

© Коллектив авторов, 2023

А.Е. АНДРЕЙЧЕНКО¹, А.С. ЛУЧИНИН², А.А. ИВШИН³, А.Д. ЕРМАК¹, Р.Э. НОВИЦКИЙ¹, А.В. ГУСЕВ^{4,5}

РАЗРАБОТКА И ВАЛИДАЦИЯ МОДЕЛЕЙ ПРОГНОЗИРОВАНИЯ ОБЩЕГО РИСКА ПРЕЭКЛАМПСИИ И РИСКА РАННЕЙ ПРЕЭКЛАМПСИИ С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ В ПЕРВОМ ТРИМЕСТРЕ БЕРЕМЕННОСТИ

¹ООО «К-Скай», Петрозаводск, Россия

²ФГБУН «Кировский научно-исследовательский институт гематологии и переливания крови
Федерального медико-биологического агентства», Киров, Россия

³ФГБОУ ВО «Петрозаводский государственный университет», Петрозаводск, Россия

⁴ФГБУ «Центральный научно-исследовательский институт организации и информатизации здравоохранения»
Минздрава России, Москва, Россия

⁵ГБУЗ города Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий
Департамента здравоохранения города Москвы», Москва, Россия

Актуальность: Преэклампсия (ПЭ) представляет собой грозное, трудно прогнозируемое осложнение беременности, характеризующееся полиорганный дисфункцией. Заболевание встречается с частотой 2–8% среди всех беременностей и является одной из ведущих причин перинатальной и материнской смертности, в особенности при раннем дебюте.

Цель: Разработка моделей прогнозирования общего риска ПЭ и риска ранней ПЭ с использованием технологий машинного обучения (МО) на основе данных реальной клинической практики в I триместре беременности.

Материалы и методы: По сведениям из электронных медицинских карт, полученных через платформу Webiomed, были проанализированы 21 092 записи, соответствующие 12 434 уникальным беременностям 12 283 женщин в возрасте от 11 до 60 лет на сроке до 16 недель включительно. В качестве потенциальных факторов прогноза ПЭ отобраны анамнестические, конституциональные, клинические, инструментальные и лабораторные данные, широко используемые в рутинной врачебной практике, всего — 53 переменные. Для создания моделей применялись логистическая регрессия (LR), методы градиентного бустинга (LightGBM, XGBoost, CatBoost) и методы, основанные на деревьях решений (RandomForest и ExtraTrees).

Результаты: Наилучшие результаты при решении задачи прогнозирования ПЭ показала модель ExtraTrees — площадь под кривой (AUC) 0,858 (95% ДИ 0,827–0,890). Точность модели составила 0,634 (95% ДИ 0,616–0,652), чувствительность — 0,897 (95% ДИ 0,837–0,953), специфичность — 0,624 (95% ДИ 0,605–0,643). В числе моделей оценки риска ранней ПЭ наилучшие результаты продемонстрировал алгоритм RandomForest. По результатам его валидации AUC составила 0,848 (95% ДИ 0,785–0,904), точность — 0,813 (95% ДИ 0,798–0,828), чувствительность — 0,733 (95% ДИ 0,565–0,885), специфичность — 0,814 (95% ДИ 0,799–0,828).

Заключение: Метрики полученных итоговых моделей соответствовали опубликованным ранее аналогам. Результаты внешней валидации показали относительную устойчивость моделей к новым данным, что в совокупности с показателями качества предполагает возможность их использования в реальной клинической практике. Это наш первый опыт прогнозирования сложного по механизму развития осложнения беременности на основе данных реальной клинической практики. Качество прогностической модели напрямую зависит от используемых данных и статистических алгоритмов, которые нам предстоит улучшить в будущих исследованиях.

Ключевые слова: преэклампсия, большие акушерские синдромы, ранняя диагностика, прогнозные модели, машинное обучение, искусственный интеллект.

Вклад авторов: Андрейченко А.Е. — руководство исследованием, редактирование текста статьи; Лучинин А.С. — составление текста статьи; Ившин А.А. — концепция исследования, экспертный анализ результатов, редактирование текста статьи; Ермак А.Д. — моделирование и анализ данных; Новицкий Р.Э., Гусев А.В. — концепция исследования.

Конфликт интересов: Авторы заявляют об отсутствии конфликта интересов.

Финансирование: Исследование выполнено с использованием Уникальной научной установки (УНУ)

«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» (рег. номер: 2075518).

Благодарности: Авторы выражают свою признательность Д.В. Гаврилову за ценные консультации при сборе данных и валидации результатов работы моделей, а также В.О. Баркиной и В.Ю. Борисову за помощь в обработке данных и построении моделей.

Обмен исследовательскими данными: Данные, подтверждающие выводы этого исследования, доступны по запросу у автора, ответственного за переписку, после одобрения ведущим исследователем.

