринные препараты, возраст, ишемическая болезнь сердца и сопутствующие заболевания (например, диабет) также были определены как наиболее важные факторы.

Что касается наилучшей производительности модели, то первоначальный выбор модели осуществлялся с помощью фреймворка AutoML (FLAML), который оценивал несколько алгоритмов, включая Random Forest, CatBoost и LightGBM. Процесс AutoML определил CatBoost как оптимальный базовый оценщик, подчеркнув его баланс между точностью и обобщением. Хотя Random Forest показал более высокие показатели производительности на тестовом наборе данных, это можно объяснить чрезмерной подгонкой, о чем свидетельствуют его исключительно высокие результаты. CatBoost, с другой стороны, продемонстрировал производительность без чрезмерной подгонки, что делает его более надежным выбором для нашего клинического приложения. Алгоритмы бустинга, в целом, хорошо документированы для обеспечения современной производительности для табличных данных. В частности, CatBoost с подстройкой гиперпараметров оказался в среднем самой эффективной моделью, как показано в [46]. Такая стабильность результатов подчеркивает надежность CatBoost

для нашего исследования. В конечном итоге CatBoost продемонстрировал превосходную производительность на тестовом наборе данных, что подтверждает его пригодность в качестве основного алгоритма.

Настоящее исследование демонстрирует, что генотипы VEGFR-2 могут быть использованы в качестве предиктора MACE. Хотя эти результаты могут быть специфичны для исследуемой популяции из-за отсутствия внешней валидации, полученные результаты обеспечивают прочную основу для проведения более масштабных исследований. Такие исследования могут привести к разработке универсального прогностического инструмента для MACE с использованием этого генетического биомаркера.

Насколько нам известно, генетические факторы ранее не использовались в моделях ML для прогнозирования MACE у пациентов с МИ, в основном из-за того, что секвенирование или даже однонуклеотидное генотипирование не проводится в больницах на регулярной основе. В этом отношении данное исследование является новаторским, поскольку показывает, что аллель риска варианта VEGFR-2 входит в пятерку наиболее важных факторов прогнозирования долгосрочных исходов у пациентов с МИ.

Мы видим большой потенциал для интеграции модели в текущую клиническую практику. Оценка риска MACE может быть интегрирована в телемедицинский сервис и автоматически оцениваться на основе клинических и генетических данных пациента. Телемедицинский сервис будет предоставлять обратную связь о состоянии пациента и способствовать внесению корректировок в модель, тем самым повышая прогностическую точность модели.

Направления будущих исследований включают совершенствование ML-модели путем обучения на большем наборе данных, проведения внешней валидации и изучения других генетических маркеров, которые, как известно, связаны с риском MACE. Последнее является предметом обширного исследования

Во всем мире, а генетическая информация пациентов поможет улучшить стратегии персонализированного лечения.

редактирование.

5 Заключение

Мы разработали ML-модель, способную предсказывать долгосрочные (9 лет) ишемические исходы у пациентов с МИ с ROC AUC 0,813, основанную на девяти признаках, выбранных из первоначального набора 39. Впервые мы включили генетические факторы в прогностическую модель ML для оценки риска пациентов с МИ в долгосрочной перспективе. Генотипы VEGFR-2 rs2305948 (C/C, C/T, T/T) продемонстрировали высокую прогностическую силу, при этом T-аллель увеличивал риск развития MACE. Кроме того, мы продемонстрировали, что терапия высокими дозами статинов в течение 12 месяцев после МИ, наряду с другими факторами, может минимизировать кардиологический риск у пациентов с T-аллелем риска.

Интеграция генотипов VEGFR-2 в модели прогнозирования MACE требует более масштабного исследования. Однако мы считаем эту модель подходящей для оценки риска у пациентов Сургутского окружного кардиологического диспансера. Для немедленного внедрения этой модели в клиническую практику необходимо простое в использовании приложение для оценки риска.

Заявление о доступности данных

Исходные данные, подтверждающие выводы этой статьи, будут предоставлены авторами без излишних оговорок.

Заявление об этике

Исследования с участием людей были одобрены Сургутским государственным университетом, Сургут, Россия. Исследования проводились в соответствии с местным законодательством и институциональными требованиями. Участники дали письменное информированное согласие на участие в данном исследовании.

