

Разработка и валидация моделей машинного обучения, прогнозирующих госпитализации пациентов с артериальной гипертензией в течение 12 месяцев

Андрейченко А. Е.¹, Ермак А. Д.¹, Гаврилов Д. В.¹, Новицкий Р. Э.¹, Драпкина О. М.², Гусев А. В.^{3,4}

¹ООО "К-Скай". Петрозаводск; ²ФГБУ "Национальный медицинский исследовательский центр терапии и профилактической медицины" Минздрава России. Москва; ³ФГБУ "Центральный научно-исследовательский институт организации и информатизации здравоохранения" Минздрава России. Москва; ⁴ГБУЗ города Москвы "Научно-практический клинический центр диагностики и телемедицинских технологий ДЗМ". Москва, Россия

Цель. Разработать с использованием алгоритмов машинного обучения модели прогнозирования госпитализаций пациентов с артериальной гипертензией (АГ) в течение 12 мес. и провести их валидацию на данных реальной клинической практики.

Материал и методы. По сведениям из деперсонифицированных электронных медицинских карт, полученных из платформы Webiomed, отобрано 1165770 записей 151492 пациентов с АГ. В качестве предикторов, после первоначальной селекции, были использованы анамнестические, конституциональные, клинические, инструментальные и лабораторные данные, широко применяемые в рутинной врачебной практике, всего 43 признака. Для создания моделей применялись инструменты автоматического машинного обучения. Рассматривался широкий набор алгоритмов, включая логистическую регрессию, методы, основанные на деревьях решений с использованием градиентного бустинга и бэггинга, дискриминантный анализ, алгоритм на основе нейронных сетей и наивный байесовский классификатор. Для внешней валидации использованы данные отдельного региона.

Результаты. Наилучшие результаты показала модель XGBoost, достигнув AUROC (площадь под характеристической кривой) 0,849 (95% доверительный интервал: 0,825-0,873) при внутреннем тестировании и 0,815 (95% доверительный интервал 0,797-0,835) при внешней валидации.

Закключение. В результате исследования разработана новая высокоточная модель прогнозирования госпитализации пациен-

тов с АГ по данным реальной клинической практики. Результаты внешней валидации предложенного прогностического инструмента показали относительную устойчивость к новым данным из другого региона, что в совокупности с показателями качества отражает возможность ее апробации в реальной клинической практике.

Ключевые слова: артериальная гипертензия, госпитализация, прогнозные модели, машинное обучение.

Отношения и деятельность: нет.

Поступила 31/07-2024

Рецензия получена 30/09-2024

Принята к публикации 07/11-2024



Для цитирования: Андрейченко А. Е., Ермак А. Д., Гаврилов Д. В., Новицкий Р. Э., Драпкина О. М., Гусев А. В. Разработка и валидация моделей машинного обучения, прогнозирующих госпитализации пациентов с артериальной гипертензией в течение 12 месяцев. *Кардиоваскулярная терапия и профилактика*. 2025;24(1):4130. doi: 10.15829/1728-8800-2025-4130. EDN YXVRIN

Development and validation of machine learning models predicting hospitalizations of hypertensive patients over 12 months

Andreychenko A. E.¹, Ermak A. D.¹, Gavrilo D. V.¹, Novitsky R. E.¹, Drapkina O. M.², Gusev A. V.^{3,4}

¹ООО K-Sky. Petrozavodsk; ²National Medical Research Center for Therapy and Preventive Medicine. Moscow; ³Central Research Institute for Health Organization and Informatics. Moscow; ⁴Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies. Moscow, Russia

Aim. To develop models for predicting hospitalizations of hypertensive (HTN) over 12 months using machine learning algorithms and to validate them using real-world practice data.

Material and methods. Based on the data from depersonalized electronic health records obtained from the Webiomed platform,

1165770 records of 151492 patients with HTN were selected. After the initial selection, a total of 43 anamnestic, constitutional, clinical, and paraclinical features were used as predictors. Automatic machine learning tools were used to create the models. A wide range of algorithms was considered, including logistic regression, decision tree-

*Автор, ответственный за переписку (Corresponding author):

e-mail: aermak@webiomed.ru

[Андрейченко А. Е. — к.ф.-м.н., руководитель направления искусственного интеллекта, ORCID: 0000-0001-6359-0763, Ермак А. Д.* — аналитик данных направления искусственного интеллекта, ORCID: 0000-0002-0513-8557, Гаврилов Д. В. — эксперт по медицине, ORCID: 0000-0002-8745-857X, Новицкий Р. Э. — генеральный директор, ORCID: 0000-0002-2350-977X, Драпкина О. М. — д.м.н., профессор, академик РАН, директор, ORCID: 0000-0002-4453-8430, Гусев А. В. — к.т.н., с.н.с. отдела научных основ организации здравоохранения, ORCID: 0000-0002-7380-8460].

based methods using gradient boosting and bagging, discriminant analysis, a neural network algorithm and a naive Bayes classifier. Data from a single region were used for external validation.

Results. The XGBoost model showed the best results, achieving an area under the ROC curve (AUC) of 0,849 (95% confidence interval: 0,825-0,873) during internal testing and 0,815 (95% confidence interval: 0,797-0,835) during external validation.

Conclusion. A new highly accurate model for predicting hospitalization of HTN patients based on real-world data was developed. The results of external validation of the final model showed relative resistance to new data from another region that in combination with quality metrics presents the possibility of its approval for application in clinical practice.

Keywords: hypertension, hospitalization, predictive models, machine learning.

Relationships and Activities: none.

Andreychenko A. E. ORCID: 0000-0001-6359-0763, Ermak A. D.* ORCID: 0000-0002-0513-8557, Gavrilov D. V. ORCID: 0000-0002-8745-857X, Novitsky R. E. ORCID: 0000-0002-2350-977X, Drapkina O. M. ORCID: 0000-0002-4453-8430, Gusev A. V. ORCID: 0000-0002-7380-8460.

*Corresponding author: aermak@webiomed.ru

Received: 31/07-2024

Revision Received: 30/09-2024

Accepted: 07/11-2024

For citation: Andreychenko A. E., Ermak A. D., Gavrilov D. V., Novitsky R. E., Drapkina O. M., Gusev A. V. Development and validation of machine learning models predicting hospitalizations of hypertensive patients over 12 months. *Cardiovascular Therapy and Prevention*. 2025;24(1):4130. doi: 10.15829/1728-8800-2025-4130. EDN YXVRIN

АГ — артериальная гипертензия, АД — артериальное давление, ДАД — диастолическое артериальное давление, ДИ — доверительный интервал, ИБС — ишемическая болезнь сердца, МКБ-10 — международная классификация болезней 10 пересмотра, МО — машинное обучение, ПЦОР — прогностическая ценность отрицательного результата, ПЦПР — прогностическая ценность положительного результата, САД — систолическое артериальное давление, СН — сердечная недостаточность, ЭМК — электронная медицинская карта, AUROC — площадь под характеристической кривой.

Ключевые моменты

Что известно о предмете исследования?

- Методы машинного обучения хорошо зарекомендовали себя в разработке прогностических инструментов для определения исходов различных многофакторных заболеваний.
- Прогнозирование течения артериальной гипертензии, а также оценка риска внеплановой госпитализации пациентов с этим заболеванием и своевременное вмешательство в тактику ведения таких пациентов особенно важны для всего здравоохранения и для профилактики развития осложнений у отдельного пациента.

Что добавляют результаты исследования?

- Сформирован набор данных, включающий записи более чем 150 тыс. пациентов с артериальной гипертензией.
- С использованием общепризнанных технологий, на основе различных алгоритмов машинного обучения, был разработан ряд предиктивных моделей для прогнозирования внеплановых госпитализаций этих пациентов.
- Модель на основе XGBoost показала лучшие метрики точности и стабильность на внешних данных.

