



¹ Медицинский
исследовательский
центр терапии
и профилактической
медицины
Минздрава России

² Государственное
учреждение
здравоохранения
Тульской области
«Киреевская
центральная
районная больница»

³ ПАО «Группа
Ренессанс
Страхование»

⁴ Кыргызская
государственная
медицинская академия
им. И.К. Ахунбаева

Возможности оценки сердечно-сосудистого риска с применением искусственного интеллекта без использования предикторов, нуждающихся в инвазивных методах определения

И.А. Мишкин^{1, 2}, А.В. Концевая¹, А.А. Сахаров³, А.Г. Полупанов⁴,
О.М. Драпкина¹

Адрес для переписки: Илья Александрович Мишкин, Ilya.mischckin@yandex.ru

Для цитирования: Мишкин А.И., Концевая А.В., Сахаров А.А. и др. Возможности оценки сердечно-сосудистого риска с применением искусственного интеллекта без использования предикторов, нуждающихся в инвазивных методах определения. Эффективная фармакотерапия. 2025; 21 (8): 8–14.

DOI: 10.33978/2307-3586-2025-21-8-8-14

Сердечно-сосудистые заболевания (ССЗ) остаются основной причиной смертности в России, несмотря на то что достижения профилактической медицины и мобильного здравоохранения (mHealth) способствуют снижению заболеваемости. Объединение искусственного интеллекта (ИИ), носимых устройств и методов неинвазивной оценки риска позволяет применять эти технологии для массового скрининга и раннего выявления групп повышенного риска.

Цель – изучить и сравнить точность оценки сердечно-сосудистого риска с использованием и без использования факторов риска, связанных с инвазивными методами определения, на примере населения регионов России и Кыргызстана. Для этого была применена одна из моделей ИИ.

Материал и методы. В анализе использовались данные исследования «ИНТЕРЭПИД», проведенного в 2011–2016 гг. В исследовании приняли участие 2392 пациента из Самарской области (Российская Федерация) и Чуйской области (Кыргызская Республика). Для обработки данных применялся алгоритм искусственного интеллекта – GradientBoostingClassifier, который прошел процедуру 5-кратной кросс-валидации. Основным показателем качества модели стала ROC-AUC.

Результаты. Мы изучили, насколько эффективно градиентный бустинг может прогнозировать риск ССЗ. В ходе исследования были построены модели, которые продемонстрировали ROC-AUC от 0,564 (Самарская когорта) до 0,806 (Чуйская когорта), в общей когорте – 0,766. В Самарской когорте ключевыми предикторами стали наличие в анамнезе ССЗ, уровень креатинина и липопротеидов высокой плотности, в Чуйской — возраст, уровень тревожности, потребление колбас и избыточный вес. Исключение инвазивных переменных улучшило качество моделей: ROC-AUC повысился на 6,4% в Самарской когорте, на 0,9% в Чуйской когорте, но не изменился в общей когорте. Прогноз, основанный на ИИ, превзошел традиционные шкалы SCORE и Framingham. В общей когорте он оказался на 6,4% лучше, чем по Framingham, и на 16,9% лучше, чем по SCORE ($p < 0,05$).

Выводы. Полученные данные демонстрируют, что алгоритмы ИИ потенциально могут заменять традиционные шкалы при отсутствии лабораторных данных. Это открывает перспективы для применения ИИ в первичной профилактике ССЗ в условиях ограниченного доступа к медицинским ресурсам.

Ключевые слова: сердечно-сосудистый риск, оценка риска, искусственный интеллект, машинное обучение



Введение

За последние годы в России удалось значительно снизить уровень сердечно-сосудистых заболеваний (ССЗ), однако эта патология по-прежнему остается ведущей причиной смертности в стране [1]. В основе здравоохранения в РФ лежат методы профилактической медицины, которые помогают предотвратить развитие заболеваний на ранних стадиях или до их развития [2, 3]. Это позволяет минимизировать прямой и косвенный экономический ущерб, который государство несет в связи с утратой трудоспособного населения из-за хронических неинфекционных заболеваний, включая ССЗ [4].

В последнее время все большую популярность набирает мобильная медицина (mHealth). В частности, это относится к мобильным приложениям, позволяющим отслеживать состояние здоровья (wellness-приложения). Их использование помогает врачу получить предварительные результаты о состоянии пациента до того, как он придет на осмотр в медицинское учреждение [5].