Авторский вклад

АК: Формальный анализ, Исследование, Методология, Валидация, Написание - первоначальный проект, Написание - рецензирование и редактирование, Визуализация. KB: Формальный анализ, Расследование, Методология, Валидация, Визуализация, Написание - первоначальный проект, Написание - рецензирование и редактирование. AV: Формальный анализ, Исследование, Методология, Валидация, Написание - первоначальный проект, Написание - рецензирование и редактирование, Концептуализация, Курирование данных, Получение финансирования, Администрирование проекта, Надзор. EZb: Формальный анализ, Расследование, Администрирование проекта, Написание - рецензирование и редактирование. GL: Формальный анализ, Написание - рецензирование и

KN: Методология, ресурсы, написание - рецензирование и редактирование, исследование, получение финансирования, наблюдение, обработка данных, формальный анализ. EZe: Курирование данных, Формальный анализ, Методология, Написание - проверка и редактирование. MD: Формальный анализ, Написание - рецензирование и редактирование, Расследование. LK: Формальный анализ, Написание - рецензирование и редактирование, Получение финансирования, Ресурсы, Наблюдение. IU: Получение финансирования, Ресурсы, Надзор, Написание - обзор и редактирование, Расследование, Методология, Администрирование проекта. MP: Приобретение финансирования, Исследование, Методология, Администрирование проекта, Надзор, Написание - обзор и редактирование, Концептуализация, Курация данных, Формальный анализ, Проверка, Написание - первоначальный проект.

Финансирование

Автор(ы) заявляют о получении финансовой поддержки для проведения исследований, авторства и/или публикации данной статьи. Статья подготовлена в рамках проекта "Зеркальные лаборатории" Университета ВШЭ.

Благодарности

Ссылки

1. Hall M, Smith L, Wu J, Hayward C, Batty JA, Lambert PC, et al. Исходы для здоровья после инфаркта миокарда: популяционное исследование 56 миллионов человек в Англии. *PLoS Med.* (2024) 21:e1004343. doi: 10.1371/journal.pmed.1004343
2. Chi G, Lee JJ, Kazmi SH, Fitzgerald C, Memar Montazerin S, Kalayci A, et al. Ранние и поздние рецидивы сердечно-сосудистых событий среди пациентов высокого риска с острым коронарным синдромом: мета-анализ исследований фазы iii и последствия для дизайна испытаний. *Clin Cardiol.* (2022) 45:299-307. doi: 10.1002/clc.23773

Мы благодарим рецензентов за полезные комментарии, которые помогли значительно улучшить рукопись.

Конфликт интересов

Авторы заявляют, что исследование проводилось в отсутствие каких-либо коммерческих или финансовых отношений, которые могли бы быть истолкованы как потенциальный конфликт интересов.

Примечание издателя

Все утверждения, высказанные в этой статье, принадлежат исключительно авторам и не обязательно отражают точку зрения их аффилированных организаций, а также издателя, редакторов и рецензентов. Любой продукт, который может быть оценен в этой статье, или утверждение, которое может быть сделано его производителем, не гарантируется и не поддерживается издателем.

Дополнительный материал

Дополнительные материалы к этой статье можно найти в Интернете по адресу: <https://www.frontiersin.org/articles/10.3389/fmed.2024.1452239/full#supplementary-material>

3. Grundy SM, Stone NJ, Bailey AL, Beam C, Birtcher KK, Blumenthal RS, et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APHA/ASPC/NLA/PCNA Руководство по управлению уровнем холестерина в крови: резюме: отчет целевой группы Американской коллегии кардиологов/Американской ассоциации сердца по руководству клинической практикой. *Circulation.* (2019) 139:e1046-81. doi: 10.1161/CIR.0000000000000624
4. Liang S, Wang C, Zhang J, Liu Z, Bai Y, Chen Z, et al. Triglyceride-glucose index and coronary artery disease: a systematic review and meta-analysis of risk, severity, and

