

Прогнозирование наличия субклинического каротидного атеросклероза у пациентов с избыточным весом и ожирением при помощи модели машинного обучения

Гаврилов Д. В.¹, Кузнецова Т. Ю.², Дружилов М. А.², Корсаков И. Н.¹, Гусев А. В.^{1,3,4}

Цель. Разработать модель прогнозирования наличия субклинического каротидного атеросклероза (СКА) с целью уточнения сердечно-сосудистого риска (ССР) при помощи методов машинного обучения у пациентов с избыточным весом и ожирением без артериальной гипертензии, сахарного диабета и/или сердечно-сосудистых заболеваний (ССЗ).

Материал и методы. Использована база обезличенных данных (БД) Webiomed (2,9 млн пациентов). Критерии включения: возраст ≥ 18 лет, индекс массы тела ≥ 25 кг/м², наличие результатов выполненного ультразвукового исследования брахиоцефальных артерий (БЦА). Исключались из анализа пациенты с артериальной гипертензией, сахарным диабетом и/или ССЗ. Отобраны данные о 5750 пациентах, из которых атеросклеротические бляшки БЦА выявлены у 385 человек. Окончательный набор данных (НД) содержал сведения о 447 пациентах, у 197 (44,1%) из них был выявлен СКА. Количественные и категориальные признаки для обучения модели взяты с заполненностью в БД $\geq 40\%$, число окончательных признаков для машинного обучения составило 28. При создании модели использовались 3 алгоритма Random Forest, AdaBoostClassifier, KNeighborsClassifier и библиотека Scikit-learn. Для улучшения работоспособности модели применялась функция заполнения пропущенных значений. Целевыми параметрами модели были заданы предсказательная прогнозная способность (accuracy) не ниже 75%, площадь под ROC-кривой не $< 0,75$.

Результаты. Полученный НД был разделен на тренировочную и тестовую части в соотношении 80:20. В зависимости от примененных алгоритмов обученная модель характеризуется предсказательной способностью 75-97%, чувствительностью 77-92%, специфичностью 80-98%, площадью под ROC-кривой 0,88-0,97. С учетом метрик точности лучшие результаты были получены для модели, обученной алгоритмом Random Forest (95%, 92%, 98% и 0,95, соответственно).

Заключение. Разработанная модель может помочь врачу принимать решение о направлении пациента с избыточным весом и ожирением без ССЗ на ультразвуковое исследование БЦА, что способствует более точной стратификации ССР. Внедрение в практику таких алгоритмов риск-стратификации позволит увеличить точность и качество прогнозирования ССР и оптимизировать систему проводимых профилактических мероприятий.

Ключевые слова: ожирение, сердечно-сосудистый риск, субклинический каротидный атеросклероз, машинное обучение, искусственный интеллект.

Отношения и деятельность. Исследование выполнено на уникальной научной установке "Многокомпонентный программно-аппаратный комплекс для автоматизированного сбора, хранения, разметки научно-исследовательских и клинических биомедицинских данных, их унификации и анализа на базе ЦОД с ис-

пользованием технологий искусственного интеллекта" (регистрационный номер 2075518), при финансовой поддержке Министерства науки и высшего образования Российской Федерации в рамках Соглашения № 075-15-2021-665.

¹К-Скай, Петрозаводск; ²ФГБОУ ВО ПетрГУ, Петрозаводск; ³ФГБУ Центральный научно-исследовательский институт организации и информатизации здравоохранения Минздрава России, Москва; ⁴ГБУЗ Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы, Москва, Россия.

Гаврилов Д. В.* — руководитель медицинского направления, ORCID: 0000-0002-8745-857X, Кузнецова Т. Ю. — д.м.н., доцент, зав. кафедрой факультетской терапии, фтизиатрии, инфекционных болезней и эпидемиологии, ORCID: 0000-0002-6654-1382, Дружилов М. А. — к.м.н., доцент кафедры факультетской терапии, фтизиатрии, инфекционных болезней и эпидемиологии, ORCID: 0000-0002-3147-9056, Корсаков И. Н. — к.ф.-м.н., эксперт по машинному обучению, ORCID: 0000-0003-2343-9641, Гусев А. В. — к.т.н., директор по развитию, ORCID: 0000-0002-7380-8460.

*Автор, ответственный за переписку (Corresponding author):
dgavrilov@webiomed.ai

АБ — атеросклеротическая бляшка, БД — база данных, БЦА — брахиоцефальные артерии, ИМТ — индекс массы тела, НД — набор данных, СКА — субклинический каротидный атеросклероз, ССЗ — сердечно-сосудистое заболевание, ССР — сердечно-сосудистый риск, УЗИ — ультразвуковое исследование, ЭМК — электронные медицинские карты.

Рукопись получена 02.02.2022

Рецензия получена 03.02.2022

Принята к публикации 15.04.2022



Для цитирования: Гаврилов Д. В., Кузнецова Т. Ю., Дружилов М. А., Корсаков И. Н., Гусев А. В. Прогнозирование наличия субклинического каротидного атеросклероза у пациентов с избыточным весом и ожирением при помощи модели машинного обучения. *Российский кардиологический журнал*. 2022;27(4):4871. doi:10.15829/1560-4071-2022-4871. EDN CNXIPK

Predicting the subclinical carotid atherosclerosis in overweight and obese patients using a machine learning model

Gavrilov D. V.¹, Kuznetsova T. Yu.², Druzhilov M. A.², Korsakov I. N.¹, Gusev A. V.^{1,3,4}

Aim. To develop a model for predicting the subclinical carotid atherosclerosis (SCA) in order to refine cardiovascular risk (CVR) using machine learning methods in overweight and obese patients without hypertension, diabetes and/or cardiovascular disease (CVD).