С увеличением доступности смартфонов и мобильного Интернета технологии mHealth открывают новые горизонты в области профилактической медицины. Они позволяют проводить массовые скрининги без необходимости использования специального оборудования, что значительно повышает вероятность выявления групп людей с факторами риска на ранних стадиях. Кроме того, предварительная оценка рисков позволяет оптимизировать затраты на здравоохранение [6].

Таким образом, для создания организационной технологии предварительной оценки уровня риска среди широких слоев населения целесообразно рассмотреть возможность применения технологий определения сердечно-сосудистого риска, не требующих инвазивной диагностики. Кроме того, актуальность этой темы обусловлена стремительным развитием технологии искусственного интеллекта (ИИ) в области здравоохранения и растущей популярностью носимых электронных устройств, позволяющих в реальном времени отслеживать состояние организма [7].

Цель данной работы – изучить и сравнить точность оценки сердечно-сосудистого риска с использованием и без использования факторов риска, определяемых инвазивными методами обследования, на популяции регионов РФ и Кыргызстана с использованием одной из моделей ИИ.

Материал и методы

Работа выполнена с использованием данных исследования «ИНТЕРЭПИД», проведенного в 2011–2016 гг. База данных содержала записи о 2392 пациентах из двух когорт – из Самарской области РФ и Кыргызской Республики, – прошедших наблюдение в течение семи лет [8, 9].

Во время исследования были проанализированы факторы риска развития ССЗ (всего 192 переменные), в том числе:

- социальный статус;
- семейный анамнез;
- наличие факторов риска;
- вопросник о наличии жалоб и симптомов;
- личный анамнез;
- объективные данные;
- данные лабораторных и инструментальных исследований;
- данные об обращаемости за медицинской помощью;
- госпитальная шкала тревоги и депрессии;
- опросник по уровню стресса;
- опросник по качеству жизни.

Самарская когорта включала 1050 (43,9%) человек:

- 447 (42,6%) мужчин, средний возраст ($M \pm \text{сигма}$) $49,8 \pm 9,2$ года;
- 603 (57,4%) женщины, средний возраст ($M \pm \text{сигма}$) $49 \pm 9,2$ года.

Курили 21,2% участников исследования, повышенные цифры артериального давления отмечались у 49,4%, наличие в анамнезе инфаркта миокарда (ИМ) – у 1,3%, инсульта – у 1%, сахарного диабета (СД) 2-го типа – у 2,9%, уровень общего холестерина плазмы крови (ОХС) выше 5,0 ммоль/л – у 36%.

Чуйская когорта составила 1341 (56,1%) человек:

- 575 (42,9%) мужчин, средний возраст ($M \pm \text{сигма}$) $40 \pm 13,2$ года;
- 766 (57,1%) женщин, средний возраст ($M \pm \text{сигма}$) $41 \pm 13,2$ года.

Курили 24,5% участников исследования, повышенное артериальное давление отмечалось у 39,1%, наличие ИМ в анамнезе – у 1,9%, инсульта – у 3,1%, ХСН – у 6,1%, СД 2-го типа – у 3,7%, уровень ОХС выше 5,0 ммоль/л – у 40,8%.

Всего по двум когортам:

- мужчин – 1022 (42,7%), средний возраст ($M \pm \text{сигма}$) $44,3 \pm 12,8$ года;
- женщин – 1369 (57,2%), средний возраст ($M \pm \text{сигма}$) $44,6 \pm 12,2$ года.

Курили 24,5% участников, повышенное артериальное давление (АД) отмечалось у 39,1%, наличие ИМ в анамнезе – у 1,9%, инсульта – у 3,1%, ХСН – у 6,1%, СД 2-го типа – у 3,7%, уровень ОХС выше 5,0 ммоль/л – у 40,8%.

В рамках проспективного этапа наблюдения продолжительностью семь лет были зафиксированы следующие конечные точки:

- случаи смерти от всех причин;
- случаи смерти от ССЗ;
- нефатальные сердечно-сосудистые события, включая развитие ИМ и острого нарушения мозгового кровообращения.

Для формирования датасета для обучения ИИ базу данных исследования «ИНТЕРЭПИД» разделили на две группы.

Первая группа – пациенты, перенесшие на этапе проспективного наблюдения фатальные и нефатальные сердечно-сосудистые события, включая ишемическую болезнь сердца (ИБС) (МКБ 10: I20–I25); острые нарушения мозгового кровообращения (МКБ 10: I60–I64). За период наблюдения исследованы 253 (24,1%) человека в Самарской когорте и 280 (20,9%) – в Чуйской когорте, всего 533 (22,5%) человека.