прогноз. *Cardiovasc Diabetol.* (2023) 22:170. doi: 10.1186/s12933-023- 01906-4

5. Хунг Дж, Роос А, Кадешё Е, Макалистер ДА, Кименаи ДМ, Шах АС, и др. Эффективность оценки по шкале Grace 2.0 у пациентов с инфарктом миокарда 1 и 2 типа. *Eur Heart J.* (2021) 42:2552-61. doi: 10.1093/eurheartj/ehaa375

6. Scirica BM. Острый коронарный синдром: новые инструменты для диагностики и оценки риска. *J Am Coll Cardiol.* (2010) 55:1403-15. doi: 10.1016/j.jacc.2009.09.071

7. Zhou Y, Zhu X, Cui H, Shi J, Yuan G, Shi S, et al. Роль семейства Vegf в развитии ишемической болезни сердца. *Front Cardiovasc Med.* (2021) 8:738325. doi: 10.3389/fcvm.2021.738325

8. Парадовска-Горыцка А, Штыпинска Б, Павлик А, Малиновски Д, Романовска-Прочница К, Манчак М, и др. Генетические варианты Kdr (Vegfr2) и уровень в сыворотке крови у пациентов с ревматоидным артритом. *Biomol Ther.* (2019) 9:355. doi: 10.3390/biom9080355

9. Камаре К., Пусель М., Герпе-Сальвайр А., Сальвайр Р. Антигенез в атеросклеротической бляшке. *Redox Biol.* (2017) 12:18-34. doi: 10.1016/j.redox.2017.01.007

10. Qiu S, Wu T, Wang P, Li J, Li Q, Du J. Ассоциация между полиморфизмами гена Vegfr и инсультом: мета-анализ. *PLoS One.* (2016) 11:e0151371. doi: 10.1371/journal.pone.0151371

11. Лю Д, Сонг Ж, Джи Х, Лю З, Конг М, Ху Б. Ассоциация генетических полиморфизмов Vegfa и Vegfr2 с риском ишемической болезни сердца. *Медицина.* (2016) 95:e3413. doi: 10.1097/MD.0000000000003413

12. Jin F, Hagemann N, Brockmeier U, Schäfer ST, Zechariah A, Hermann DM. Ldl ослабляет Vegf-индуцированный ангиогенез через механизмы, включающие интернализацию и деградацию Vegfr2 после перемещения по сети эндосом-транс-Голги. *Ангиогенез.* (2013) 16:625-37. doi: 10.1007/s10456-013-9340-2

13. Холмс К., Робертс О. Л., Томас А. М., Кросс М. Дж. Рецептор-2 фактора роста эндотелия сосудов: структура, функция, внутриклеточная сигнализация и терапевтическое ингибирование. *Cell Signal.* (2007) 19:2003-12. doi: 10.1016/j.cellsig.2007.05.013

14. Marks EC, Wilkinson TM, Frampton CM, Skelton L, Pilbrow AP, Yandle TG, et al. Плазменные уровни растворимых изоформ рецептора Vegf, циркулирующих птерино и Snps системы Vegf как прогностические биомаркеры у пациентов с острыми коронарными синдромами. *BMC Cardiovasc Disord.* (2018) 18:1-10. doi: 10.1186/s12872-018-0894-1

15. Д'Аскенцо Ф, Де Филиппо О, Галлоне Г, Миттоне Г, Дериу МА, Ианнаконе М, и др. Прогнозирование неблагоприятных событий после острого коронарного синдрома (praise) на основе машинного обучения: модельное исследование объединенных наборов данных. *Lancet.* (2021) 397:199-207. doi: 10.1016/S0140-6736(20)32519-8

16. Chicco D, Jurman G. Машинное обучение может предсказывать выживаемость пациентов с сердечной недостаточностью только на основе сыровороточного креатинина и фракции выброса. *BMC Med Inform Decis Mak.* (2020) 20:1-16. doi: 10.1186/s12911-020-1023-5

17. Angraal S, Mortazavi BJ, Gupta A, Khara R, Ahmad T, Desai NR, et al. Машинное обучение прогнозирует смертность и госпитализацию при сердечной недостаточности с сохраненной фракцией выброса. *JACC Heart Fail.* (2020) 8:12-21. doi: 10.1016/j.jchf.2019.06.013