Key messages

What is already known about the subject?

- Machine learning methods have proven effectiveness in developing predictive tools for determining outcomes of various multifactorial diseases.
- Predicting the progression of hypertension, along with non-elective hospitalization risk for patients with this condition, and implementing timely interventions in their management are crucial for the healthcare system as a whole and for preventing complications in individual patients.

What might this study add?

- A data set was formed, including records of more than 150 thousand patients with hypertension.
- Using generally accepted technologies, based on various machine learning algorithms, a number of predictive models were developed to predict non-elective hospitalizations of these patients.
- The XGBoost-based model showed the best accuracy metrics and stability on external data.

Введение

Артериальная гипертензия (АГ) — это синдром повышения систолического (САД) артериального давления (АД) ≥ 140 мм рт.ст. и/или диастолического АД (ДАД) ≥ 90 мм рт.ст., обусловленный известной причиной (вторичная АГ), либо не связанный с выявлением явных причин, приводящих к развитию вторичных форм АГ (гипертоническая болезнь) [1].

АГ относится к категории заболеваний с неуклонными темпами роста распространенности и представляет большую угрозу за счет ранней инвалидизации и высокой смертности пациентов. Согласно приблизительным оценкам, ~10% мировых расходов на здравоохранение приходится на лечение и профилактику АГ [2]. При этом, на примере системы здравоохранения США, показано, что ежегодные расходы на госпитализацию по поводу

АГ увеличились с 40 млрд за период с 1979 по 1982гг до 113 млрд за период с 2003 по 2006гг [3]. Таким образом, стабильный популяционный рост, увеличение продолжительности жизни, распространенность АГ и тяжесть ее осложнений приводят к значительному социальному ущербу и колоссальным экономическим расходам.

В научной литературе опубликованы результаты нескольких исследований, связанных с разработкой моделей на основе алгоритмов машинного обучения (МО) с целью прогнозирования развития отдельных осложнений у пациентов с АГ [4–11]. В этих работах сильно варьировали размеры наборов данных (от 3395 до 2037027 записей), а также были использованы различные клинические группы пациентов с АГ. В исследовании Lee SJ, et al. [6] в анализируемую группу дополнительно вошли пациентки с повышением АД во время беременности, а в исследовании Wu X, et al. [7] акцент сделан на АГ у молодых пациентов в возрасте 14–39 лет. Значение площади под характеристической кривой (AUROC) либо C-index, как интегральной метрики качества в задаче бинарной классификации, при проведении внутреннего тестирования существующих моделей, находилось в интервале 0,60–0,932 со средним значением 0,772 согласно результатам 4-х исследований [4, 6, 10, 11]. В единственной публикации с описанием процедуры проведения внешней валидации [9] исследованы результаты модели, использующей значения 555 признаков, что делает ее неприменимой для клинической практики.

Таким образом, оценка риска внеплановой госпитализации пациентов с АГ и своевременное вмешательство в тактику их ведения на амбулаторном этапе особенно важны как для всего здравоохранения, так и для профилактики развития осложнений у отдельного пациента, а создание и внедрение многофакторной персонифицированной модели прогнозирования этих госпитализаций для дальнейшей ее практической апробации остается актуальной задачей.

Цель исследования — разработать с использованием алгоритмов МО модели прогнозирования госпитализаций пациентов с артериальной гипертензией (АГ) в течение 12 мес. и провести их валидацию на данных реальной клинической практики.

Материал и методы

Проведено многоцентровое ретроспективное обсервационное исследование. Основой для него послужила база данных платформы Webiomed, содержащая обезличенные данные электронных медицинских карт (ЭМК) 11,6 млн пациентов. Для анализа был собран набор данных, включающий 1165770 записей о 151492 пациентах с диагнозом АГ, обозначенным в ЭМК соответствующими кодами международной классификации болезней 10 пересмотра (МКБ-10) (таблица 1), в возрасте 18–100 лет за период с 2000 по 2023гг из 11 регионов РФ. Каждая за-

пись представляла собой случай оказания медицинской помощи, включая врачебные осмотры, а также инструментальные и лабораторные исследования. Из полученного набора данных были отложены 44527 записей из одного отдельного региона для проведения внешней валидации [12]. Оставшиеся данные были разделены на обучающую (80%) и тестовую выборки (10%), 10% были использованы для настройки гиперпараметров различных алгоритмов МО (рисунок 1).

В качестве целевого события в исследовании выступал факт госпитализации пациента в круглосуточный стационар по поводу самой АГ и/или ее осложнений в течение года после исследуемого случая лечения. В процессе сбора данных случаи, после которых не было зафиксировано таких госпитализаций, но отсутствовала какая-либо информация о пациенте через >12 мес., исключались из набора. Коды МКБ-10, использовавшиеся для отбора пациентов с АГ, а также определения наличия целевого события (класс 1) или его отсутствия (класс 0), представлены в таблице 1.

Платформа Webiomed поддерживает автоматическое извлечение >2900 машиночитаемых признаков из ЭМК, включая социо-демографические, антропометрические, лабораторные, инструментальные, физикальные и анамнестические данные. Из всего этого множества медицинским экспертом изначально был отобран 61 признак в качестве предикторов госпитализации пациентов с АГ (приложение 1). Количественные признаки с заполненностью менее чем в 10% записей для разработки и валидации моделей не использовались.

Основные этапы обработки данных и моделирования проводились согласно разработанной методике, описанной в предыдущих работах [13, 14]. Для статистического анализа и построения моделей МО использовали язык программирования Python версии 3.9. Результаты анализа количественных данных представлены в виде медианы (Me) и интерквартильного размаха (Q25–Q75), а также минимальных и максимальных значений (Min–Max), а категориальных — в виде долей. Сравнение количественных переменных между группами с целевым событием и без него проводилось с использованием теста Манна-Уитни, категориальных — с помощью χ^2 . Значение $p < 0,05$ принималось за статистически значимое. Для анализа эффективности и качества моделей использовали AUROC [15]. Доверительные интервалы (ДИ) метрики оценивали с помощью метода бутстрапинг на основе случайной генерации тысячи псевдовыборок из 5 тыс. наблюдений [16]. Пороговые значения для активации определялись с использованием индекса Юдена, а также целевых уровней прогностических ценностей для отрицательного (0,999) и положительного (0,5) результатов (ПЦОР/NPV/negative predictive value и ПЦПР/PPV/positive predictive value, соответственно).

При обработке наборов данных к пропущенным значениям в количественных признаках, а также значениям, выходящим за границы, установленные медицинским экспертом на основании клинической практики, применялось заполнение фиксированной константой "–10000" [17]. Упомянутые границы значений представлены в приложении 1. Заполнение пропусков в бинарных признаках проводилось с помощью нулей (отрицание наличия признака). На этапе масштабирования данных использовали несколько подходов: их трансформацию

Таблица 1

Группы кодов МКБ-10, наличие которых в ЭМК пациента учитывалось на разных этапах формирования набора данных

Группа	Коды МКБ
Критерий включения пациентов в исследование	I10*, I11*, I12*, I13*, I15*
Определение наличия целевого события	
Основное заболевание	I10*, I11*, I12*, I13*, I15*
Кардиологические осложнения	I20-I25*, I48*, I49*, I50*, I51.7, I51.8, I51.9, I46.0, I42.0
Нефрологические осложнения	N18*, N19*, N20, R80, N26
Цереброваскулярные осложнения	I61*, I63-I66*, I69*, G45*, F01.1, F01.2, F01.3, F01.8, F01.9, I67.3
Офтальмологические осложнения	H35.0
Сосудистые осложнения	I70*, I73.8, I73.9

Примечание: * — включая все подкоды. МКБ-10 — международная классификация болезней, ЭМК — электронная медицинская карта.