Вторая группа – пациенты без сердечно-сосудистых событий за период наблюдения: 797 (75,9%) в Самарской и 1061 (79,1%) в Чуйской когорте, всего 1858 (77,5%) человек.

Фатальные и нефатальные события, вызванные другими заболеваниями, в исследование не включались. Для построения модели прогнозов был использован метод ИИ – GradientBoostingClassifier (градиентный бустинг) (рис. 1).

Данный алгоритм был выбран, поскольку градиентный бустинг входил в тройку наиболее часто используемых алгоритмов для создания прогнозов на табличных данных в проведенном систематическом обзоре [10], а также имел наиболее высокую долю результатов с высокими показателями площади под кривой (Area Under the Curve (AUC)) в ранее опубликованных работах [11–13].

Поскольку оба датасета содержали относительно небольшое количество записей, для тестирования модели было решено использовать алгоритм GridSearchCV из пакета sklearn.model_selection, который реализует

метод кросс-валидации (cross-validation, CV) на пяти фолдах (5-fold cross-validation).

Для оценки эффективности моделей использован показатель ROC-AUC. Для оценки показателей мультиколлинеарности использован показатель VIF (Variance Inflation Factor) с помощью statsmodels.stats.outliers_influence.importance.variance_inflation_factor. Для оценки вклада каждого предиктора в формирование прогноза проведено ранжирование факторов риска по степени их влияния на результат с помощью функции feature importance.

Статистическую обработку данных производили с использованием прикладной программы IBM SPSS Statistics 27. Цифровой код написан на языке программирования Python.

Результаты

В ходе исследования с использованием одного из алгоритмов технологии ИИ были выявлены факторы риска, которые наиболее тесно связаны с вероятностью возникновения сердечно-сосудистых заболеваний у здоровых людей. Также были разработаны и протестированы три модели ИИ.

В частности, для Самарской когорты градиентный бустинг показал результаты значения ROC-AUC 0,564, для Чуйской когорты – 0,806, для всей когорты «ИНТЕРЭПИД» – 0,766, $p < 0,05$ (рис. 2).

Низкие результаты классификации, полученные на Самарской когорте, можно объяснить более высокой несбалансированностью классов в данной выборке и более высокими показателями мультиколлинеарности ($VIF > 10$ наблюдался у 131 (85%) переменной). Наибольшие значения VIF имели

```
In [22]: param_grid = {
    "learning_rate": [0.01, 0.02, 0.03],
    "min_samples_split": [2],
    "min_samples_leaf": [6],
    "subsample": [0.5, 0.618, 0.8, 0.85, 0.9, 0.95, 1.0]
}

cv_gbc = GridSearchCV(gbc, param_grid=param_grid, refit=False, n_jobs=-1, cv=2)
cv_gbc.fit(x, y)
cv_gbc.best_params_

Out[22]: {'learning_rate': 0.01,
 'min_samples_leaf': 6,
 'min_samples_split': 2,
 'subsample': 0.618}
```