18. Chirinos JA, Orlenko A, Zhao L, Basso MD, Cvijic ME, Li Z, et al. Множественные биомаркеры плазмы для стратификации риска у пациентов с сердечной недостаточностью и сохраненной фракцией выброса. *J Am Coll Cardiol.* (2020) 75:1281-95. doi: 10.1016/j.jacc.2019.12.069

19. Zhao H, Li P, Zhong G, Xie K, Zhou H, Ning Y, et al. Модели машинного обучения у пациентов с сердечной недостаточностью с умеренно сниженной фракцией выброса. *Front Cardiovasc Med.* (2022) 9:1042139. doi: 10.3389/fcvm.2022.1042139

20. Mranya D, Celik T, Klug E, Ntsinjana H. Прогнозирование внутрибольничной смертности по всем причинам при сердечной недостаточности с помощью машинного обучения. *Front Cardiovasc Med.* (2023) 9:1032524. doi: 10.3389/fcvm.2022.1032524

21. Lin H, Xue Y, Chen K, Zhong S, Chen L. Прогнозирование риска острого коронарного синдрома на основе градиентного усиленного древовидного отбора признаков и рекурсивного исключения признаков: исследование моделирования с учетом конкретного набора данных. *PLoS One.* (2022) 17:e0278217. doi: 10.1371/journal.pone.0278217

22. Yu T, Shen R, You G, Lv L, Kang S, Wang X, et al. Прогнозирование посттромботического синдрома на основе машинного обучения: разработка модели и валидационное исследование. *Front Cardiovasc Med.* (2022) 9:990788. doi: 10.3389/fcvm.2022.990788

23. Wang X, Zhu T, Xia M, Liu Y, Wang Y, Wang X, et al. Прогнозирование прогноза пациентов в отделении коронарной терапии: новая многокатегорийная модель машинного обучения с использованием Xgboost. *Front Cardiovasc Med.* (2022) 9:764629. doi: 10.3389/fcvm.2022.764629

24. Commandeur F, Slomka PJ, Goeller M, Chen X, Cadet S, Razipour A, et al. Machine learning to predict long-term risk of myocardial infarction and cardiac death based on clinical risk, coronary calcium, and epicardial adipose tissue: a prospective study. *Cardiovasc Res.* (2020) 116:2216-25. doi: 10.1093/cvr/cvz321

25. Van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. *J Stat Softw.* (2011) 45:1-67. doi: 10.18637/jss.v045.i03

26. Wang C, Wu Q, Weimer M, Zhu E. Flaml: быстрая и легкая библиотека Automl. *Proc Mach Learn Res.* (2021) 3:434-47. doi: 10.48550/arXiv.1911.04706

27. Акиба Т., Сано С., Янасе Т., Охтя Т., Кояма М., редакторы. Optuna: Рамка оптимизации гиперпараметров нового поколения. Материалы 25-й международной конференции ACM SIGKDD по обнаружению знаний и добыче данных; (2019).

28. Лундберг CM, Ли С-И. Единый подход к интерпретации предсказаний моделей. *Adv Neural Inf Proces Syst.* (2017) 30:4768-77.

29. Goldstein A, Kapelner A, Bleich J, Pitkin E. Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation. *J Comput Graph Stat.* (2015) 24:44-65. doi: 10.1080/10618600.2014.907095

30. Ferrara N, Gerber HP, LeCouter J. The biology of Vegf and its receptors. *Nat Med.* (2003) 9:669-76. doi: 10.1038/nm0603-669

31. Wang L, Ge H, Peng L, Wang B. Мета-анализ взаимосвязи между полиморфизмами Vegfr2 и атеросклеротическими сердечно-сосудистыми заболеваниями. *Clin Cardiol.* (2019) 42:860-5. doi: 10.1002/clc.23233

32. Li Z, Wang M, Gu J, Zhao L, Guo Y, Zhang Z, et al. Миссенс-варианты в индуцируемой гипоксией сигнализации Vegfa/Vegfr2 предсказывают исход атеросклеротического инсульта крупных артерий. *Cell Mol Neurobiol.* (2021) 41:1217-25. doi: 10.1007/s10571-020-00890-7

33. Tschiderer L, Seekircher L, Willeit P, Peters SA. Оценка сердечно-сосудистого риска у женщин: прогресс на данный момент и прогресс в будущем. *Int J Women's Health.* (2023) 15:191-212. doi: 10.2147/IJWH.S364012