Для цитирования: Андрейченко А.Е., Лучинин А.С., Ившин А.А., Ермак А.Д., Новицкий Р.Э., Гусев А.В. 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; 10: 94–107
<https://dx.doi.org/10.18565/aig.2023.101>

©A group of authors, 2023

A.E. ANDREYCHENKO¹, A.S. LUCHININ², A.A. IVSHIN³, A.D. ERMAK¹,
R.E. NOVITSKIY¹, A.V. GUSEV^{4,5}

DEVELOPMENT AND VALIDATION OF MODELS TO PREDICT TOTAL AND EARLY-ONSET PREECLAMPSIA IN THE FIRST TRIMESTER OF PREGNANCY USING MACHINE LEARNING ALGORITHMS

¹K-SkAI LLC, Petrozavodsk, Russia

²Kirov Research Institute of Hematology and Blood Transfusion under the Federal Medical Biological Agency, Kirov, Russia

³Petrozavodsk State University, Petrozavodsk, Russia

⁴Russian Research Institute of Health, Ministry of Health of Russia, Moscow, Russia

⁵Center for Diagnostics and Telemedicine, Moscow Health Department, Moscow, Russia

Background: Preeclampsia (PE) is a life-threatening and difficult-to-predict complication of pregnancy, characterized by multi-organ dysfunction. PE affects 2–8% of all pregnancies and is one of the leading causes of perinatal and maternal mortality, especially in cases of early onset PE.

Objective: To develop models for predicting total and early onset PE in the first trimester of pregnancy using machine learning (ML) technologies based on real-world clinical data.

Materials and methods: We analyzed 21,092 records obtained from electronic medical records through the Webiomed platform, corresponding to 12,434 unique pregnancies of 12,283 women aged 11 to 60 years, up to 16 weeks. Anamnestic, constitutional, clinical, instrumental, and laboratory data, commonly used in routine medical practice, were selected as potential factors for predicting PE, totaling 53 variables. To create the models, we employed logistic regression (LR), gradient boosting methods (LightGBM, XGBoost, CatBoost), and methods based on decision trees (RandomForest and ExtraTrees).

Results: The ExtraTrees model demonstrated the highest accuracy in predicting PE, with an area under the curve (AUC) of 0.858 (95% CI 0.827–0.890). The model's overall accuracy was 0.634 (95% CI 0.616–0.652), sensitivity was 0.897 (95% CI 0.837–0.953), and specificity was 0.624 (95% CI 0.605–0.643). Among the models for assessing the risk of early onset PE, the RandomForest algorithm yielded the most promising results. The AUC after validation was 0.848 (95% CI 0.785–0.904), with an accuracy of 0.813 (95% CI 0.798–0.828), sensitivity of 0.733 (95% CI 0.565–0.885), and specificity of 0.814 (95% CI 0.799–0.828).

Conclusion: The metrics of the final models align with previously published models. External validation results demonstrate the relative stability of the models with new data, indicating their potential applicability in real clinical practice. This is our first experience in predicting complex pregnancy complications based on real-world clinical data. The quality of the predictive model depends directly on the data and the statistical algorithms used, aspects that we intend to refine in future studies.

Keywords: preeclampsia, great obstetrical syndromes, early diagnosis, predictive models, machine learning, artificial intelligence.

Authors' contributions: Andreychenko A.E. – study administration, manuscript editing; Luchinin A.S. – manuscript drafting; Ivshin A.A. – conception of the study, expert analysis of results, manuscript editing; Ermak A.D. – modeling and data analysis; Novitskiy R.E., Gusev A.V. – conception of the study.

Conflicts of interest: The authors have no conflicts of interest to declare.

Funding: This study was performed using the Unique Scientific Unit (UNU) «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» (reg. number: 2075518).

Acknowledgments: The authors would like to thank D.V. Gavrilov for valuable consultations during data collection and validation of model results, and to V.O. Barkina and V.Yu. Borisov for assistance in data processing and model building.

Authors' Data Sharing Statement: The data supporting the findings of this study are available upon request from the corresponding author after approval from the principal investigator.

For citation: Andreychenko A.E., Luchinin A.S., Ivshin A.A., Ermak A.D., Novitskiy R.E., Gusev A.V. 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/Obstetrics and Gynecology*. 2023; (10): 94-107 (in Russian) <https://dx.doi.org/10.18565/aig.2023.101>

Преэклампсия (ПЭ) представляет собой грозное, трудно прогнозируемое осложнение беременности, характеризующееся полиорганной дисфункцией. Заболевание встречается с частотой 2–8% среди всех беременностей и является одной из ведущих причин материнской и младенческой смертности [1]. В настоящее время этиология и патогенез ПЭ до конца не изучены, поэтому терапия носит симптоматиче-

ский характер. У разных пациенток ПЭ протекает неодинаково, поскольку в патологический процесс вовлекаются различные системы органов; при этом клиническая картина полиорганной дисфункции широко варьирует [2]. Эта особенность затрудняет своевременную диагностику, прогноз наступления и прогрессирования ПЭ. В этой связи создание и внедрение в практическое здравоохранение много-

факторной модели прогнозирования ПЭ является актуальной задачей, позволяющей улучшить исходы беременности [3]. В недавнем опубликованном систематическом обзоре 68 прогностических моделей ПЭ показано, что наиболее часто используемыми предикторами являлись заболевания беременной в анамнезе, индекс массы тела (ИМТ), артериальное давление (АД), состояние кровотока в маточных артериях и возраст матери. В то же время, специфические биомаркеры и предикторы, полученные по результатам ультразвуковой диагностики, значительно не улучшали дискриминативные свойства моделей. Значение показателя площади под ROC-кривой (AUC) для исследованных моделей варьировало в широком диапазоне от 0,61 до 0,996. Только в 4% и 6% всех работ имелась информация о внутренней и внешней валидации, соответственно [4].