Рис. 1. Решетка гиперпараметров для алгоритма ИИ GradientBoostingClassifier

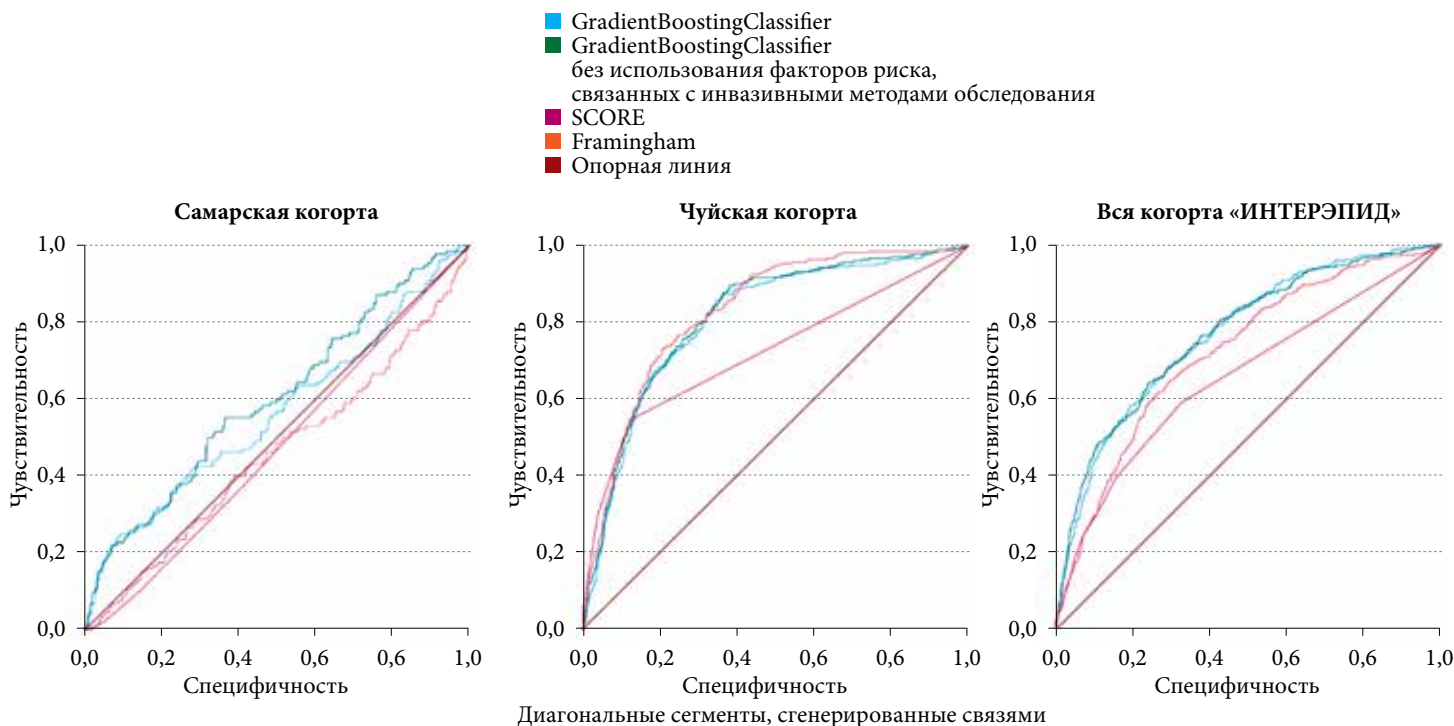


Рис. 2. ROC-AUC анализ GradientBoostingClassifier, шкал Framingham и SCORE для Самарской, Чуйской и всей когорты «ИНТЕРЭПИД»



переменные, кодирующие значения АД, а также потребление различных продуктов питания.

Для Кыргызской когорты наиболее весомыми факторами оказались возраст, ответы на утверждение: «Я стал все делать очень медленно» по шкале тревоги и депрессии и частота потребления пациентами колбас, сосисок и субпродуктов. Затем шли параметры значений АД, данных об избыточной массе тела и также об особенностях питания.

В тройку наиболее значимых факторов для Самарской когорты вошли: наличие болезней сердечно-сосудистой системы, уровень креатинина и липопротеидов высокой плотности (ЛПВП) в плазме крови. Значимую ассоциацию продемонстрировали также параметры физической активности, антропометрические показатели, злоупотребление алкоголем и оценка чувства страха и общего состояния дел по шкале тревоги и депрессии.

Наиболее значимым показателем при анализе всей когорты «ИНТЕРЭПИД» оказался возраст. На втором и третьем месте оказались уровень С-реактивного белка и факт потребления животного жира. При этом такие традиционные факторы риска, как пол и статус курения, не попали в первые 30 факторов, а показатель систолического артериального давления (САД) оказался на 24-м месте. 90% факторов риска являлись модифицируемыми и условно модифицируемыми (рис. 3).

Из всех 30 наиболее важных факторов риска, полученных на трех когортах, включая общую выборку, 10% были связаны с показателями АД, 33,3% – с особенностями питания и ассоциированными с ними факторами риска, 10% – с показателями ментального здоровья, 6,6% – с возрастом, 3,3% – с физической активностью, 3,3% – с вредными привычками, 23,3% – с лабораторными показателями, включая показатели липидов плазмы крови, и 10% – с данными анамнеза. Аналогичный анализ проведен при исключении факторов риска, связанных с инвазивными методами обследования. К таким факторам относятся:

- значения уровня ОХС;
- уровень ЛПВП в плазме крови;
- уровень триглицеридов плазмы крови;
- уровень глюкозы плазмы крови;
- уровень креатинина плазмы крови;
- уровень С-реактивного белка плазмы крови;
- показатели спирометрии.