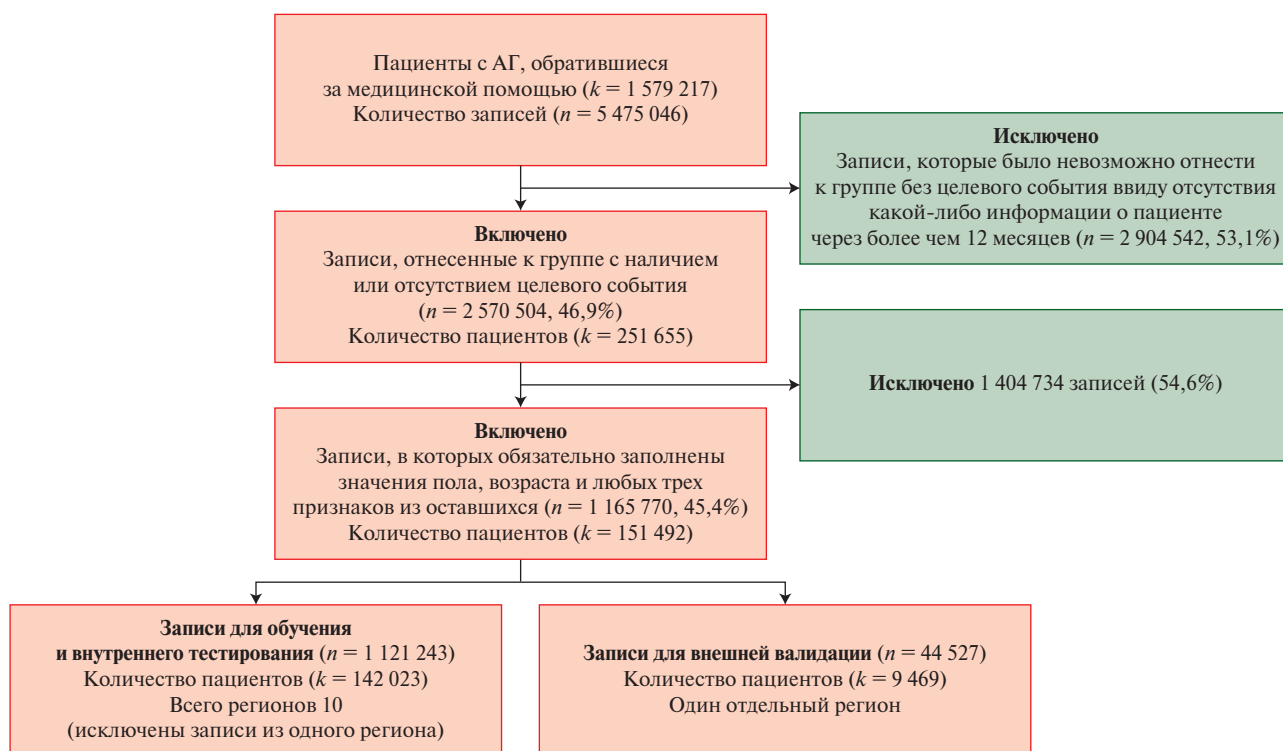


Рис. 1 Дизайн исследования.

Примечание: АГ — артериальная гипертензия.

с помощью гистограммной нормализации или стандартизации, а также использование данных в изначальной размерности [18, 19]. Для исследования целесообразности коррекции дисбаланса классов использовались следующие алгоритмы: Random Undersampling, Random Oversampling, Synthetic Minority Oversampling Technique и Adaptive Synthetic Sampling Approach [20].

Моделирование включало использование логистической регрессии (LR) с L2 регуляризацией, алгоритмов градиентного бустинга (AdaBoost, LightGBM, XGBoost, CatBoost), бэггинга (Random Forest, Extra Trees), линейного и квадратичного дискриминантного анализа, наивного байесовского классификатора и нейронной сети (Multi-layer Perceptron) с одним скрытым слоем из ста нейронов, функцией активации "Relu", L2 регуляризацией и оптимизацией на основе стохастического градиентного спуска.

Первый этап моделирования проводили с использованием всех признаков, включенных в набор после консультации с медицинским экспертом (приложение 1). Для каждой из архитектур итеративно проверяли все возможные подходы к обработке данных с использованием различных комбинаций заполнения пропусков, масштабирования, и коррекции дисбаланса классов. Для дальнейшего изучения использовали ту обработку, при которой, после обучения модели с дефолтными гиперпараметрами, было получено максимальное значение AUROC на наборе данных для внутреннего тестирования. Определение же оптимальных гиперпараметров для всех алгоритмов проводилось с использованием PyCaret [21] и Random Grid Search [22] на обработанных соответствующим образом данных, оптимизируя AUROC на отложенном наборе для настройки гиперпараметров. Сам поиск заложен в структуру модуля и происходит по заранее

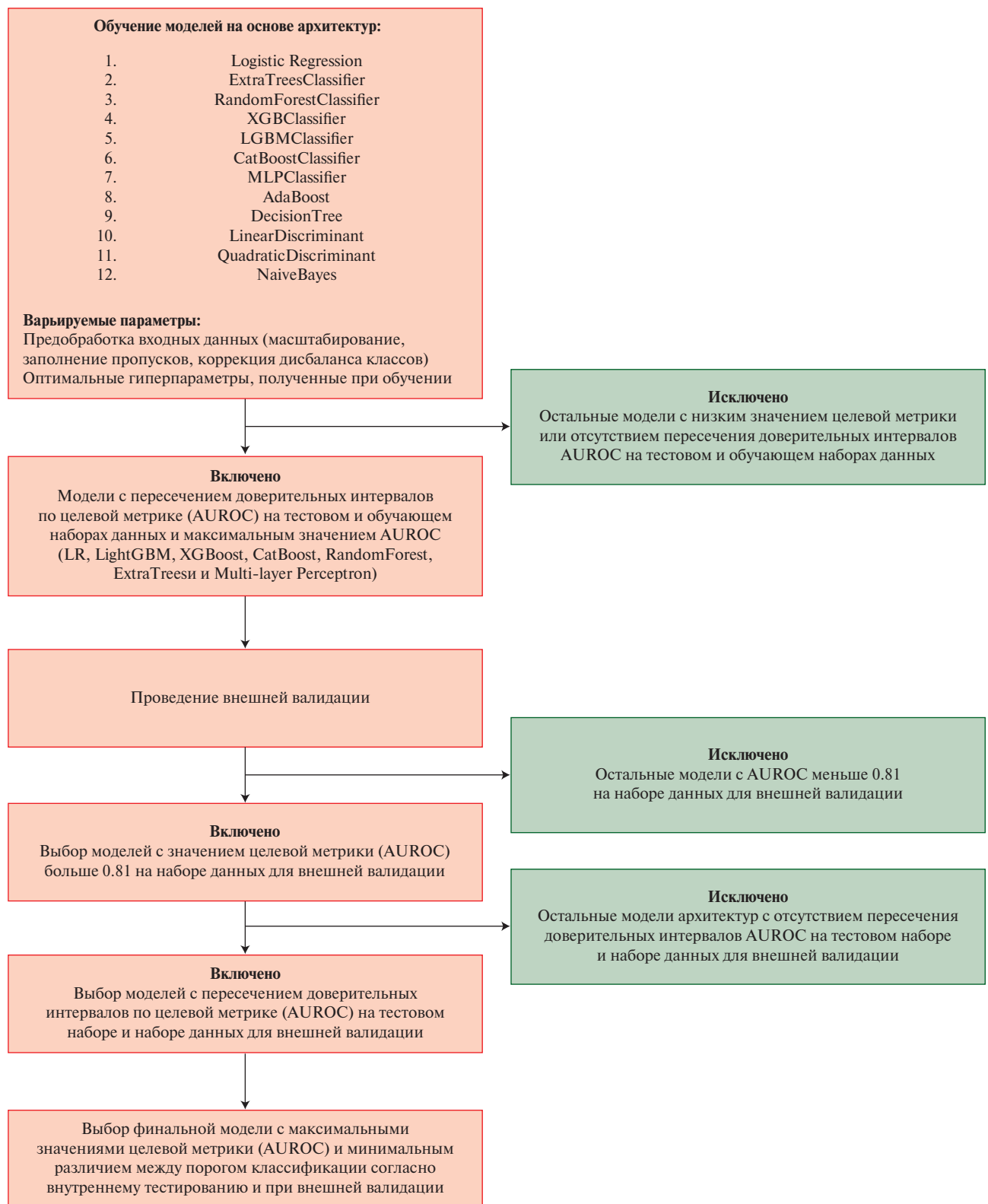


Рис. 2 Алгоритм выбора итоговой модели.