Построение многопараметрических прогностических моделей, несмотря на существующие трудности, может способствовать эффективному прогнозированию ПЭ и, как следствие, своевременному принятию клинических решений в виде результативной профилактики. В настоящее время в научной литературе опубликованы результаты исследований, связанных с разработкой прогностических моделей ПЭ. Так, Thangaratnam S. et al. построили две прогностические модели (PREP-L и PREP-S), используя регрессионный анализ Кокса и логистический регрессионный анализ, включив в них параметры, связанные с эффектом от терапии высокого АД и спазмолитиков. Из-за сложности и неудобства использования модели не нашли применения в клинической практике [5]. Одним из вариантов упрощения, удобства и интуитивности понимания прогностических моделей является их графическая визуализация в виде номограммы, которые часто используют в онкологии [6]. Применение такого подхода в сочетании с алгоритмом LR легло в основу другой модели прогноза ПЭ с хорошей дискриминативностью. По результатам внутренней валидации значение AUC составило 0,957 (0,935–0,979), при этом средняя абсолютная ошибка по данным калибровочной кривой была равна 1,4% [7].

Поскольку заболеваемость ПЭ во всем мире не снижается, качество и эффективность создаваемых моделей прогнозирования варьируется в широких пределах, а возможность их практического применения до настоящего времени отсутствует — вопрос исследований в данной области представляется актуальным.

Цель исследования: разработка и валидация моделей прогнозирования общего риска ПЭ и риска ранней ПЭ с использованием алгоритмов машинного обучения (МО) в I триместре беременности на данных реальной клинической практики.

Материалы и методы

Источник данных. Проведено ретроспективное исследование базы данных платформы прогнозной аналитики Webiomed, содержащей неперсонифицированные формализованные данные электронных медицинских карт (ЭМК) 11,6 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 в платформу Webiomed, в том числе для научно-исследовательских целей. Поскольку анализировались обезличенные медицинские данные, информированное добровольное согласие пациентов не требовалось. Медицинские данные и машиночитаемые значения признаков собирались из ЭМК с помощью различных технологий извлечения информации из неструктурированных либо слабо структурированных текстовых данных, в том числе NLP (natural language processing). На момент исследования платформа Webiomed поддерживала автоматическое извлечение более 2900 логических, категориальных и количественных признаков.

Участники. На основе имеющейся базы данных был сформирован набор данных, включающий 21 092 записи, соответствующие 12 434 уникальным беременностям 12 283 женщин, которым была оказана медицинская помощь в период с 21.12.2004 г. по 28.11.2022 г. Под записью понимается случай получения медицинской помощи, зафиксированный в ЭМК пациентки с установленным на момент обращения сроком беременности вместе с уточненными на дату этого случая значениями клинических параметров пациентки. Были использованы следующие критерии включения случая в набор данных: 1) установленный диагноз, соответствующий состоянию беременности согласно классификации МКБ-10 (табл. 1); 2) срок беременности в диапазоне 7–16 недель; 3) известный исход беременности (на основе записей в ЭМК с соответствующими кодами МКБ-10); 4) обязательное наличие значений возраста, систолического АД, диастолического АД, роста и веса.

Из итогового набора данных дополнительно формировалась отдельная выборка, в которую были включены записи, относящиеся только к беременностям с ранней ПЭ или с отсутствием ПЭ (всего 20 516 медицинских записей).

Полученные наборы данных содержали информацию о беременностях у пациенток из 11 регионов

Таблица 1. Группы кодов МКБ-10, наличие которых в ЭМК пациентки учитывалось на разных этапах формирования набора данных

Группа	Коды МКБ
Факт обращения за медицинской помощью по поводу беременности	O10*, O12*, O13*, O16*, O20–26*, O28–36*, O40–41*, O43–48*, O88*, O98*, O99*, Z32–36*
Определение исходов беременности	O61–75*, O80–87*, O89–92*, O95*, O36.5, P95, Z37–39*
Определение наличия целевого события	O11, O11.9, O14, O14.0, O14.1, O14.9, O15, O15.0, O15.1, O15.2, O15.9

*включая все подкоды.