По данным С-статистики, средние значения ROC-AUC для Самарской когорты увеличились на 6,4% и составили 0,6. Для Чуйской когорты показатели возросли на 0,9% и составили 0,813. Для всей когорты значения не изменились и составили 0,767, $p < 0,05$ (см. рис. 2). Аномальное повышение показателей ROC-AUC для Самарской когорты можно объяснить большим процентом пропущенных значений среди исключенных переменных (22,3% по сравнению с 0,03% у Чуйской когорты). Таким образом, исключение этих переменных повысило качество обучения, так как переменные с большим количеством пропусков создавали шум в данных.

При этом, несмотря на удовлетворительное качество данных, исключение факторов риска, требующих инвазивных



Примечание. ЛПВП – липопротеиды высокой плотности, ОХС – общий холестерин плазмы крови, ФЖЕЛ – форсированная жизненная емкость легких.

Рис. 3. Ранжирование по Feature Importance первых десяти наиболее значимых переменных для модели ИИ с использованием всех факторов риска для всей когорты «ИНТЕРЭПИД»



Примечание. ССЗ – сердечно-сосудистые заболевания, АД – артериальное давление, САД – систолическое артериальное давление.

Рис. 4. Ранжирование по вкладу в прогноз первых десяти наиболее значимых переменных для модели ИИ без использования инвазивных факторов риска для всей когорты «ИНТЕРЭПИД»

методов обследования, не привело к значительным изменениям в классификационных качествах модели.

Для Самарской когорты в тройку наиболее значимых факторов вошли:

- наличие ССЗ;
- оценка своего состояния по шкале тревоги и депрессии;
- частота сердечных сокращений.

В то время как для Чуйской когорты наиболее важными оказались два фактора:

- возраст;
- АД.

Для всей когорты «ИНТЕРЭПИД» наиболее значимыми оказались:

- возраст;
- заболевания почек в анамнезе;
- высокое АД.

В целом комбинация факторов риска сохранилась. 77% факторов риска являлись модифицируемыми и условно модифицируемыми (рис. 4). Из всех 30 наиболее важных факторов риска, выявленных в трех когортах, включая общую выборку, 30% были связаны с показателями АД, 10% – с особенностями питания и ассоциированными с ними факторами риска, 3,3% – с показателями ментального здоровья,



6,6% – с возрастом, 10% – с социальным статусом, 6,6% – с вредными привычками и 33,3% – с данными анамнеза. Чтобы сравнить качество оценки сердечно-сосудистого риска между ИИ (который не учитывает инвазивные факторы) и традиционными шкалами, мы сопоставили их с прогностическими возможностями шкал SCORE и Framingham. Для оценки риска с помощью SCORE и Framingham расчет производился на основе пяти предикторов:

- пол;
- возраст;
- статус курения;
- уровень ОХС (ммоль/л);
- показатели среднего САД (мм рт. ст.).

Хотя в Самарской когорте значения ROC-AUC для алгоритма GradientBoostingClassifier были в среднем на 18% выше, чем у традиционных шкал, разница не была статистически значимой. В Чуйской когорте наивысший уровень ROC-AUC показала шкала Framingham – 0,828, а градиентный бустинг занял второе место. Для всей когорты «ИНТЕРЭПИД» значения ROC-AUC алгоритма ИИ оказались на 6,4% выше, чем у шкалы Framingham, и на 16,9% выше, чем у шкалы SCORE, при уровне значимости $p < 0,05$ (см. рис. 2).

Обсуждение

В настоящем исследовании мы достигли высоких результатов в области классификации с использованием одного из алгоритмов ИИ – градиентного бустинга. Результаты были получены как при применении всех доступных предикторов, так и после исключения факторов, связанных с инвазивными методами обследования.

Кроме того, исследование показало, что факторы, ассоциированные с прогнозом ССЗ, могут отличаться в зависимости от конкретной группы пациентов, на которой проводился анализ.

В ранее проведенной нами работе мы также сравнили возможности использования градиентного бустинга в прогнозе риска ССЗ [14]. Изначально в исследовании использовались факторы риска:

- пол;
- возраст;
- индекс массы тела;
- частота сердечных сокращений;
- САД;
- ДАД;
- данные о наличии отягощенной наследственности по ССЗ у близких родственников;
- ОХС;
- показатели глюкозы в плазме крови.