Примечание: AUROC — площадь под характеристической кривой.

сформированным на основе опыта создателей библиотеки сеткам значений, которые могут, при необходимости, быть модифицированы. После этой процедуры модели были повторно обучены на обучающем наборе с использованием отобранных гиперпараметров.

На основании чисел Шепли [23] были отобраны признаки, совокупная значимость которых составила не <95% от значимости всех признаков для лучшей по итогам первой фазы обучения модели на основе XGBoost (приложение 2). С использованием только этих предикторов,

Таблица 2

AUROC [95% ДИ] на наборах для обучения, внутреннего тестирования и внешней валидации для исследованных методов МО

Модель	Обучение	Внутреннее тестирование	Внешняя валидация
XGBoost	0,861 [0,840-0,882]	0,856 [0,833-0,878]	0,816 [0,797-0,834]
LightGBM	0,865 [0,841-0,888]	0,860 [0,834-0,883]	0,814 [0,795-0,830]
Random Forest	0,921 [0,901-0,942]	0,918 [0,896-0,938]	0,811 [0,792-0,829]
Extra Trees	0,897 [0,871-0,920]	0,895 [0,871-0,919]	0,780 [0,760-0,800]
CatBoost	0,916 [0,898-0,934]	0,912 [0,893-0,930]	0,777 [0,756-0,798]
LR	0,783 [0,750-0,814]	0,779 [0,747-0,809]	0,775 [0,754-0,795]
Multi-layer Perceptron	0,669 [0,633-0,703]	0,671 [0,636-0,705]	0,665 [0,640-0,690]

Примечание: ДИ — доверительный интервал, МО — машинное обучение, AUROC — площадь под характеристической кривой. Проведена сортировка по значению метрики при проведении внешней валидации.

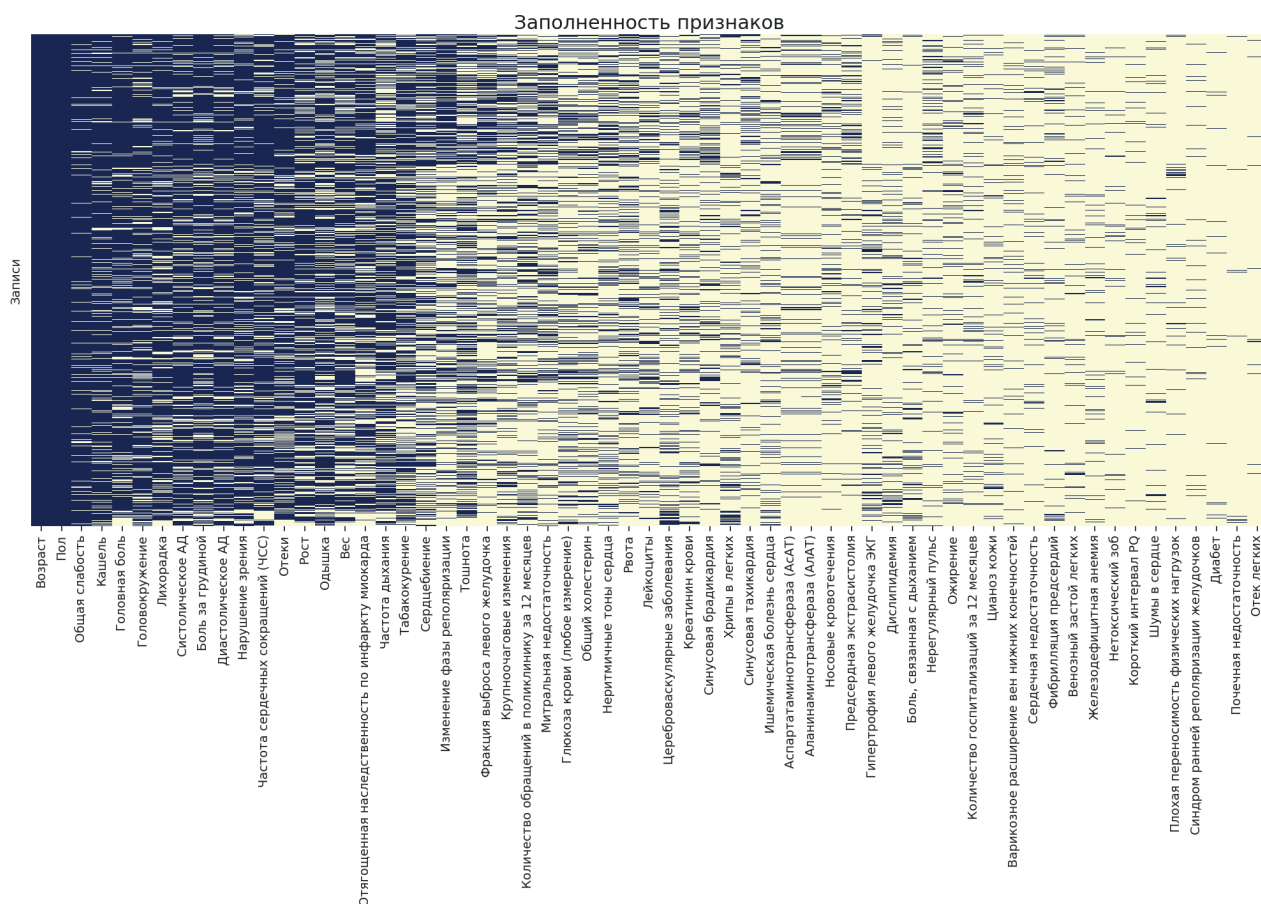


Рис. 3 Заполненность значений в итоговом наборе данных.

Примечание: поля, выделенные синим цветом, обозначают заполненные значения признаков; поля, выделенные жёлтым цветом, обозначают пропуски. Цветное изображение доступно в электронной версии журнала.

в уже определенной ранее соответствующей обработке, для всех алгоритмов был проведен второй этап моделирования, в котором были повторены шаги с настройкой гиперпараметров и непосредственным обучением моделей. До внешней валидации допускались те модели, которые показали пересечение 95% ДИ AUROC на наборах данных для обучения и внутреннего тестирования (рисунок 2). В этот список вошли модели на основе LR, LightGBM, XGBoost, CatBoost, RandomForest, ExtraTrees и Multi-layer Perceptron. Выбор итоговой модели основыв-

вался на максимальном значении AUROC по итогам проведенной внешней валидации, пересечении ДИ метрики и минимальном различии порогов активации на основе индекса Юдена при внутреннем тестировании и внешней валидации.

Результаты

Заполненность признаков в наборе данных до разделения его на выборки для разработки и внеш-

ней валидации отражена на рисунке 3. Среди жалоб чаще всего встречались слабость, кашель, головная боль, головокружение, боль за грудиной, нарушение зрения, отеки и одышка. Самыми распространенными осложнениями были цереброваскулярные

катастрофы, ишемическая болезнь сердца (ИБС), аритмии, варикозное расширение вен нижних конечностей и сердечная недостаточность (СН).