Material and methods. Anonymized database (DB) Webiomed (2.9 million patients) was used. There were following inclusion criteria: age ≥ 18 years, body mass index ≥ 25 kg/m², availability of data on ultrasound of extracranial arteries. Patients with hypertension, diabetes and/or CVD were excluded from the analysis. Data on 5750 patients were selected, of which atherosclerotic plaques were detected in 385 people. The final data set contained information on 447 patients, 197 (44,1%) of which had SCA. Quantitative and categorical traits for model training were taken with $\geq 40\%$ occupancy

in the database. The number of final traits for machine learning was 28. When creating the model, 3 Random Forest algorithms, AdaBoostClassifier, KNeighborsClassifier and the Scikit-learn library were used. To improve the model performance, the fill missing function was used. The target parameters of the model were given a predictive ability (accuracy) of at least 75%, while the area under the ROC curve was at least 0,75.

Results. The resulting dataset was divided into training and test parts in a ratio of 80:20. Depending on the applied algorithms, the learned model was characterized by a predictive ability of 75-97%, sensitivity of 77-92%, specificity of 80-98%, and area under the ROC-curve of 0,88-0,97. Taking into account the accuracy metrics, the best results were obtained for the model learned by the Random Forest algorithm (95%, 92%, 98% and 0,95, respectively).

Conclusion. The developed model can help a physician make a decision to refer an overweight and obese patient without cardiovascular diseases for ultrasound of extracranial arteries, which contributes to a more accurate CVR stratification. The introduction of such risk stratification algorithms into practice will increase the accuracy and quality of CVR prediction and optimize the system of preventive measures.

Keywords: obesity, cardiovascular risk, subclinical carotid atherosclerosis, machine learning, artificial intelligence.

Relationships and Activities. The study was carried out on a unique scientific installation “Multicomponent software and hardware system for automated collection, storage, markup of research and clinical biomedical data, their unification and analysis based on Data Center with Artificial Intelligence technologies” (№ 2075518) and financially supported by Ministry of Science and Higher Education of the Russian Federation within the Agreement № 075-15-2021-665.

Наблюдаемое на протяжении последних десятилетий значительное увеличение распространенности избыточного веса и ожирения среди населения подавляющего большинства стран мира [1] определяет одной из основных задач системы здравоохранения необходимость улучшения стратификации сердечно-сосудистого риска (ССР) и интенсификации профилактических мероприятий для предупреждения развития и прогрессирования сердечно-сосудистых заболеваний (ССЗ) [2].

Невысокая предсказательная способность различных шкал-рискометров, в первую очередь, у лиц с исходно низким или умеренным ССР, в основном, связана с ограниченным числом используемых параметров в качестве предикторов сердечно-сосудистого события, поэтому требуется реклассификация риска с помощью дополнительных методов исследования [3].

При использовании традиционных шкал-рискометров для оценки ССР пациенты с избыточным весом и ожирением относятся чаще к “невысокой” категории риска. При этом, известно, что избыточный вес и ожирение отнесены к факторам, реклассифицирующим величину исходно определенного риска [4]. Кроме того, взаимосвязь ожирения и ассоциированных с ним заболеваний и их осложнений демонстрирует наличие феномена гетерогенности, проявляющегося в различных “метаболических фенотипах” ожирения, отличающихся величиной ССР [5].

Одним из основных подходов к выделению фенотипа ожирения с высоким ССР является прямая визуализация абдоминальной и/или эктопической висцеральной жировой ткани с помощью ультразвуковых (УЗИ) и томографических исследований, существенным образом увеличивающая чувствительность и специфичность верификации висцерального ожирения [5], однако данные методики не являются широкодоступными в клинической практике.

Альтернативным направлением совершенствования прогнозирования ССР у данной категории может стать оптимизация диагностических алгоритмов по выявлению субклинического каротидного атеросклероза (СКА), являющегося отражением влияния

¹K-Sky, Petrozavodsk; ²Petrozavodsk State University, Petrozavodsk; ³Central Research Institute for Health Organization and Informatics, Moscow; ⁴Scientific and Practical Clinical Center for Diagnostics and Telemedicine Technologies, Moscow, Russia.

Gavrilov D. V.* ORCID: 0000-0002-8745-857X, Kuznetsova T. Yu. ORCID: 0000-0002-6654-1382, Druzhilov M. A. ORCID: 0000-0002-3147-9056, Korsakov I. N. ORCID: 0000-0003-2343-9641, Gusev A. V. ORCID: 0000-0002-7380-8460.

*Corresponding author: dgavrilov@webiomed.ai

Received: 02.02.2022 **Revision Received:** 03.02.2022 **Accepted:** 15.04.2022

For citation: Gavrilov D. V., Kuznetsova T. Yu., Druzhilov M. A., Korsakov I. N., Gusev A. V. Predicting the subclinical carotid atherosclerosis in overweight and obese patients using a machine learning model. *Russian Journal of Cardiology*. 2022;27(4):4871. doi:10.15829/1560-4071-2022-4871. EDN CNXIPK

на сосудистую стенку всего спектра неблагоприятных факторов при избыточном весе и ожирении [3]. По данным многочисленных срезовых и проспективных исследований выявление СКА имеет высокое прогностическое значение в отношении риска развития сердечно-сосудистых осложнений, сопоставимое с оценкой коронарного кальциевого индекса [6-8].

В свою очередь, данные диагностические алгоритмы должны быть основаны на различных моделях оценки вероятности выявления СКА, создаваемых при анализе совокупности доступных антропометрических, клинических и лабораторных параметров, определяющих высокую вероятность получения положительного результата. В качестве примеров таких прогнозных моделей у асимптомных в отношении ССЗ пациентов с избыточным весом и ожирением с исходно “невысоким” риском по шкале SCORE можно привести регрессионные уравнения, полученные в результате проведения бинарного многофакторного логистического регрессионного анализа данных [9, 10].