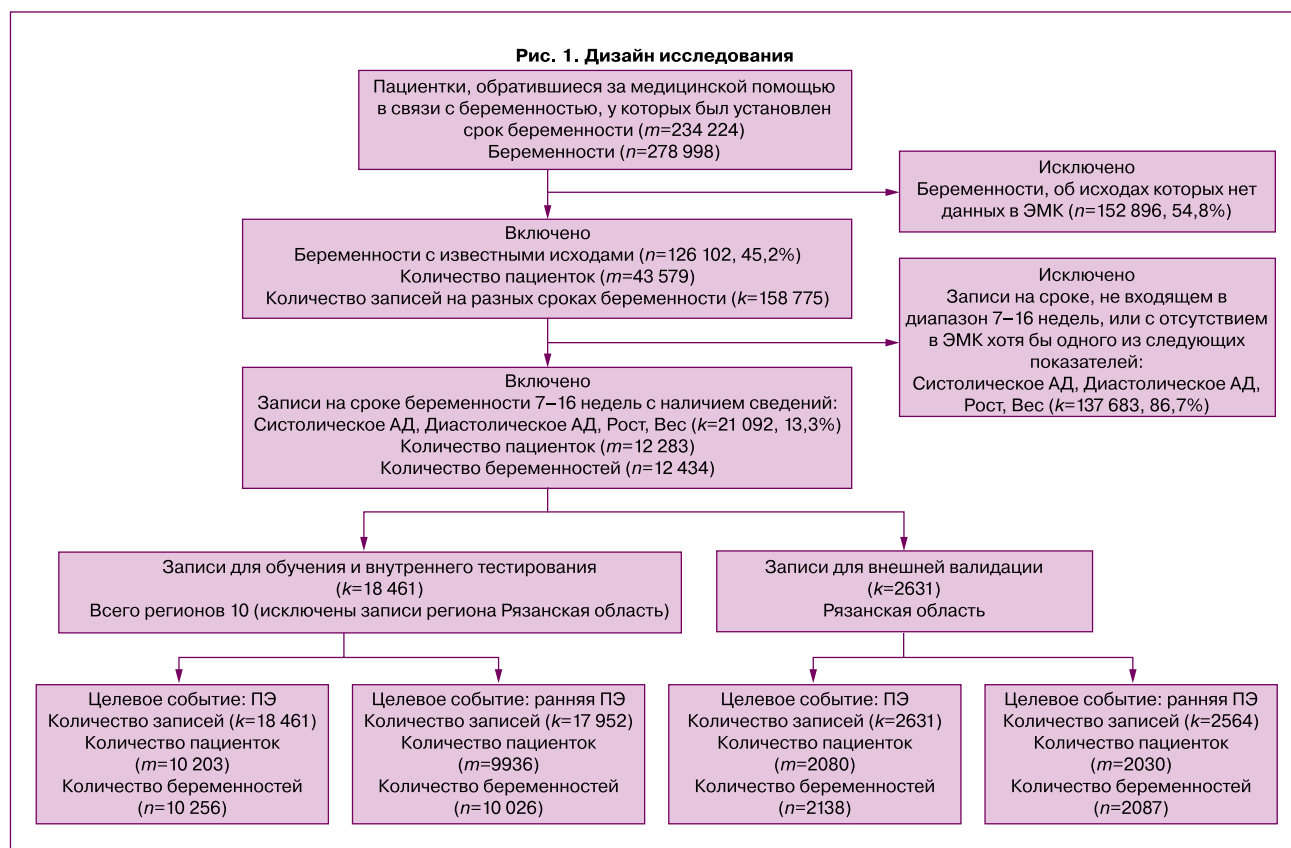
Российской Федерации. На основании территориального распределения для обеих групп были исключены выборки для проведения внешней валидации [8], в которые вошли записи беременных из Рязанской области: 2564 записи для прогнозирования ранней ПЭ и 2631 запись для определения общего риска ПЭ в течение беременности. Выборка для внешней валидации была сформирована согласно дизайну типа 2b рекомендаций TRIPOD [8]. В соответствии с обозначенными рекомендациями, формирование выборки для внешней валидации на основе территориальной принадлежности либо временного промежутка является более сильным дизайном внешней валидации предиктивных моделей, чем случайное разделение. Общий дизайн исследования представлен на рисунке 1. После исключения данных Рязанской области, оставшиеся данные были использованы для разработки моделей и разделены случайным образом на обучающую выборку (80%), выборку для подбора гиперпараметров (10%) и внутреннего тестирования (10%). Выборки для внутреннего тестирования и внешней валидации не использовались при обучении алгоритмов.

В проведенном исследовании модели прогнозирования ПЭ создавались на основе многоцентровых данных реальной клинической практики из различных регионов Российской Федерации, не накладывая требования к обязательному наличию результатов отдельных лабораторных тестов, имеющих известную прогностическую ценность в отношении предикции ПЭ, но крайне редко применяемых в повседневной практике.