В результате формирования набора данных для разработки в группу с целевым событием были включены 49634 (4%) записи. Оставшиеся 1071609 (96%) были отнесены к группе без целевого события (класс 0). При сравнении двух классов в этом наборе были выявлены статистически значимые различия ряда предикторов (приложение 1), таких как масса тела, возраст, количество госпитализаций и обращений в поликлинику за последние 12 мес., уровни глюкозы, холестерина, САД, ДАД, частота дыхательных движений и частота сердечных сокращений. В классе 1 чаще встречались жалобы на боль за грудиной, головную боль, головокружение, нарушение зрения, носовые кровотечения, одышку, отеки, тошноту и рвоту. Пациенты класса 1 чаще имели отягощенную наследственность по инфаркту миокарда, а также ожирение, сахарный диабет (СД), ИБС, СН, аритмии и цереброваскулярные заболевания.

В набор данных для внешней валидации вошло 44527 записей пациентов: класс 1 — 4335 (9,7%), класс 0 — 40192 (90,3%). При сравнении распределения и частоты выявления предикторов между двумя наборами обнаружены значимые различия для СД, дислипидемии, железодефицитной анемии, ИБС, почечной недостаточности, фибрилляции предсердий, цереброваскулярных заболеваний, СН и ряда других признаков.

Значения AUROC, полученные на наборах для внешней валидации, внутреннего тестирования

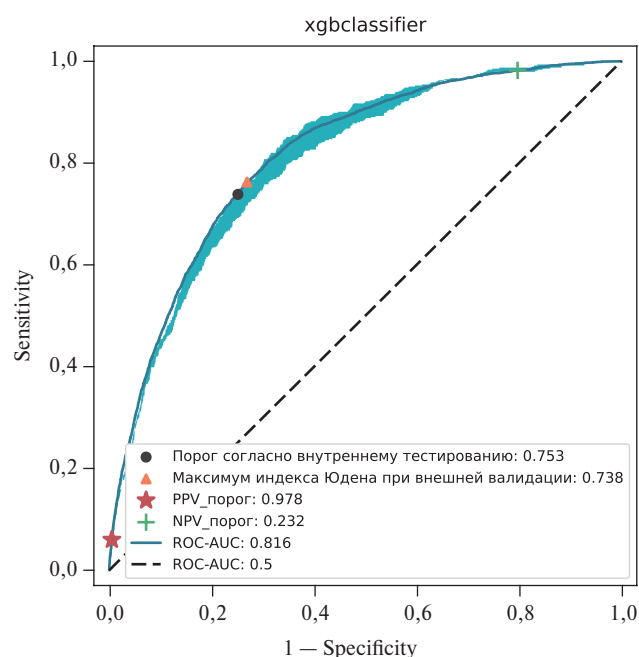


Рис. 4 ROC-кривая с 95% ДИ итоговой модели XGBoost, полученная по результатам прогнозирования на наборе данных для внешней валидации.

Примечание: ДИ — доверительный интервал, PPV_порог — порог при целевом значении ПЦПР (прогностическая ценность положительного результата), NPV_порог — порог при целевом значении ПЦОР (прогностическая ценность отрицательного результата).



Рис. 5 Топ-10 значимых признаков лучшей модели XGBoost.

Примечание: АД — артериальное давление.

и обучения для всех отобранных архитектур, представлены в таблице 2. Модель XGBoost продемонстрировала лучшую способность различать классы 1 и 0, а также стабильность на внешних данных, используя итоговые 43 признака. Значение метрики AUROC составило 0,849 (95% ДИ: 0,825-0,873) при внутреннем тестировании и 0,815 (95% ДИ: 0,797-0,835) при внешней валидации. Модель на основе LightGBM также успешно прошла внешнюю валидацию и соответствовала всем критериям алгоритма отбора (рисунок 2). Однако предпочтение было отдано XGBoost на основании наименьшего различия между рассчитанными максимумами индекса Юдена согласно внутреннему тестированию и при внешней валидации.

Для разделения записей на три группы риска [24] на наборе данных для тестирования нами дополнительно рассчитаны два порога активации в зависимости от целевых ПЦОР (0,999) и ПЦПР (0,5). ROC-кривая модели XGBoost при проведении внешней валидации с указанием трех порогов представлена на рисунке 4. Для внешних данных точность модели с порогом классификации 0,232, при котором достигнуто целевое значение ПЦОР на внутреннем тестировании (0,999), составила 0,2 (95% ДИ: 0,189-0,021). Чувствительность при этом составила 0,996 (95% ДИ: 0,986-1,000), а специфичность — 0,163 (95% ДИ: 0,152-0,173). При использовании второго порога (0,978) с ожидаемой ПЦПР при тестировании, равной 0,5, метрики качества были следующими: точность — 0,956 (95% ДИ: 0,950-0,961), чувствительность — 0,083 (95% ДИ: 0,047-0,012), специфичность — 0,996 (95% ДИ: 0,994-0,998). Значимость 10-и наиболее значимых для итоговой модели признаков, определённых по модулю чисел Шепли, показана на рисунке 5.

Обсуждение

Методы МО хорошо зарекомендовали себя в разработке прогностических инструментов для определения исходов многофакторных заболеваний. При этом в качестве предикторов, используются рутинные клинические и лабораторно-инструментальные параметры, которые можно легко отслеживать и контролировать при визите пациента. Однако среди опубликованной на настоящий момент литературы, нами не было найдено ни одной публикации, где в качестве целевого события использовался именно факт госпитализации пациента по поводу осложнений АГ. Среди немногих работ, связанных с изучением течения этого заболевания, в качестве цели исследования чаще всего использовалось прогнозирование развития сердечно-сосудистых катастроф или заболеваний почек [4-11].

В ряде публикаций, как и в нашей работе, был отмечен выраженный дисбаланс классов в сфор-

мированных наборах данных. При этом только несколько коллективов описали использование алгоритмов балансировки при разработке моделей [4, 6, 8], однако без проведения аналогичных нашему сравнений влияния различных подходов на значение целевой метрики качества. Только в двух работах был описан сам алгоритм определения целевого события, и основывался он на наличии выбранных кодов по МКБ-10 в ЭМК пациента и датах их регистрации [5, 6]. Данная особенность является недостатком как настоящего, так и упомянутых исследований, ввиду возможности оказания медицинской помощи пациенту в медицинской организации, данные от которой не использовались, а также риска ошибки с кодированием диагнозов в реальной практике. Другими важными ограничениями являются изменение подходов к лекарственной терапии за период сбора данных в 21 год, невысокая точность извлечения препаратов и их доз, а также невозможность оценки приверженности пациентов терапии, которые не позволили нам включить лекарственную терапию в обучение моделей. Вместе с тем необходимо отметить, что лекарственная терапия напрямую влияет на значения признаков, участвующих в прогнозе модели (показатели АД, поражение органов-мишеней), и таким образом — на целевое событие разработанной модели.

Отдельного внимания заслуживает присутствие пропусков среди значений количественных признаков. Использование таких данных, с одной стороны, позволяет подготовить модель к условиям реальной практики и работе с ограниченной информацией о пациенте, а с другой — может приводить к смещениям оценок. На примере этого набора нами была выявлена примечательная особенность интерпретации данных линейными алгоритмами МО, обученными при подавляющем количестве пропусков. В этом случае они оказались чувствительны к используемому методу заполнения и переставали оценивать количественные признаки как дискретные, а вместо этого выделяли две категории: записи с реальным значением и заполненной константой. Данная проблема, с нашей точки зрения, значима ввиду потери большого объёма информации, которую можно получить из изменчивости признака, и, возможно, стала основной причиной слабых результатов методов, основанных не на деревьях решений. Все это открывает перспективу для новых исследований, в которых стоит определить оптимальную пороговую заполненность количественных признаков в наборах данных для обучения моделей.