Так, в работе Шенковой Н. Н. и др. предикторами вероятности выявления СКА выступали уровни грелина, лептина и С-реактивного белка крови и эхокардиографическая толщина эпикардальной жировой ткани с предсказательной точностью модели 89,7% [9], в ранее проведенном нами исследовании — среднесуточные скорость пульсовой волны и систолическое артериальное давление в аорте, уровни гликемии натощак и мочевого кислоты крови с общим процентом верных предсказаний 91,7% [10]. Вместе с тем определение указанных предикторов также сопряжено с проблемой доступности данных методов, что определяет актуальность создания и внедрения в клиническую практику более простых, доступных в использовании и более точных моделей прогноза.

В настоящее время подобные модели создаются при помощи машинного обучения и обработки больших данных, что позволяет существенно упростить и одновременно улучшить систему стратификации риска, реализовав тем самым переход к персонализированной медицине и высокотехнологичному здравоохранению [11].

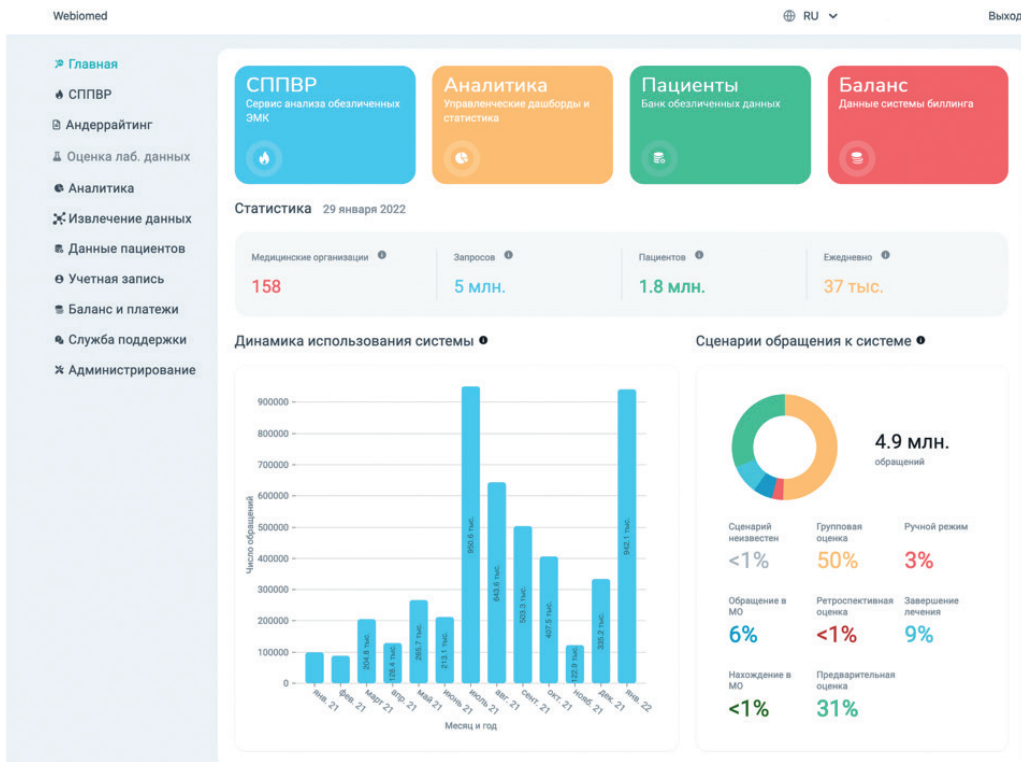


Рис. 1. Общий вид платформы Webiomed.

По данным различных авторов создаваемые таким способом прогнозные модели существенно превосходят существующие алгоритмы и шкалы в точности оценки наступления того или иного события [12, 13]. Методы машинного обучения позволяют создавать прогнозные модели, способные включать в качестве предикторов неограниченное количество признаков, что повышает значимость ведения практикующими врачами электронных медицинских карт (ЭМК) и обеспечивает комплексный подход к оценке ССР [11].

Целью данного исследования стала разработка модели прогнозирования вероятности наличия СКА при помощи методов машинного обучения системы у пациентов с избыточным весом и ожирением без артериальной гипертензии, сахарного диабета и/или ССЗ.

Материал и методы

Источником информации для создания модели прогноза стала база данных (БД) платформы прогнозной аналитики Webiomed, содержащая деперсонифицированные формализованные данные ЭМК 2,9 млн пациентов, проходивших обследование и лечение в медицинских организациях различных регионов Российской Федерации. Для сбора данных компанией-разработчиком платформы Webiomed были подписаны соглашения с соответствующими операторами персональных медицинских данных на их обезличивание на стороне оператора и передачу результатов такой обработки данных для анализа в платформу Webiomed, в т.ч. для научно-иссле-

довательских целей. Поскольку анализировались обезличенные медицинские данные, информированное добровольное согласие не использовалось. Медицинские данные и выявляемые признаки получены рефрактивно из ЭМК с помощью различных технологий извлечения информации из “сырых” данных, в т.ч. NLP (natural language processing). БД содержит >2600 категориальных и количественных признаков. Общий вид и детализация платформы Webiomed представлены на рисунках 1 и 2.

В качестве критериев отбора информации о пациентах для последующего анализа стали возраст ≥ 18 лет, индекс массы тела (ИМТ) ≥ 25 кг/м², наличие результатов выполненного УЗИ брахиоцефальных артерий (БЦА). Исключались из анализа пациенты с артериальной гипертензией, сахарным диабетом и/или ССЗ.