34. Shah AS, Lee KK, Pérez JAR, Campbell D, Astengo F, Logue J, et al. Клиническое бремя, влияние факторов риска и исходы после инфаркта миокарда и инсульта: 25-летнее исследование связи на уровне отдельных пациентов. *Lancet Reg Health Eur.* (2021):7. doi: 10.1016/j.lanre.2021.100141

35. Li L, Scott CA, Rothwell PM, Study OV. Тенденции заболеваемости инсультом в странах с высоким уровнем дохода в XXI веке: популяционное исследование и систематический обзор. *Stroke.* (2020) 51:1372-80. doi: 10.1161/STROKEAHA.119.028484

36. Сафи С., Сети Н.Дж., Коранг С.К., Нильсен Е.Е., Файнберг Дж, Глууд С. и др. Бета-блокаторы у пациентов без сердечной недостаточности после инфаркта миокарда. *Cochrane Database Syst Rev.* (2021) 11:CD012565. doi: 10.1002/14651858.CD012565.pub2

37. Yu D, Liao JK. Новые взгляды на плеiotропию статинов и снижение уровня холестерина. *Cardiovasc Res.* (2022) 118:413-23. doi: 10.1093/cvr/cvab032

38. Patel ND, Williams JA, Nwakanma LU, Weiss ES, Conte JV. Влияние инфаркта миокарда боковой стенки на исходы после хирургического восстановления желудочков. *Ann Thorac Surg.* (2007) 83:2017-28. doi: 10.1016/j.athoracsur.2007.02.026

39. G ripenberg T, Johhaji F, Östlund-Papadogeorgos N, Ekenbäck C, Linder R, Samad B, et al. Outcome and selection of revascularization strategy in left main coronary artery stenosis. *Scand Cardiovasc J.* (2018) 52:100-7. doi: 10.1080/14017431.2018.1429648

40. Puymirat E, Nakache A, Saint Etienne C, Marcollet P, Fichaux O, Decomis MP, et al. Связано ли многососудистое коронарное заболевание у пациентов с острым инфарктом миокарда с худшими клиническими исходами в течение 1 года? *Clin Cardiol.* (2021) 44:429-37. doi: 10.1002/clc.23567

41. Persson J, Yan J, Angeräs O, Venetsanos D, Jeppsson A, Sjögren I, et al. Pci или Cabg для лечения заболевания левой главной коронарной артерии: регистр Swedeheart. *Eur Heart J.* (2023) 44:2833-42. doi: 10.1093/eurheartj/ehad369

42. Баннинг А.П. Стратегии ревазуляризации у пациентов с МИ и МПП: Есть ли у нас ответы? *J Am Coll Cardiol.* (2024) 84:295-7. doi: 10.1016/j.jacc.2024.05.019

43. Jeong JH, Lee KS, Park SM, Kim SR, Kim MN, Chae SC, et al. Прогнозирование продольных клинических исходов после острого инфаркта миокарда с помощью динамического алгоритма машинного обучения. *Front Cardiovasc Med.* (2024) 11:1340022. doi: 10.3389/fcvm.2024.1340022

44. Wang J, Wang S, Zhu MX, Yang T, Yin Q, Hou Y. Прогнозирование риска возникновения основных неблагоприятных сердечно-сосудистых событий в течение 6 месяцев после коронарной ревазуляризации: исследование с применением машинного обучения. *JMIR Med Inform.* (2022) 10:e33395. doi: 10.2196/33395

45. You J, Guo Y, Kang JJ, Wang HF, Yang M, Feng JF, et al. Разработка моделей на основе машинного обучения для прогнозирования 10-летнего риска сердечно-сосудистых заболеваний: проспективное когортное исследование. *Stroke Vasc Neurol.* (2023) 8:475-85. doi: 10.1136/svn-2023-002332

46. McElfresh D, Khandagale S, Valverde J, Prasad CV, Ramakrishnan G, Goldblum M, et al. Когда нейронные сети превосходят ускоренные деревья на табличных данных? В материалах 37-й Международной конференции по нейронным системам обработки информации (NIPS'23). Orleans, LA, USA (2023) 76336-76369.