Исходы. В качестве целевого события выступала ПЭ. Все записи пациентки, относящиеся к текущей

беременности, если за время ее течения была диагностирована ПЭ, относились к группе с целевым событием. При отсутствии в ЭМК пациентки в течение текущей беременности кодов МКБ-10, отобранных для определения целевого события, все записи, соответствующие данной беременности, относили к группе без целевого события. Коды МКБ-10, использовавшиеся для отбора пациенток, обратившихся за медицинской помощью по поводу беременности, определения исходов беременностей, а также наличия или отсутствия целевого события представлены в таблице 1.

Предикторы. В качестве потенциальных факторов прогнозирования ПЭ выбрали анамнестические, конституциональные, клинические и лабораторные параметры (всего 53 переменные). В набор данных включали значения параметров из ЭМК пациенток с датами регистрации, не превышающими дату окончания 16-й недели беременности. Принимая во внимание, что ПЭ диагностируется не ранее 20-й недели беременности, изученные в период с 7-й по 16-ю неделю беременности параметры рассматривались в качестве прогностических в отношении ПЭ. Анамнестическими факторами считались сопутствующие или прошлые заболевания и патологические состояния, особенности становления менструального цикла, паритет и вредные привычки, которые могли повлиять на исход текущей беременности, включая в себя табакокурение, возраст менархе, предстоящие первые роды, бесплодие, экстракорпоральное оплодотворение (ЭКО), маловодие, ПЭ, плацентарную недостаточность, задержку роста плода, дистресс плода и гибель плода в анамнезе, интергенетический интер-



вал, мертворождение, сердечно-сосудистые заболевания (ССЗ), тромбозы, неврологические заболевания, болезни мочевыделительной системы (МВС), сахарный диабет (СД), гестационный СД и инфекции передаваемые половым путем (ИППП). К конституциональным параметрам отнесли возраст на момент наступления беременности, рост, вес во время и до беременности, ИМТ во время и до беременности и срок беременности на момент анализа. Клинические факторы характеризовали текущую беременность и включали в себя цианоз кожи, нарушение зрения, головную боль, боль в животе, судороги, тошноту, рвоту, систолическое АД (САД), диастолическое АД (ДАД), среднее АД, тяжелую артериальную гипертензию (АГ), умеренную АГ, отеки, отек легких, гемолиз, анурию или олигурию, полиорганную недостаточность, анемию и тромбоцитопению, многоплодие. К лабораторным параметрам отнесли количество тромбоцитов, уровень креатинина крови, аланин-аминотрансферазы (АЛТ), аспартатаминотрансферазы (АСТ), активированное частичное тромбопластиновое время (АЧТВ) и фибриноген.

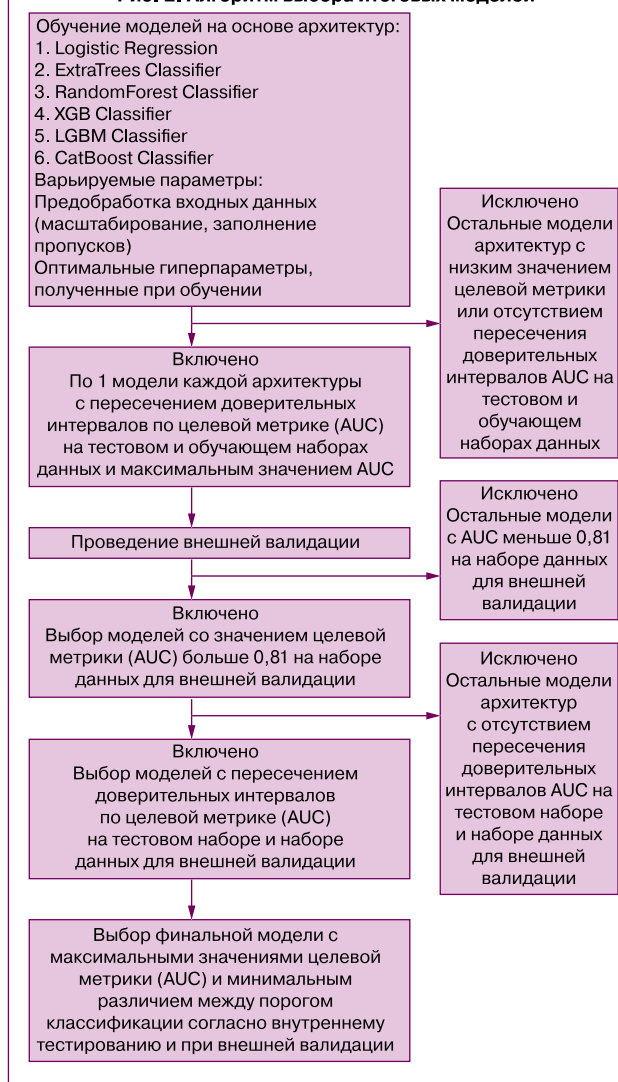
Коррекция выбросов и заполнение пропущенных значений. Заполнение пропущенных значений в численных параметрах выполняли с помощью постоянного значения «-10000» [9], в бинарных — с помощью нулевых значений (отрицаний наличия признака). При обработке количественных признаков значения, выходящие за установленные на основании клинической практики границы, исключались. Используемые границы значений параметров для удаления выбросов представлены в Приложении. Также, перед моделированием численные параметры трансформировали с помощью гистограммной нормализации [10].