В настоящей работе предложено разделение пациентов на три группы риска с помощью двух порогов активации, что имеет значительную клиническую ценность. Первый порог позволяет точно и эффективно выделить пациентов с низким

риском госпитализации, что дает возможность избежать излишней медицинской нагрузки и проведения ненужных обследований. Второй порог, с высокой специфичностью, помогает с уверенностью выявить пациентов с высоким риском, что позволяет заранее корректировать стратегию лечения и более агрессивно вмешиваться в процессе их наблюдения. Это может существенно улучшить прогноз и качество жизни таких пациентов, снизив вероятность развития осложнений, требующих госпитализации. Такой подход способствует более точному и персонализированному оказанию помощи, что является критически важным в условиях перегрузки системы здравоохранения. Эти факты подчеркивают значительный потенциал разработанной модели как дополнительного инструмента для скрининга пациентов с АГ. С учетом устойчивости модели, продемонстрированной при внешней валидации, ее можно использовать в практических проспективных исследованиях.

Литература/References

- Kobalava ZD, Konradi AO, Nedogoda SV, et al. Arterial hypertension in adults. Clinical guidelines 2020. Russian Journal of Cardiology. 2020;25(3):3786. (In Russ.) Кобалава Ж.Д., Конради А.О., Недогода С.В. и др. Артериальная гипертензия у взрослых. Клинические рекомендации 2020. Российский кардиологический журнал. 2020;25(3):3786. doi:10.15829/1560-4071-2020-3-3786.
- Gaziano TA, Bitton A, Anand S, et al. The global cost of nonoptimal blood pressure. J Hypertens. 2009;27:1472-7. doi:10.1097/HJH.0b013e32832a9ba3.
- Wang G, Fang J, Ayala C. Hypertension-associated hospitalizations and costs in the United States, 1979-2006. Blood Pressure. 2014;23:126-33. doi:10.3109/08037051.2013.814751.
- Lee W, Lee J, Lee H, et al. Prediction of hypertension complications risk using classification techniques. Ind Eng Manag Syst. 2014;13:449-53. doi:10.7232/iems.2014.13.4.449.
- Feng Y, Leung AA, Lu X, et al. Personalized prediction of incident hospitalization for cardiovascular disease in patients with hypertension using machine learning. BMC Med Res Methodol. 2022;22:325. doi:10.1186/s12874-022-01814-3.
- Lee SJ, Lee SH, Choi HI, et al. Deep learning improves prediction of cardiovascular disease-related mortality and admission in patients with hypertension: analysis of the Korean National Health Information Database. J Clin Med. 2022;11:6677. doi:10.3390/jcm11226677.
- Wu X, Yuan X, Wang W, et al. Value of a machine learning approach for predicting clinical outcomes in young patients with hypertension. Hypertension. 2020;75:1271-8. doi:10.1161/HYPERTENSIONAHA.119.13404.
- Ren Y, Fei H, Liang X, et al. A hybrid neural network model for predicting kidney disease in hypertension patients based on electronic health records. BMC Med Inform Decis Mak. 2019;19:51. doi:10.1186/s12911-019-0765-4.
- Park J, Kim JW, Ryu B, et al. Patient-level prediction of cerebrovascular events in hypertension using Nationwide Claims Data. J Med Intern Res. 2019;21:11757. doi:10.2196/11757.
- Lacson RC, Baker B, Suresh H, et al. Use of machine-learning algorithms to determine features of systolic blood pressure variability that predict poor outcomes in hypertensive patients. Clin Kidney J. 2019;12:206-12. doi:10.1093/ckj/sfy049.
- Chen R, Yang Y, Miao F, et al. 3-year risk prediction of coronary heart disease in hypertension patients: a preliminary study. 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2017;1182-5. doi:10.1109/EMBC.2017.8037041.
- Moons KGM, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015;162:W1-73. doi:10.7326/M14-0698.
- Andreychenko AE, Ermak AD, Gavrilov DV, et al. Development and validation of machine learning models to predict unplanned hospitalizations of patients with diabetes within the next 12 months. Diabetes mellitus. 2024;27(2):142-57. (In Russ.) Андрейченко А.Е., Ермак А.Д., Гаврилов Д.В. и др. Разработка и валидация моделей машинного обучения, прогнозирующих риск госпитализации пациентов с сахарным диабетом в течение последующих 12 месяцев. Сахарный диабет. 2024;27(2):142-57. doi:10.14341/DM13065.
- Andreychenko AE, Luchinin AS, Ivshin AA, et al. Development and validation of models to predict total and early-onset preeclampsia in the first trimester of pregnancy using machine learning algorithms. Akusherstvo i Ginekologiya. 2023;2:94-107. (In Russ.) Андрейченко А.Е., Лучинин А.С., Ившин А.А. и др. Разработка и валидация моделей прогнозирования общего риска преэклампсии и риска ранней преэклампсии с использованием алгоритмов машинного обучения в первом триместре беременности. Акушерство и гинекология. 2023;2:94-107. doi:10.18565/aig.2023.101.
- Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. Information Processing & Management. 2009;45:427-37. doi:10.1016/j.ipm.2009.03.002.
- Zoubir AM, Iskander DR. Bootstrap Methods and Applications: A Tutorial for the Signal Processing Practitioner. IEEE Signal Processing Magazine. 2007;24:10-9. doi:10.1109/MSP.2007.4286560.
- Ding Y, Simonoff JS. An investigation of missing data methods for classification trees applied to binary response data. J Mach Learn Res. 2010;11:131-70. doi:10.1145/1756006.1756012.
- Cao XH, Stojkovic I, Obradovic Z. A robust data scaling algorithm to improve classification accuracies in biomedical data. BMC Bioinformatics. 2016;17. doi:10.1186/s12859-016-1236-x.

Заключение

В результате исследования была успешно разработана и валидирована модель МО на основе XGBoost, метрики которой соответствовали результатам инструментов, представленных в ранее опубликованных исследованиях. Результаты внешней валидации подтвердили, что модель сохраняет свою стабильность при обработке данных из других регионов, что, в сочетании с её качественными характеристиками, свидетельствует о возможности ее применения в реальной клинической практике. Подход, включающий выделение трех групп риска, значительно увеличивает ценность настоящей разработки, позволяя формировать группы пациентов с реально высоким или низким риском госпитализации.

Отношения и деятельность: все авторы заявляют об отсутствии потенциального конфликта интересов, требующего раскрытия в данной статье.

19. Amorim LB, Cavalcanti GD, Cruz RM. The choice of scaling technique matters for classification performance. Appl Soft Comput. 2023;133. doi:10.1016/j.asoc.2022.109924.
20. Weiss GM. Foundations of Imbalanced Learning. In: Haibo H, Yunqian M. Imbalanced Learning: Foundations, Algorithms, and Applications. USA: John Wiley & Sons. 2013:13-41. ISBN: 9781118074626.
21. Gain U, Hotti V. Low-code AutoML-augmented data pipeline — a review and experiments. JPCS. 2021;1828. doi:10.1088/1742-6596/1828/1/012015.
22. Bergstra J, Bengio Y. Random search for hyper-parameter optimization. J Mach Learn Res. 2012;13:281-305.
23. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell. 2020;2:56-67. doi:10.1038/s42256-019-0138-9.
24. Fischer BG, Evans AT. SpPin and SnNout are not enough. It's time to fully embrace likelihood ratios and probabilistic reasoning to achieve diagnostic excellence. J Gen Inter Med. 2023;38:2202-4. doi:10.1007/s11606-023-08177-5.