На данных предикторах AUC модели составил 0,83–0,84. Затем были исключены лабораторные показатели, и AUC модели вырос до 0,85, что согласуется с результатами текущего исследования. В подобном исследовании 2008 г. авторы использовали оригинальную когорту Фрамингемского исследования для построения двух моделей прогноза сердечно-сосудистого риска с использованием и без использования показателей

липидов плазмы крови на основе логистической регрессии [15]. В первую модель были включены возраст, ОХС, ЛПВП, САД, курение, антигипертензивная терапия и наличие СД 2-го типа. Во второй модели был исключен липидный статус.

По данным С-статистики, показатели ROC-AUC для первой модели составили:

- для женщин – 0,763;
- для мужчин – 0,793.

Без учета липидных показателей значения ROC-AUC оказались следующими:

- для женщин – 0,785;
- для мужчин – 0,749.

Эти показатели в среднем лишь на 2,9% ниже, чем результаты, полученные с использованием липидных значений и нашей безлипидной модели.

В исследовании Stella Aslibekyan и соавт. [16] использовали факторы, связанные с питанием, снижением физической активности, курением, индексом талии и бедер, злоупотреблением алкоголем и низким социально-экономическим статусом, для прогноза риска развития ИМ на основании логистической регрессии. Показатели ROC-AUC составили от 0,63 до 0,67, что на 14,5% ниже, чем у нашей модели без липидов.

По мнению авторов представленных работ, неинвазивные методы оценки риска являются предпочтительными для первичного скрининга, особенно в условиях ограниченных ресурсов, где анализ крови может быть недоступен по различным причинам. Однако анализ, включающий показатели липидов плазмы крови, как правило, обладает более высокой прогностической точностью. Таким образом, выбор шкалы зависит от конкретных клинических условий и доступных данных.

Если рассматривать важность факторов риска, влияющих на прогноз ССЗ, полученных с помощью использования ИИ, то на сегодняшний день существует множество научных исследований, подтверждающих связь между особенностями питания и риском развития ССЗ. Эти связи были выявлены и в нашем исследовании.

Например, исследование, проведенное на выборке из 17 175 человек в России, показало, что ежедневное употребление творога, сыра и красного мяса связано с лучшей выживаемостью до развития ССЗ. С другой стороны, избыточное потребление соли и привычка досаливать пищу увеличивают риск сердечно-сосудистых осложнений [17].

Работа [18] представляет собой обзор различных систем питания, включая средиземноморскую, скандинавскую и диету DASH. Было установлено, что эти диеты, богатые растительными продуктами, снижают факторы риска, связанные с ССЗ, в частности уровень холестерина липопротеинов низкой плотности в крови.

В ноябре 2024 г. в «Российском медицинском журнале» было опубликовано исследование, в котором изучались факторы риска, связанные с нерациональным питанием. Исследование показало, что отношение окружности талии к росту может быть более точным показателем риска ССЗ, чем другие антропометрические параметры. Было установлено, что у участников с индексом талия/рост 0,5 и более



вероятность развития ССЗ была в 3,1 раза выше, чем у тех, у кого этот показатель был ниже 0,5 [19].

Также существует значительное количество научных исследований, посвященных изучению связи между психическим здоровьем и риском развития ССЗ.

Исследование [20] выявило, что повышенная тревожность увеличивает вероятность фатального ИМ в 1,9 раза и внезапной смерти в 4,5 раза. Авторы подчеркивают необходимость коррекции психосоциальных факторов, включая тревожную симптоматику, в терапии пациентов с ССЗ.

Метаанализ, объединивший 20 исследований, показал, что тревожные пациенты имеют повышенный риск развития ИБС и смерти от ССЗ независимо от других факторов риска [21].

Обзор [22] посвящен как положительным, так и отрицательным влияниям физической активности на профилактику ССЗ. Авторы анализируют данные многочисленных исследований, демонстрирующие, что регулярная физическая активность способствует снижению общей и сердечно-сосудистой смертности. Однако чрезмерные или неправильно подобранные нагрузки могут иметь негативные последствия для сердечно-сосудистой системы. Статья подчеркивает важность индивидуального подхода к назначению физических упражнений для профилактики ССЗ.

Многочисленные исследования также подтвердили влияние традиционных факторов риска, таких как артериальная гипертензия, коморбидность,

отягощенная наследственность и чрезмерное употребление алкоголя [23].

Исходя из полученных нами данных, можно сказать, что прогностические возможности шкал зависят:

- от качества данных, на которых они были созданы;
- комбинации определенных факторов риска, включая показатели мультиколлинеарности;
- типа математической модели, которая была использована для их построения.