В качестве возможных компонентов модели были отобраны клинические и лабораторно-инструментальные признаки, изучавшиеся в ранее выполненных исследованиях, посвященных прогнозированию СКА [8-10], но с учетом их доступности для определения в широкой клинической практике. Окончательный набор признаков устанавливался совместно со специалистами по анализу данных на стадии определения значимости признаков для прогнозирования целевого события — атеросклеротической бляшки (АБ) БЦА.

Из общей БД платформы Webiomed была извлечена информация о 6438 пациентах, у которых имелись результаты УЗИ БЦА. Критериям включения без критериев исключения соответствовали данные о 5750

Социальная информация

ID пациента: 3356811. 40 лет

Hash №1
e12f2e72f5f32dd603196d78f1b14f088c8aea8a12bb453d9b4282826ecf7fce

Hash №2
a57c679b869adfb8c1cebb08348dd676d51d11878f63ab11237978cc0c57f320

Hash №3
cf2e91de575e77ecc9adde5f98d6a822f80647e076f0ce68e9555cdafb066cf2

Hash №4
2c3ddb74a20c70d4db93f28361ae1fb87051be72da413a9167b5c28e97992c4

[Подробнее](#)

Анамнез и сигнальная информация

Рост 156.0 см 01.04.19	Вес 59.0 кг 01.04.19	АД 120.0/80.0 29.03.19	ЧСС —	Глюкоза 4.81 ммоль/л 29.03.19	Холестерин 4.47 ммоль/л 29.03.19	Эритроциты —
ИМТ 24.2	Талия 76.0 см 29.03.19	ЧД —	Группа крови —	Лейкоциты —	Тромбоциты —	
		Фенотип —	Резус —			

Курение
Нет
02.04.21

Злоупотребление алкоголем
—

Наркомания
—

Отягощенная наследственность
Нет
02.04.21

Лист окончательных диагнозов

Код МКБ-10	Формулировка	Дата регистрации	Дата обращения	Тип
O04.9	Медикаментозный аборт при беременности раннего срока. 3 рубца на матке.	07 июля 2021	07 июля 2021	Острое
N80.8	Другой эндометриоз	17 марта 2021	17 марта 2021	Ранее установленное хроническое
I82.8	Варикозное расширение вен нижних конечностей без язвы	11 марта 2021	11 марта 2021	Ранее установленное хроническое

[Показать ещё](#)

Заболевания пациента
Заболевания не указаны

Извлечённые признаки пациента

Значение	Дата регистрации	Документ	Развернуть все значения
Вес			Показать значения
Гастрит (источник: анкета диспансеризации)			Показать значения
Глюкоза крови (любое измерение)			Показать значения
Диастолическое АД			Показать значения
Нарушение сознания			Показать значения
Общий холестерин			Показать значения
Окружность талии			Показать значения
Отягощенная наследственность по инфаркту миокарда в возрасте до 60 лет у ближайших родственников			Показать значения
Пребывания в домах престарелых			Показать значения
Рациональное питание (источник: анкета диспансеризации)			Показать значения
Рост			Показать значения
Систолическое АД			Показать значения
Табакокурение			Показать значения
Уровень употребления алкоголя (источник: анкета диспансеризации)			Показать значения
Язвенная болезнь (источник: анкета диспансеризации)			Показать значения

[Свернуть](#)

Рис. 2. Обезличенные социальные, категориальные и количественные признаки пациента в БД Webiomed.

пациентах, из которых АБ БЦА были выявлены у 385 человек. Так как в сформированном наборе данных (НД) пациентов без АБ БЦА было гораздо больше, чем с АБ БЦА, то была применена функция балансировки данных: уравнивание количества пациентов с и без АБ БЦА, а также учет максимального количества признаков в соответствии с исходно запланированными для изучения предикторами. Окончательный НД содержит

данные о 447 пациентах, у 197 (44,1%) из которых был выявлен СКА. Этапы с критериями отбора пациентов и формирование окончательного НД для машинного обучения представлены в таблице 1.

Количественные и категориальные признаки взяты с заполненностью в БД $\geq 40\%$, в итоге был получен 51 клинико-социальный признак. Из данного списка признаков были исключены те, которые не встречались