Статистический анализ

Статистический анализ и построение моделей МО выполняли на языке программирования Python, версия 3.9. Нормальность распределения количественных переменных оценивали при помощи критерия Шапиро—Уилка, их данные отражены в виде медианы с указанием межквартильного размаха, категориальных — в виде долей (N , %). Сравнение количественных переменных в группах с целевым событием (ПЭ) и без проводили при помощи теста Манна—Уитни, категориальных — при помощи χ^2 , значение $p < 0,05$ считалось статистически значимым.

В качестве алгоритмов МО применяли логистическую регрессию (LR), методы градиентного бустинга (LightGBM, XGBoost, CatBoost) и методы, основанные на деревьях решений (RandomForest и ExtraTrees). Внутреннюю валидацию полученных моделей проводили на тестовом наборе данных. Критериями оценки эффективности и качества моделей служили AUC, чувствительность (Recall), специфичность, точность (Accuracy), прогностическая ценность положительного результата (Precision), F-мера, матрица ошибок и калибровочные кривые [11, 12]. Доверительные интервалы значений выбранных статистик оценивались с помощью метода бутстрап путем случайной генерации множества псевдовыборок [13]. В качестве порога классификации результата работы моделей

Рис. 2. Алгоритм выбора итоговых моделей



использовали максимум индекса Юдена. Значимость предикторов, входящих в модели МО определяли по методу SHAP [14].

Отобранные согласно нашему алгоритму модели проходили внешнюю валидацию. Расчет матрицы ошибок и всех метрик качества в этом случае проводился с использованием порога классификации согласно внутреннему тестированию и максимума индекса Юдена, определенного на наборе данных для внешней валидации. Выбор финальной модели основывался на максимальном значении AUC при внешней валидации, пересечении доверительных интервалов метрики и минимальном различии между порогом классификации согласно внутреннему тестированию и при внешней валидации. Полная процедура отбора моделей представлена на рисунке 2.

Результаты

Описательная статистика. По итогам формирования набора данных для обучения и внутреннего тестирования (т.е. разработки) в группу с наличием целевого события были включены 672 (3,6%) записи, отно-

сящихся к 358 беременностям, в течение которых была диагностирована ПЭ (класс 1). Оставшиеся 17 789 (96,4%) записей (9 938 беременностей) отнесли к группе без целевого события (класс 0). При формировании подвыборки для прогнозирования риска развития ранней ПЭ из общего количества записей с наличием целевого события были исключены записи, относящиеся к беременностям, в которых ПЭ развилась после 34-й недели. Общее число записей с целевым событием в этой группе составило 163 (0,9%).

Распределение по частоте встречаемости и величине ряда признаков в общем наборе данных между двумя классами имело статистически значимые отличия. Классы отличались между собой по таким характеристикам, как вес и ИМТ до и вовремя беременности, уровни САД и ДАД, среднего АД и АЛТ. Также среди

записей с ПЭ (класс 1) чаще встречались указания на головную боль, тошноту, рвоту, боль в животе, отеки, гемолиз, нарушения зрения, умеренную и тяжелую АГ, маловодие, задержку роста и дистресс плода, плацентарную недостаточность, ССЗ, неврологические заболевания и СД. Среди пациенток, формировавших записи 1 класса, было больше первородящих женщин, а также женщин с отягощенным анамнезом по ПЭ и бесплодию. Статистические показатели величины и частоты признаков в наборе данных для обучения и тестирования в зависимости от наличия целевого события представлены в таблице 2.

В набор данных для внешней валидации в общей сложности вошла 2631 запись пациенток из Рязанской области. Было отмечено соответствующее набору для разработки разделение запи-