Наше исследование и другие подобные работы показали, что добавление или исключение факторов риска, определяемых инвазивными методами диагностики, не оказывает существенного влияния на классификационные качества модели. Эти факторы могут быть полезны только в определенных клинических ситуациях и служат лишь дополнительным инструментом.

Заключение

Результаты исследования показали, что алгоритмы ИИ потенциально могут успешно заменить традиционные шкалы, даже при отсутствии лабораторных данных. Это открывает перспективы для применения ИИ в первичной профилактике ССЗ, особенно в регионах с ограниченным доступом к медицинским услугам. ☺

Финансирование. Авторы заявляют,

что исследование проведено без спонсорской поддержки.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Литература

1. Федеральная служба государственной статистики. Здравоохранение.
2. Полупина Н.В., Пивоваров Ю.П., Милушкина О.Ю. Профилактическая медицина – основа сохранения здоровья населения. Вестник Российского государственного медицинского университета. 2018; (5): 5–13.
3. Драпкина О.М., Шальнова С.А., Калинина А.М. и др. Профилактика основных хронических неинфекционных заболеваний и формирование здорового образа жизни. Учебно-методическое пособие. 2024.
4. Концевая А.В., Драпкина О.М. Экономика профилактики неинфекционных заболеваний. Профилактическая медицина. 2018; 21 (2): 4–10.
5. Мишкин И.А., Гусев А.В., Концевая А.В. и др. Эффективность использования mHealth в качестве инструмента профилактики сердечно-сосудистых заболеваний. Систематический обзор. Врач и информационные технологии. 2022; 4: 12–27.
6. Global strategy on digital health 2020–2025.
7. Честнов О.П., Бойцов С.А., Куликов А.А. и др. Мобильное здравоохранение: мировой опыт и перспективы. Профилактическая медицина. 2014; 17 (4): 3–9.
8. Мырзаматова А.О., Концевая А.В., Полупанов А.Г. и др. Результаты 7-летнего проспективного наблюдения в исследовании «ИНТЕРЭПИД»: факторы, влияющие на общую и сердечно-сосудистую смертность сельских жителей России и Кыргызской Республики. Российский кардиологический журнал. 2022; 27 (5): 38–48.
9. Концевая А.В., Мырзаматова А.О., Халматов А.Н. и др. Результаты 4-летнего проспективного наблюдения в исследовании «ИНТЕРЭПИД»: факторы, влияющие на заболеваемость и смертность популяции в сельских регионах России и Кыргызской Республики. Кардиоваскулярная терапия и профилактика. 2018; 17 (2): 49–56.
10. Мишкин И.А., Концевая А.В., Гусев А.В. и др. Прогнозирование сердечно-сосудистых событий с помощью моделей пропорциональных рисков и моделей машинного обучения: систематический обзор. Современные проблемы здравоохранения и медицинской статистики. 2023; 2: 804–829.
11. Мишкин И.А., Концевая А.В., Гусев А.В. и др. Разработка и тестирование новых методических подходов прогнозирования сердечно-сосудистых событий у здоровых людей с использованием технологии машинного обучения на базе международного исследования «ИНТЕРЭПИД». Профилактическая медицина. 2024; 27 (3): 72–79.
12. Мишкин И.А. и др. Сравнение прогностических возможностей традиционных методов оценки сердечно-сосудистого риска с использованием шкал Score и Framingham, технологий машинного обучения «ИНТЕРЭПИД». Профилактическая медицина. М.: Медиа Сфера. 2024; 27 (2): 96–102.