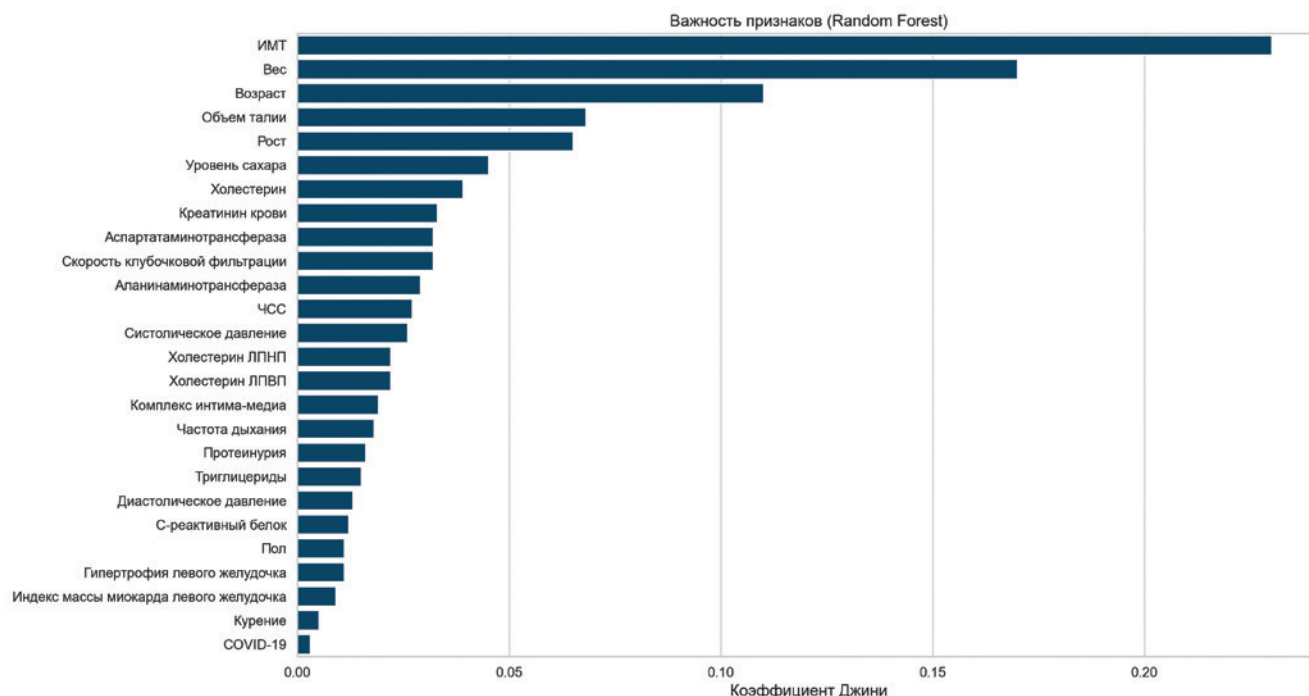


Рис. 3. Значимость входящих признаков при обучении модели алгоритмом Random Forest.

Сокращения: ИМТ — индекс массы тела, ЛПВП — липопротеины высокой плотности, ЛПНП — липопротеины низкой плотности, ЧСС — частота сердечных сокращений.

Таблица 1

Этапы формирования НД для машинного обучения

Критерии отбора	Число, n
Пациенты БД Webiomed	2915468
Из них пациенты, у которых имеются документы о выполненном УЗИ БЦА	7135
Из них уникальных пациентов	6438
Из них пациенты с критериями включения и без критериев исключения	5750
Из них пациенты с целевым событием — АБ БЦА	385
Окончательный НД после балансирования	447

Сокращения: АБ — атеросклеротическая бляшка, БД — база данных, БЦА — брахиоцефальные артерии, УЗИ — ультразвуковое исследование, НД — набор данных.

в техническом задании для обучения модели или частота их встречаемости была признана недостаточной. В окончательном НД часть признаков встречалась не в 100% случаев, поскольку в информации о пациентах с АБ БЦА отмечались пропуски некоторых значений признаков. Так как таких пациентов было не много, было решено не удалять их из НД, а пропущенные значения заменить средним значением, рассчитанным при анализе тренировочного набора для обучения модели. Количество окончательных признаков для машинного обучения составило 28. Их значения представлены в виде средней величины со стандартным отклонением и частотами, принимая во внимание нормальный характер их распределения. Описание окончательно отобранных признаков (n=28) представлено в таблице 2.

При создании модели использовались 3 алгоритма машинного обучения (Random Forest, AdaBoostClassifier, KNeighborsClassifier) и библиотека Scikit-learn. Для улучшения работоспособности модели применялась функция заполнения пропущенных значений. Целевыми параметрами модели были заданы предсказательная прогностическая способность (accuarcy) не <75%, площадь под ROC-кривой не <0,75.

Результаты

Для обучения модели НД был разделен на тренировочную и тестовую части в соотношении 80:20. В ходе машинного обучения была исследована функция перестановок признаков (permutation importance), ее применение не улучшило окончательный результат. В ходе обучения модели была определена значимость признаков, влияющая на прогнозируемое моделью целевое событие, которая представлена на рисунке 3. Согласно обучающему алгоритму определены самые важные признаки, влияющие на целевое событие: ИМТ, вес, возраст, окружность талии, гликемический профиль, общий холестерин и креатинин крови. Распределение значений признаков, наиболее повлиявших на выходное событие, представлено на рисунке 4.

Для обучения модели были использованы 3 алгоритма машинного обучения, применение других алгоритмов для решения этой задачи было расценено как избыточное.

Полученная модель оценки вероятности наличия СКА характеризуется различными параметрами прогностической значимости в зависимости от использу-