Таблица 2. Описательная статистика набора данных

Признаки	Характеристика признаков в наборе для разработки в зависимости от наличия или отсутствия целевого события			Сравнение распределения признаков в наборе для разработки и наборе для внешней валидации		
	Класс 0 (N=17 789)	Класс 1 (N=672)	p	Внешняя валидация (N=2631)	Разработка (N=18 461)	p
Анамнестические						
Табакокурение	6417 (36.1%)	342 (50.9%)		837 (31.8%)	6759 (36.6%)	<0.0001
Менархе	13.0 (12.0, 14.0)	13.0 (12.0, 14.0)		13.0 (12.0, 14.0)	13.0 (12.0, 14.0)	
Первородящая	7980 (44.9%)	402 (59.8%)	<0.0001	1892 (71.9%)	8382 (45.4%)	<0.0001
Бесплодие в анамнезе	1222 (6.9%)	64 (9.5%)	0.009	185 (7.0%)	1286 (7.0%)	
ЭКО	1711 (9.6%)	71 (10.6%)		213 (8.1%)	1782 (9.7%)	0.011
Маловодие	873 (4.9%)	64 (9.5%)	<0.0001	181 (6.9%)	937 (5.1%)	<0.0001
Преэклампсия	120 (0.7%)	14 (2.1%)	<0.0001	59 (2.2%)	134 (0.7%)	<0.0001
Плацентарная недостаточность	236 (1.3%)	44 (6.5%)	<0.0001	210 (8.0%)	280 (1.5%)	<0.0001
Задержка роста плода	2209 (12.4%)	204 (30.4%)	<0.0001	522 (19.8%)	2413 (13.1%)	<0.0001
Дистресс плода	5416 (30.4%)	280 (41.7%)	<0.0001	1276 (48.5%)	5696 (30.9%)	<0.0001
Гибель плода	61 (0.3%)	5 (0.7%)		20 (0.8%)	66 (0.4%)	0.004
Интергенетический интервал	1.7 (1.0, 2.8)	2.0 (0.8, 2.1)		1.9 (1.1, 3.2)	1.7 (1.0, 2.7)	
Мертворождение	164 (0.9%)	10 (1.5%)		42 (1.6%)	174 (0.9%)	0.002
ССЗ	2869 (16.1%)	202 (30.1%)	<0.0001	986 (37.5%)	3071 (16.6%)	<0.0001
Тромбозы	2029 (11.4%)	68 (10.1%)		304 (11.6%)	2097 (11.4%)	
Неврологические заболевания	2614 (14.7%)	159 (23.7%)	<0.0001	303 (11.5%)	2773 (15.0%)	<0.0001
Болезни МВС	4082 (22.9%)	165 (24.6%)		903 (34.3%)	4247 (23.0%)	<0.0001
СД	185 (1.0%)	22 (3.3%)	<0.0001	12 (0.5%)	207 (1.1%)	0.002
Гестационный СД	5150 (29.0%)	269 (40.0%)	<0.0001	198 (7.5%)	5419 (29.4%)	<0.0001
ИППП	60 (0.3%)	2 (0.3%)		14 (0.5%)	62 (0.3%)	
Конституциональные						
Возраст	29.0 (26.0, 33.0)	30.0 (25.0, 34.0)		30.0 (26.0, 34.0)	30.0 (26.0, 33.0)	
Рост	164.0 (160.0, 168.0)	164.0 (159.0, 168.0)		165.0 (161.0, 168.0)	164.0 (159.5, 168.0)	
Вес	63.0 (56.0, 73.8)	70.0 (60.0, 88.0)	<0.0001	66.0 (58.0, 77.0)	63.0 (56.0, 74.0)	
Вес до беременности	62.0 (55.0, 73.0)	68.0 (59.0, 80.0)	<0.0001	65.0 (60.0, 75.0)	62.0 (55.0, 73.0)	
ИМТ	23.6 (20.9, 27.4)	26.5 (22.7, 31.9)	<0.0001	24.2 (21.4, 28.3)	23.7 (20.9, 27.6)	
ИМТ до беременности	23.3 (20.5, 27.0)	25.2 (21.7, 30.5)	<0.0001	23.6 (21.8, 27.7)	23.4 (20.6, 27.1)	
Срок беременности	12.0 (10.0, 14.0)	12.0 (10.0, 14.0)	0.025	12.0 (10.0, 13.0)	12.0 (10.0, 14.0)	

Клинические

Цианоз кожи	781 (4.4%)	36 (5.4%)		19 (0.7%)	817 (4.4%)	<0.0001
Нарушение зрения	10852 (61.0%)	443 (65.9%)	0.011	964 (36.6%)	11295 (61.2%)	<0.0001
Головная боль	2691 (15.1%)	154 (22.9%)	<0.0001	670 (25.5%)	2845 (15.4%)	<0.0001
Боль в животе	10980 (61.7%)	447 (66.5%)	0.013	1859 (70.7%)	11427 (61.9%)	<0.0001
Судороги	1141 (6.4%)	47 (7.0%)		176 (6.7%)	1188 (6.4%)	
Тошнота	5767 (32.4%)	259 (38.5%)	0.001	469 (17.8%)	6026 (32.6%)	<0.0001
Рвота	5951 (33.5%)	282 (42.0%)	<0.0001	383 (14.6%)	6233 (33.8%)	<0.0001
САД	110.0 (105.0, 120.0)	117.0 (110.0, 120.0)	<0.0001	115.0 (110.0, 120.0)	110.0 (105.0, 120.0)	
ДАД	70.0 (66.5, 80.0)	75.0 (70.0, 80.0)	<0.0001	72.0 (70.0, 80.0)	70.0 (67.0, 80.0)	
Среднее АД	83.0 (80.0, 90.0)	88.0 (83.0, 93.0)	<0.0001	86.0 (82.5, 93.0)	83.0 (80.0, 90.0)	
Умеренная АГ	350 (2.0%)	51 (7.6%)	<0.0001	159 (6.0%)	401 (2.2%)	<0.0001
Тяжелая АГ	37 (0.2%)	6 (0.9%)	0.001	10 (0.4%)	43 (0.2%)	
Отеки	10426 (58.6%)	553 (82.3%)	<0.0001	1602 (60.9%)	10979 (59.5%)	
Отек легких	77 (0.4%)	19 (2.8%)	<0.0001	4 (0.2%)	96 (0.5%)	0.015
Гемолиз	167 (0.9%)	27 (4.0%)	<0.0001	3 (0.1%)	194 (1.1%)	<0.0001
Анурия или олигурия	391 (2.2%)	19 (2.8%)		6 (0.2%)	410 (2.2%)	<0.0001
Полиорганная недостаточность	3 (0.0%)	1 (0.1%)		4 (0.2%)	4 (0.0%)	0.007
Анемии и тромбоцитопении	8734 (49.1%)	347 (51.6%)		1234 (46.9%)	9081 (49.2%)	0.029
Многоплодная беременность	3929 (22.1%)	170 (25.3%)		649 (24.7%)	4099 (22.2%)	0.005