13. Мишкин И.А., Концевая А.В., Гусев А.В. и др. Разработка и тестирование новых методических подходов прогнозирования 6-месячной постгоспитальной летальности у пациентов, перенесших инфаркт миокарда, с использованием технологии машинного обучения на базе международного исследования «Острый инфаркт миокарда в Российской Федерации: существующая практика и препятствия на пути эффективного лечения на разных уровнях системы здравоохранения». Современные проблемы здравоохранения и медицинской статистики. 2023; 4: 936–955.
14. Мишкин И.А. Сахаров А.А. Изучение использования алгоритмов машинного обучения в оценке риска развития сердечно-сосудистых заболеваний. Инновации. Наука. Образование. 2020; 17: 187–202.
15. D'Agostino R.B.Sr., Vasan R.S., Pencina M.J., et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation. 2008; 117 (6): 743–753.
16. Aslibekyan S., Campos H., Loucks E.B., et al. Development of a cardiovascular risk score for use in low- and middle-income countries. J. Nutr. 2011; 141 (7): 1375–1380.
17. Карамнова Н.С., Капустина А.В., Куценко В.А. и др. Характер питания и риск развития сердечно-сосудистых осложнений. Результаты 6-летнего проспективного наблюдения за когортой исследования ЭССЕ-РФ. Российский кардиологический журнал. 2024; 29 (9): 5999.
18. Барбараш О.Л., Шибанова И.А., Помешкина С.А. и др. Роль питания в профилактике сердечно-сосудистых заболеваний. Доктор.Ру. 2019; 10 (165); 6–10.
19. Викторова И.А., Моисеева М.В., Стасенко В.Л. и др. Взаимосвязь между индексом талия/рост и сердечно-сосудистыми заболеваниями по данным исследования ЭССЕ-РФ2 в Омском регионе. РМЖ. Медицинское обозрение. 2024; 8 (9): 508–512.
20. Драпкина О.М., Федин А.И., Дорофеева О.А. и др. Влияние психосоциальных факторов риска на течение и прогноз сердечно-сосудистых заболеваний. Кардиоваскулярная терапия и профилактика. 2022; 21 (5): 64–70.
21. Акимова Е.В., Каюмова М.М., Бессонова М.И. Психосоциальные факторы риска развития сердечно-сосудистых заболеваний в женских популяциях с позиции исследований тревоги. Профилактическая медицина. 2022; 25 (7): 97–102.
22. Саргсян В.Д. Физическая нагрузка в профилактике заболеваний сердечно-сосудистой системы: две стороны одной медали. Профилактическая медицина. 2023; 26 (10): 117–122.
23. Драпкина О.М., Концевая А.В., Калинина А.М. и др. Коморбидность пациентов с хроническими неинфекционными заболеваниями в практике врача-терапевта. Евразийское руководство. Кардиоваскулярная терапия и профилактика. 2024; 23 (3): 113–118.

The Possibilities of Assessing Cardiovascular Risk without Predictors That Require Invasive Detection Methods with the Use of Artificial Intelligence

I.A. Mishkin^{1,2}, A.V. Koncevaya¹, A.A. Saharov³, A.G. Polupanov⁴, O.M. Drapkina¹

¹ Medical Research Center for Therapy and Preventive Medicine of the Ministry of Health of the Russian Federation

² The State Healthcare Institution of the Tula region «Kireevskaya Central District Hospital»

³ Renaissance Insurance Group PJSC

⁴ Kyrgyz State Medical Academy named after I.K. Akhunbaev

Contact person: Ilya A. Mishkin, Ilya.mischkin@yandex.ru

Cardiovascular diseases (CVD) remain the leading cause of death in the Russian Federation, despite the fact that advances in preventive medicine and mHealth contribute to reducing the burden of these diseases. The combined use of artificial intelligence (AI), wearable devices, and noninvasive risk assessment methods makes it possible to use these technologies for mass screening and early detection of high-risk groups.

Purpose was to study and compare the accuracy of the cardiovascular risk assessment with and without the use of risk factors associated with invasive detection methods in the population of the Russian Federation and Kyrgyzstan regions with the use of an AI model.

Material and methods. The analysis was based on data from the INTEREPID study (2011–2016), which included 2,392 patients from the Samara Region (RF) and the Chui region (Kyrgyzstan). The AI algorithm Gradient Boosting Classifier with 5-fold cross-validation was used. The main metric of the model's quality was ROC-AUC.

Results. The effectiveness of using gradient boosting to determine the risk of developing CVD was analyzed. The constructed models demonstrated a ROC-AUC from 0.564 (Samara cohort) to 0.806 (Kyrgyz Republic cohort), in the general cohort – 0.766. In the Samara cohort, the key predictors were the presence of CVD in the anamnesis, the level of creatinine and high-density lipoproteins (HDL), in the Kyrgyz – age, anxiety level, sausage consumption and overweight. The exclusion of invasive variables improved the quality of the models: ROC-AUC increased by 6.4% in the Samara cohort, 0.9% in the Kyrgyz Republic cohort, but did not change in the general cohort. The AI prediction surpassed the traditional SCORE and Framingham scales, showing an advantage of 6.4% over Framingham and 16.9% over SCORE in the general cohort ($p < 0.05$).

Conclusions. The data obtained demonstrate that AI algorithms can potentially replace traditional scales in the absence of laboratory data. This opens up prospects for the use of AI in primary prevention of CVD in conditions of limited access to medical resources.

Keywords: cardiovascular risk, risk assessment, artificial intelligence, machine learning