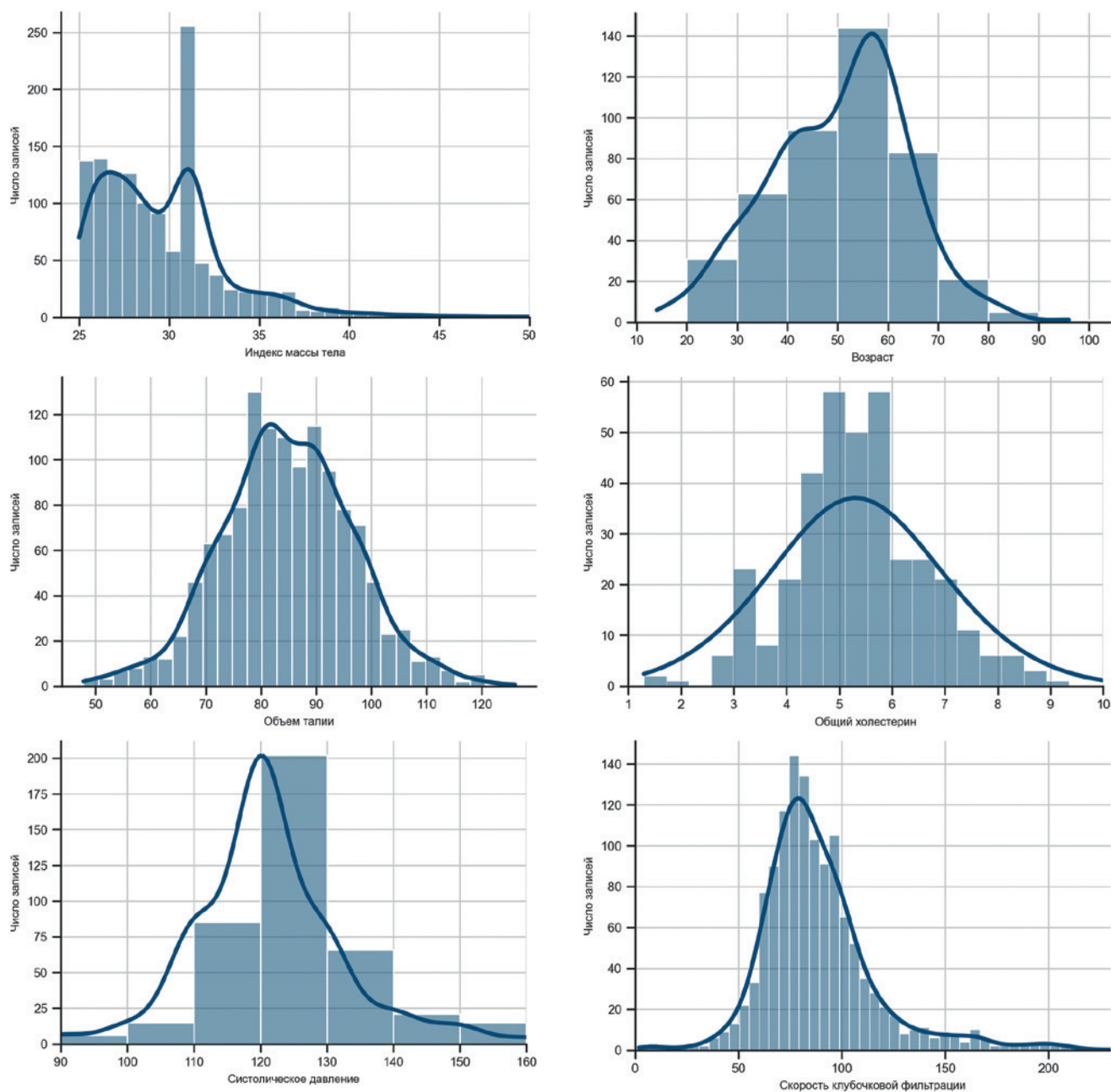


Рис. 4. Распределение значений признаков, рассчитанных алгоритмом как значимые: ИМТ, возраст, окружность талии, общий холестерин, систолическое артериальное давление, скорость клубочковой фильтрации.

емого алгоритма машинного обучения (табл. 3): предсказательная способность от 75% до 97%, чувствительность от 77% до 92%, специфичность от 80% до 98%, площадь под ROC-кривой от 0,88 до 0,97. С учетом полученных результатов точности модели для работы был выбран алгоритм Random Forest (рис. 5).

Таким образом, выходным признаком обученной модели является вероятность наличия АБ БЦА в виде числа от 0 до 1. Интерпретация выхода осуществляется с применением порогового значения 0,55, выходное значение больше которого оценивается как высокая вероятность наличия СКА, меньше — как низкая вероятность наличия СКА.

Обсуждение

Выявление СКА у пациентов с избыточным весом и ожирением, отнесенного по применяемым сегодня шкалам-рискометрам к категории “невысокого” ССР, позволяет провести реклассификацию риска и интенсифицировать проводимые профилактические мероприятия, в т.ч. инициировать назначение липидснижающей терапии [4]. Более того, непосредственная визуализация субклинического атеросклеротического заболевания дополнительно к представляемой пациенту информации о классических факторах риска существенно повышает его приверженность к соблюдению мероприятий по формированию здо-

Таблица 2
Характеристика признаков окончательного НД (n=447) для создания прогнозной модели

Признаки	Значения
Возраст, лет	49,7±13,4
Мужчины, %	29,5
Рост, см	163,3±10,3
Вес, кг	76,1±12,1
Индекс массы тела, кг/м ²	29,1±11,5
Окружность талии, см	85,9±11,1
Курение, %	13,4
Систолическое артериальное давление, мм рт.ст.	121,4±12,9
Диастолическое артериальное давление, мм рт.ст.	77,4±7,6
Частота сердечных сокращений, уд./мин	72,3±11,8
Частота дыхательных движений, в мин	17,7±6,7
Общий холестерин, ммоль/л	5,8±3,6
Холестерин липопротеинов низкой плотности, ммоль/л	3,6±1,1
Холестерин липопротеинов высокой плотности, ммоль/л	1,6±0,5
Триглицериды, ммоль/л	1,7±1,4
Аспарагиновая аминотрансфераза, ЕД/л	24,5±12,1
Аланиновая аминотрансфераза, ЕД/л	21,7±14,0
С-реактивный белок, мг/л	4,4±4,7
Протеинурия, г/л	0,12±0,7
Креатинин крови, мкмоль/л	86,4±125,7
Скорость клубочковой фильтрации, мл/мин	84,0±22,9
Гипертрофия левого желудочка, %	17,7
Индекс массы миокарда левого желудочка, г/м ²	88,2±25,6
Анамнез псориаза, %	1,1
Анамнез ревматоидного артрита, %	0,4
Анамнез системной красной волчанки, %	0
Анамнез подагры, %	0
Анамнез перенесенной COVID-19, %	3,8

рового образа жизни и приему лекарственных препаратов [14].