Приложение 1. Описательная статистика набора данных

Результаты анализа количественных данных представлены в виде медианы и интерквартильного размаха (Q25, Q75), а также Min-Max значений, а категориальных — в виде долей

Признак	Границы для идентификации выбросов	Класс 0 (n=1071609)	Класс 1 (n=49634)	p	Набор для внешней валидации (n=44527)	Набор для разработки (n=1121243)	p
Анамнестические							
Варикозное расширение вен нижних конечностей, n (%)		92076 (8,6)	4136 (8,3)	0,046	2774 (6,2)	96212 (8,6)	<0,001
Сахарный диабет, n (%)		22433 (2,1)	1656 (3,3)	<0,001	2427 (5,5)	24089 (2,1)	<0,001
Дислипидемия, n (%)		133965 (12,5)	5874 (11,8)	<0,001	6729 (15,1)	139839 (12,5)	<0,001
Железодефицитная анемия, n (%)		64413 (6,0)	3208 (6,5)	<0,001	3098 (7,0)	67621 (6,0)	<0,001
ИБС, n (%)		167040 (15,6)	15666 (31,6)	<0,001	5559 (12,5)	182706 (16,3)	<0,001
Количество госпитализаций за 12 мес.		0,0 (0,0, 1,0)	1,0 (0,0, 1,0)	<0,001	1,0 (0,0, 1,0)	0,0 (0,0, 1,0)	<0,001
Min-Max значения		(0,0; 7,0)	(0,0; 11,0)		(0,0; 8,0)	(0,0; 11,0)	
Количество обращений в поликлинику за 12 мес.		1,0 (0,0, 2,0)	1,0 (0,0, 2,0)	<0,001	1,0 (0,0, 4,0)	1,0 (0,0, 2,0)	<0,001
Min-Max значения		(0,0; 17,0)	(0,0; 27,0)		(0,0; 25,0)	(0,0; 27,0)	
Нетоксический зуб, n (%)		59117 (5,5)	2450 (4,9)	<0,001	1922 (4,3)	61567 (5,5)	<0,001
Ожирение, n (%)		108737 (10,1)	4612 (9,3)	<0,001	8175 (18,4)	113349 (10,1)	<0,001
Отягощенная наследственность по ИМ, n (%)		134592 (12,6)	4986 (10,0)	<0,001	4078 (9,2)	139578 (12,4)	<0,001
Почечная недостаточность, n (%)		12627 (1,2)	2272 (4,6)	<0,001	4484 (10,1)	14899 (1,3)	<0,001
СН, n (%)		82286 (7,7)	6732 (13,6)	<0,001	1956 (4,4)	89018 (7,9)	<0,001
ФП, n (%)		74992 (7,0)	11102 (22,4)	<0,001	4968 (11,2)	86094 (7,7)	<0,001
Цереброваскулярные заболевания, n (%)		238582 (22,3)	16318 (32,9)	<0,001	14772 (33,2)	254900 (22,7)	<0,001
Отек легких, n (%)		11558 (1,1)	2588 (5,2)	<0,001	800 (1,8)	14146 (1,3)	<0,001
Табакокурение, n (%)		321932 (30,0)	17637 (35,5)	<0,001	15398 (34,6)	339569 (30,3)	<0,001
Конституциональные							
Масса тела, кг	[20, 200]	77,0 (67,0, 88,0)	78,0 (67,0, 89,5)	<0,001	83,0 (71,0, 96,0)	77,0 (67,0, 88,0)	<0,001
Min-Max значения		(21,0; 175,0)	(34,0; 170,0)		(31,0; 165,0)	(21,0; 175,0)	
Возраст, лет	[18, 100]	62,0 (54,0, 70,0)	69,0 (61,0, 77,0)	<0,001	54,0 (45,0, 61,0)	62,0 (54,0, 70,0)	<0,001
Min-Max значения		(18,0; 90,0)	(18,0; 100,0)		(18,0; 93,0)	(18,0; 90,0)	
Пол мужской, n (%)		310367 (29,0)	19964 (40,2)	<0,001	16210 (36,4)	330331 (29,5)	<0,001
Рост, см	[100, 250]	163,0 (157,0, 169,0)	163,0 (156,0, 170,0)	<0,001	165,0 (160,0, 172,0)	163,0 (157,0, 169,0)	<0,001
Min-Max значения		(100,0; 195,0)	(100,0; 205,0)		(100,0; 198,0)	(100,0; 205,0)	
Клинические							
Боль за грудиной, n (%)		858487 (80,1)	43119 (86,9)	<0,001	33465 (75,2)	901606 (80,4)	<0,001

Признак	Границы для идентификации выбросов	Класс 0 (n=1071609)	Класс 1 (n=49634)	p	Набор для внешней валидации (n=44527)	Набор для разработки (n=1121243)	p
Боль, связанная с дыханием, n (%)		120803 (11,3)	7197 (14,5)	<0,001	12969 (29,1)	128000 (11,4)	<0,001
Головная боль, n (%)		907050 (84,6)	38348 (77,3)	<0,001	34086 (76,6)	945398 (84,3)	<0,001
Головокружение, n (%)		877094 (81,8)	43974 (88,6)	<0,001	33669 (75,6)	921068 (82,1)	<0,001
Кашель, n (%)		909625 (84,9)	39791 (80,2)	<0,001	37923 (85,2)	949416 (84,7)	<0,001
Лихорадка, n (%)		872381 (81,4)	35446 (71,4)	<0,001	37555 (84,3)	907827 (81,0)	<0,001
Митральная недостаточность, n (%)		347646 (32,4)	22589 (45,5)	<0,001	11038 (24,8)	370235 (33,0)	<0,001
Нарушение зрения, n (%)		843412 (78,7)	36908 (74,4)	<0,001	32425 (72,8)	880320 (78,5)	<0,001
Нерегулярный пульс, n (%)		122590 (11,4)	12378 (24,9)	<0,001	3020 (6,8)	134968 (12,0)	<0,001
Неритмичные тоны сердца, n (%)		320240 (29,9)	25264 (50,9)	<0,001	12146 (27,3)	345504 (30,8)	<0,001
Носовые кровотечения, n (%)		149210 (13,9)	9169 (18,5)	<0,001	7473 (16,8)	158379 (14,1)	<0,001
Общая слабость, n (%)		997748 (93,1)	47899 (96,5)	<0,001	41215 (92,6)	1045647 (93,3)	<0,001
Одышка, n (%)		740394 (69,1)	43829 (88,3)	<0,001	29196 (65,6)	784223 (69,9)	<0,001
Отеки, n (%)		814809 (76,0)	39570 (79,7)	<0,001	35520 (79,8)	854379 (76,2)	<0,001
Плохая переносимость физических нагрузок, n (%)		46481 (4,3)	3410 (6,9)	<0,001	6630 (14,9)	49891 (4,4)	<0,001
Рвота, n (%)		296477 (27,7)	15796 (31,8)	<0,001	10167 (22,8)	312273 (27,9)	<0,001
Сердцебиение, n (%)		507898 (47,4)	30394 (61,2)	<0,001	20113 (45,2)	538292 (48,0)	<0,001
Тошнота, n (%)		485899 (45,3)	23170 (46,7)	<0,001	20733 (46,6)	509069 (45,4)	<0,001
Хрипы в легких, n (%)		166064 (15,5)	15030 (30,3)	<0,001	22919 (51,5)	181094 (16,2)	<0,001
Цианоз кожи, n (%)		85794 (8,0)	12028 (24,2)	<0,001	7143 (16,0)	97822 (8,7)	<0,001
Шумы в сердце, n (%)		52277 (4,9)	6523 (13,1)	<0,001	932 (2,1)	58800 (5,2)	<0,001
Инструментальные							
Венозный застой легких по данным рентгеновских исследований, n (%)		56797 (5,3)	8733 (17,6)	<0,001	6379 (14,3)	65530 (5,8)	<0,001
Гипертрофия ЛЖ, n (%)		144358 (13,5)	9916 (20,0)	<0,001	11467 (25,8)	154274 (13,8)	<0,001
ДАД, мм рт.ст.	[20, 150]	80,0 (80,0, 85,0)	80,0 (80,0, 90,0)	<0,001	80,0 (80,0, 90,0)	80,0 (80,0, 85,0)	<0,001
Min-Max значения		(20,0; 150,0)	(20,0; 150,0)		(21,0; 150,0)	(20,0; 150,0)	
Изменение фазы реполяризации ЭКГ, n (%)		497968 (46,5)	24277 (48,9)	<0,001	11812 (26,5)	522245 (46,6)	<0,001
Короткий интервал PQ, n (%)		57173 (5,3)	3545 (7,1)	<0,001	1173 (2,6)	60718 (5,4)	<0,001
Крупноочаговые изменения на ЭКГ, n (%)		360114 (33,6)	22105 (44,5)	<0,001	15514 (34,8)	382219 (34,1)	<0,001
Предсердная экстрасистолия, n (%)		146658 (13,7)	12088 (24,4)	<0,001	5901 (13,3)	158746 (14,2)	<0,001
Синдром ранней реполяризации желудочков, n (%)		45963 (4,3)	2179 (4,4)	0,088	1510 (3,4)	48142 (4,3)	<0,001
Синусовая брадикардия, n (%)		221257 (20,6)	10950 (22,1)	<0,001	5944 (13,3)	232207 (20,7)	<0,001
Синусовая тахикардия, n (%)		184558 (17,2)	10079 (20,3)	<0,001	6717 (15,1)	194637 (17,4)	<0,001
САД, мм рт.ст.	[40, 250]	130,0 (120,0, 140,0)	130,0 (125,0, 145,0)	<0,001	127,5 (120,0, 140,0)	130,0 (120,0, 140,0)	<0,001
Min-Max значения		(40,0; 250,0)	(40,0; 250,0)		(41,0; 250,0)	(40,0; 250,0)	
ФВ ЛЖ, %	[0, 100]	64,0 (60,0, 68,0)	60,0 (53,0, 64,0)	<0,001	62,0 (58,0, 66,0)	64,0 (60,0, 68,0)	<0,001
Min-Max значения		(20,0; 78,0)	(30,0; 72,0)		(32,0; 77,0)	(20,0; 78,0)	
Частота дыхательных движений, в мин	[5, 65]	16,5 (16,0, 18,0)	17,0 (16,0, 18,0)	<0,001	17,0 (16,0, 18,0)	16,5 (16,0, 18,0)	0,057
Min-Max значения		(5,0; 65,0)	(8,0; 62,0)		(5,0; 60,0)	(5,0; 65,0)	