Учитывая феномен гетерогенности фенотипов ожирения в отношении ассоциированного риска [5], проявляющийся в т.ч. различной частотой выявления СКА [15], для повышения точности риск-стратификации и оптимизации алгоритмов направления асимптомного пациента для проведения УЗИ БЦА требуется внедрение в широкую клиническую практику доступных и точных прогнозных моделей.

Безусловно, если включать в качестве предикторов оценки вероятности СКА параметры, получаемые в ходе выполнения УЗИ, магнитно-резонансной и компьютерной томографии, анализов показателей артериальной жесткости, лабораторных маркеров воспаления, фиброза, нейрогуморальной активности висцеральной жировой ткани и других высоко специализированных методов обследования пациентов с избыточным весом и ожирением, возможно создание моделей с высокой прогностической способностью в результате классического многофакторного логистического регрессионного анализа [9, 10]. Однако реалии клинической практики, в первую очередь ее амбулаторного звена, не позволяют в рамках первичной профилактики провести в большинстве случаев комплекс

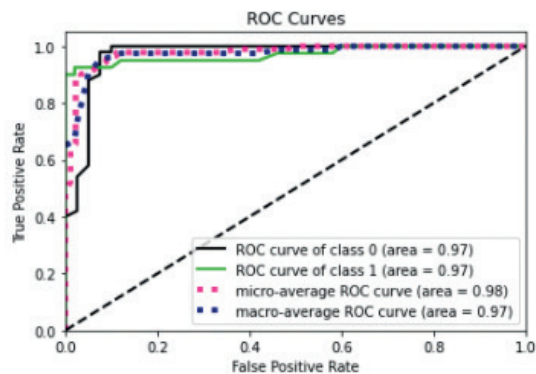


Рис. 5. Графики ROC-кривой для алгоритма Random Forest.

Таблица 3
Метрики для различных алгоритмов машинного обучения при создании модели оценки вероятности наличия СКА

Метрика	Алгоритм машинного обучения		
	Random Forest	AdaBoostClassifier	KNeighborsClassifier
Accuracy	0,95	0,90	0,78
Precision	0,97	0,89	0,75
Sensitivity	0,92	0,87	0,77
Specificity	0,98	0,92	0,80
F1	0,95	0,88	0,76
ROC-AUC	0,97	0,94	0,88
Порог отсечения	0,55	0,5	0,5

специализированных методов обследования, и данные прогнозные модели зачастую остаются предметом интереса научных исследований. Кроме того, их прогностическая способность не всегда проходит валидацию в независимой от исходной выборки.

Перспективным является моделирование вероятности того или иного события при помощи методов машинного обучения с использованием большого количества данных. Получаемые таким образом модели способны выполнять свою функцию и в случае ограниченного количества предикторов, отличающегося от исходного набора, используемого на этапе обучения и валидации, при незначительном снижении своей прогностической способности [11].

Нами была разработана модель прогнозирования вероятности наличия СКА у асимптомных в отношении ССЗ пациентов с избыточным весом и ожирением, которая включает доступные для определения в широкой клинической практике предикторы: возраст, пол, рост, вес, систолическое и диастолическое артериальное давление, частота сердечных сокращений и дыхания, параметры липидного обмена и функции почек, С-реактивный белок, уровни аланиновой и аспарагиновой трансаминаз, оценка гипертрофии миокарда левого желудочка, сопутствующие воспалительные и метаболические заболевания.

При анализе вклада различных параметров в развитие атеросклеротических изменений в сосудах бы-

ли выявлены наиболее значимые факторы. У пациентов без ССЗ и “не высокого” ССР именно ИМТ и другие параметры, характеризующие ожирение (рост, вес, объем талии), оказались лидирующими и в большей степени влияли на СКА по сравнению с показателями липидного спектра, курением, гликемическим профилем и другими анализируемыми маркерами риска. Можно высказать предположение, что применение данной модели в реальной клинической практике в значительной степени улучшит оценку ССР у пациентов с избыточным весом и ожирением.

Полученная модель характеризуется высокой предсказательной способностью: точность (accuracy) составила 95%, по ROC-AUC 97%.

Для практического применения необходимо учитывать ограничения, имеющиеся у модели, а также связанные с особенностями используемых данных. Конечный НД для обучения модели включал данные о 447 пациентах, что может рассматриваться недостаточным для описания генеральной совокупности. Однако с учетом того, что модель отвечает на “узкий” клинический вопрос, имеет много входящих признаков и для формирования НД была использована БД в 2,9 млн пациентов, данный набор признан достаточным для обучения. Модель предназначена для лиц с исходно “невысоким” ССР, в возрасте от 36 до 62 лет. Она работает при отсутствии некоторых значимых показателей (функция пропуска значений признаков), но точность работы модели при включении этой функции не проверялась и требует последующего анализа.

Литература/References

1. NCD Risk Factor Collaboration. Trends in adult body mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 128.9 million participants. *Lancet*. 2016;387(10026):1377-96. doi:10.1016/S0140-6736(16)30054-X.
2. The GBD 2015 Obesity Collaborators. Health Effects of Overweight and Obesity in 195 Countries over 25 Years. *N Engl J Med*. 2017;377(1):13-27. doi:10.1056/NEJMoa1614362.
3. Rossello X, Dorresteijn J, Janssen A, et al. Risk prediction tools in cardiovascular disease prevention: A report from the ESC Prevention of CVD Programme led by the European Association of Preventive Cardiology (EAPC) in collaboration with the Acute Cardiovascular Care Association (ACCA) and the Association of Cardiovascular Nursing and Allied Professions (ACNAP). *Eur J Prev Cardiol*. 2019;26(14):1534-44. doi:10.1177/2047487319846715.
4. Mach F, Baigent C, Catapano A, et al. 2019 ESC/EAS Guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk. *Eur Heart J*. 2020;41(1):111-88. doi:10.1093/eurheartj/ehz455.
5. Druzhilova MA, Kuznetsova TY. Heterogeneity of obesity phenotypes in relation to cardiovascular risk. *Cardiovascular Therapy and Prevention*. 2019;18(1):161-7. (In Russ.) Дружилова М.А., Кузнецова Т.Ю. Гетерогенность фенотипов ожирения в отношении сердечно-сосудистого риска. Кардиоваскулярная терапия и профилактика. 2019;18(1):161-7. doi:10.15829/1728-8800-2019-1-162-168.
6. Baber U, Mehran R, Sartori S, et al. Prevalence, impact, and predictive value of detecting subclinical coronary and carotid atherosclerosis in asymptomatic adults: the BiImage study. *J Am Coll Cardiol*. 2015;65(11):1065-74. doi:10.1016/j.jacc.2015.01.017.
7. Nezu T, Hosomi N. Usefulness of carotid ultrasonography for risk stratification of cerebral and cardiovascular disease. *J Atheroscler Thromb*. 2020;27(10):1023-35. doi:10.5551/jat.RV17044.
8. Li H, Xu X, Luo B, Zhang Y. The Predictive Value of Carotid Ultrasonography With Cardiovascular Risk Factors — A “SPIDER” Promoting Atherosclerosis. *Front Cardiovasc Med*. 2021;8:706490. doi:10.3389/fcvm.2021.706490.
9. Shenkova NN, Veselovskaya NG, Chumakova GA, et al. Risk prediction for subclinical atherosclerotic lesion of brachiocephalic arteries in obese women. *Russian Journal of Cardiology*. 2017;(4):54-60. (In Russ.) Шенкова Н.Н., Веселовская Н.Г., Чумакова Г.А. и др. Прогнозирование риска субклинического атеросклероза брахиоцефальных артерий у женщин с ожирением. Российский кардиологический журнал. 2017;(4):54-60. doi:10.15829/1560-4071-2017-4-54-60.
10. Druzhilova OY, Druzhilov MA, Otmakhov VV, et al. Role of assessment of arterial wall stiffness in predicting carotid artery atherosclerosis in patients with abdominal obesity. *Terapevticheskiy Arkhiv*. 2016;88(4):24-8. (In Russ.) Дружилова О.Ю., Дружилов М.А., Отмахов В.В. и др. Роль оценки жесткости артериальной стенки при прогнозировании атеросклероза сонной артерии у пациентов с абдоминальным ожирением. *Терапевтический архив*. 2016;4(88):24-8. doi:10.17116/terarkh201688424-28.
11. Gusev AV, Gavrilov DV, Novitsky RE, et al. Improvement of cardiovascular risk assessment using machine learning methods. *Russian Journal of Cardiology*. 2021;26(12):4618. (In Russ.) Гусев А.В., Гаврилов Д.В., Новицкий Р.Э. и др. Совершенствование возможностей оценки сердечно-сосудистого риска при помощи методов машинного обучения. Российский кардиологический журнал. 2021;26(12):4618. doi:10.15829/1560-4071-2021-4618.
12. Narain R, Saxena S, Goyal A. Cardiovascular risk prediction: a comparative study of Framingham and quantum neural network based approach. *Patient Prefer Adherence*. 2016;10:1259-70. doi:10.2147/PPA.S108203.
13. Dimopoulos A, Nikolaidou M, Caballero F, et al. Machine learning methodologies versus cardiovascular risk scores, in predicting disease risk. *BMC Med Res Methodol*. 2018;18(1):179. doi:10.1186/s12874-018-0644-1.
14. Bengtsson A, Norberg M, Ng N, et al. The beneficial effect over 3 years by pictorial information to patients and their physician about subclinical atherosclerosis and cardiovascular risk: Results from the VIPVIZA randomized clinical trial. *Am J Prev Cardiol*. 2021;7:100199. doi:10.1016/j.ajpc.2021.100199.
15. Rossello X, Fuster V, Oliva B, et al. Association Between Body Size Phenotypes and Subclinical Atherosclerosis. *J Clin Endocrinol Metab*. 2020;105(12):3734-44. doi:10.1210/clinem/dgaa620.

Заключение

Разработанная модель — пример того, как методы машинного обучения могут помочь практикующему врачу принять клиническое решение в области “серых зон”, когда в клинической медицине не имеется однозначного ответа на поставленный вопрос. В данном случае при проведении риск-стратификации у пациента с избыточным весом и ожирением без анамнеза и симптомов ССЗ, прогнозная модель отвечает на вопрос, насколько целесообразно направление такого пациента на УЗИ БЦА для верификации СКА. Принятие решения может быть основано на автоматическом анализе доступных в ЭМК параметров, а при интеграции модели в системы поддержки принятия врачебных решений сбор и интерпретация информации могут проходить в фоновом режиме без отвлечения врача от работы.

Внедрение в практику алгоритмов риск-стратификации, в основу которых будут заложены подобные предсказательные модели, позволит увеличить точность и качество прогнозирования ССР и оптимизировать систему проводимых профилактических мероприятий.

Отношения и деятельность. Исследование выполнено на уникальной научной установке “Многокомпонентный программно-аппаратный комплекс для автоматизированного сбора, хранения, разметки научно-исследовательских и клинических биомедицинских данных, их унификации и анализа на базе ЦОД с использованием технологий искусственного интеллекта” (регистрационный номер 2075518), при финансовой поддержке Министерства науки и высшего образования Российской Федерации в рамках Соглашения № 075-15-2021-665.