Артериальная гипертония

Признак	Границы для идентификации выбросов	Класс 0 (n=1071609)	Класс 1 (n=49634)	p	Набор для внешней валидации (n=44527)	Набор для разработки (n=1121243)	p
Частота сердечных сокращений, в мин	[25, 300]	72,0 (68,0, 77,0)	72,0 (68,0, 78,0)	<0,001	74,0 (70,0, 79,0)	72,0 (68,0, 77,0)	<0,001
Min-Max значения		(25,0; 250,0)	(25,0; 200,0)		(28,0; 195,0)	(25,0; 250,0)	
Лабораторные							
АлАТ, МЕ/л	[5, 5000]	19,7 (14,2, 28,8)	18,0 (13,0, 26,9)	<0,001	24,5 (17,5, 38,0)	19,6 (14,1, 28,6)	<0,001
Min-Max значения		(5,0; 2585,1)	(5,0; 601,0)		(5,0; 2116,0)	(5,0; 2585,1)	
АсАТ, МЕ/л	[5, 5000]	22,0 (18,0, 28,0)	22,0 (17,5, 29,0)	0,054	25,8 (20,0, 34,0)	22,0 (18,0, 28,0)	<0,001
Min-Max значения		(5,0; 1822,0)	(5,0; 1296,2)		(6,0; 3846,0)	(5,0; 1822,0)	
Глюкоза, ммоль/л	[1, 50]	5,7 (5,1, 6,7)	6,0 (5,2, 7,5)	<0,001	5,8 (5,1, 7,1)	5,7 (5,1, 6,7)	<0,001
Min-Max значения		(2,0; 50,0)	(2,7; 49,5)		(1,0; 50,0)	(2,0; 50,0)	
Креатинин, мкмоль/л	[10, 500]	84,0 (73,7, 97,0)	92,0 (76,8, 112,0)	<0,001	81,0 (70,0, 96,0)	84,0 (74,0, 97,1)	<0,001
Min-Max значения		(40,0; 500,0)	(30,0; 500,0)		(45,0; 500,0)	(30,0; 500,0)	
Лейкоциты, 10 ⁹ /л	[0,1, 75]	6,6 (5,4, 8,0)	7,2 (5,8, 8,8)	<0,001	6,7 (5,5, 8,2)	6,6 (5,4, 8,0)	<0,001
Min-Max значения		(0,6; 75,0)	(1,2; 74,5)		(1,6; 56,5)	(0,6; 75,0)	
Общий холестерин, ммоль/л	[1, 25]	5,2 (4,4, 6,2)	4,8 (3,9, 5,9)	<0,001	5,2 (4,3, 6,1)	5,2 (4,3, 6,2)	0,062
Min-Max значения		(1,0; 25,0)	(1,0; 25,0)		(1,0; 23,0)	(1,0; 25,0)	

Примечание: АлАТ — аланинаминотрансфераза, АсАТ — аспартатаминотрансфераза, ДАД — диастолическое артериальное давление, ИБС — ишемическая болезнь сердца, ИМ — инфаркт миокарда, ЛЖ — левый желудочек, САД — систолическое артериальное давление, СН — сердечная недостаточность, ФВ — фракция выброса, ФП — фибрилляция предсердий, ЭКГ — электрокардиограмма(-фия). Критерии дислипидемии: общий холестерин (ХС) $\geq 5,0$ ммоль/л, ХС липопротеинов низкой плотности $\geq 3,0$ ммоль/л, ХС липопротеинов высокой плотности $\leq 1,1$ ммоль/л для женщин, ХС липопротеинов высокой плотности $\leq 1,0$ ммоль/л для мужчин, триглицериды $\geq 1,7$ ммоль/л.

Приложение 2. Список предикторов, использованных в разработке итоговых версий моделей (в алфавитном порядке)

	Название признака	Относительная значимость признака
1	Аланинаминотрансфераза	0,039
2	Аспаратаминотрансфераза	0,026
3	Боль за грудиной, в т.ч. купирующаяся нитроглицерином	0,042
4	Венозный застой легких	0,081
5	Масса тела	0,099
6	Возраст	0,300
7	Гипертрофия левого желудочка по данным ЭКГ	0,038
8	Глюкоза крови	0,035
9	Головная боль	0,069
10	Головокружение	0,112
11	Диастолическое артериальное давление	0,034
12	Изменение фазы реполяризации	0,115
13	Ишемическая болезнь сердца	0,076
14	Кашель	0,098
15	Количество госпитализаций за 12 мес.	0,129
16	Количество обращений в поликлинику за 12 мес.	0,170
17	Креатинин крови	0,082
18	Крупноочаговые изменения	0,044
19	Лейкоциты крови	0,042
20	Лихорадка	0,136
21	Митральная недостаточность	0,011
22	Нарушение зрения	0,091
23	Неритмичные тоны сердца	0,143
24	Общая слабость	0,059
25	Общий холестерин	0,032
26	Одышка	0,226
27	Отеки	0,071
28	Пол	0,183
29	Предсердная экстрасистолия	0,117
30	Рвота	0,085
31	Рост	0,097
32	Сердечная недостаточность	0,020
33	Сердцебиение	0,058
34	Систолическое артериальное давление	0,132
35	Табакокурение	0,064
36	Фибрилляция предсердий	0,074
37	Фракция выброса левого желудочка	0,368
38	Хрипы в легких	0,126
39	Цереброваскулярные заболевания	0,056
40	Цианоз кожи	0,140
41	Частота дыхания	0,113
42	Частота сердечных сокращений	0,051
43	Шумы в сердце	0,046