Лабораторные

Тромбоциты	239.0 (204.0, 278.0)	243.5 (207.5, 280.0)		206.5 (184.6, 236.5)	239.0 (204.0, 278.0)	
АЧТВ	29.5 (27.2, 31.5)	28.8 (26.1, 31.5)		30.0 (26.7, 31.5)	29.4 (27.1, 31.5)	
АЛТ	13.1 (10.0, 18.9)	15.0 (10.1, 26.4)	0.032	16.2 (12.2, 20.9)	13.2 (10.0, 19.0)	
АСТ	17.0 (14.7, 20.6)	17.0 (14.9, 22.5)		19.0 (14.0, 30.0)	17.0 (14.7, 20.7)	
Креатинин	61.0 (55.0, 68.0)	61.0 (51.6, 69.6)		67.0 (62.0, 74.6)	61.0 (55.0, 68.0)	
Фибриноген	3.6 (3.1, 4.1)	3.6 (3.1, 4.3)		4.4 (4.0, 5.1)	3.6 (3.1, 4.1)	

сей по классам: класс 1 – 97 (3,6%), класс 0 – 2534 (96,3%). В наборах для разработки и внешней валидации моделей прогнозирования ранней ПЭ также бы одинаковое распределение на группы по развитию заболевания – доля записей с целевым событием составила 1% (163 и 30 записей, соответственно)

При сравнении распределения и частоты встречаемости отобранных нами предикторов между двумя наборами были обнаружены статистически значимые различия в частоте встречаемости болезней МВС, СД, ИППП, неврологических заболеваний, ССЗ, плацентарной недостаточности, дистресса плода, маловодия, табакокурения и ряда факторов, характеризующих течение текущей беременности (боль в животе, головная боль, тошнота, нарушение зрения, полиорганная недостаточность, цианоз кожи, гемолиз, многоплодие). Общая характеристика использовавшихся наборов данных, а также их сравнение представлены в таблице 2.

Разработка моделей. Создание моделей прогнозирования ПЭ и ранней ПЭ в течение беременности проводилось с помощью 6 алгоритмов МО: LR, LightGBM, XGBoost, CatBoost, RandomForest и ExtraTrees. На основании абсолютных значений

векторов Шепли, полученных по результатам обучения алгоритмов на первоначальных наборах данных, проводили селекцию признаков, отбирая для дальнейшего анализа предикторы, вносящие 95% процентов кумулятивного вклада в результат работы моделей. Также, вне зависимости от значимости по SHAP, для исследования оставляли перечень обязательных признаков, определенный на основании общепринятых материнских факторов риска ПЭ [15].

Следующим этапом проводили повторное обучение алгоритмов с учетом только отобранных признаков. Для моделей прогнозирования ПЭ в итоговый список вошли 34 предикторов, для ранней ПЭ – 36. Списки признаков, использованных в качестве предикторов в работе моделей представлены в приложении. Для дальнейшего изучения оставляли лишь те модели, которые показали максимальное значение целевой метрики качества (AUC) на тестовом наборе данных при пересечении 95% доверительных интервалов данной метрики на обучающей и тестовой выборках, чтобы избежать переобучения. На основании результатов внешней валидации, согласно алгоритму, представленному на рисунке 2, выбирали конечные модели.

Производительность моделей. Характеристика эффективности каждой модели после применения порога классификации на наборе для внутреннего тестирования и внешней валидации представлены в таблицах 3 и 4, соответственно. Наиболее высокую дискриминативную способность и устойчивость

к внешним данным в задаче прогнозирования на сроках до 16 недель включительно развития ПЭ в течение беременности показала модель ExtraTrees со значением целевой метрики AUC 0,858 (95% ДИ 0,827–0,890) при внешней валидации и 0,862 (95% ДИ 0,800–0,914). Для внешнего набора данных точность

Рис. 3. AUC на наборах для внутреннего тестирования и внешней валидации для всех отобранных моделей.
А: прогнозирование общего риска преэклампсии, Б: прогнозирование риска ранней преэклампсии.
Черные вертикальные линии указывают пределы 95% доверительных интервалов

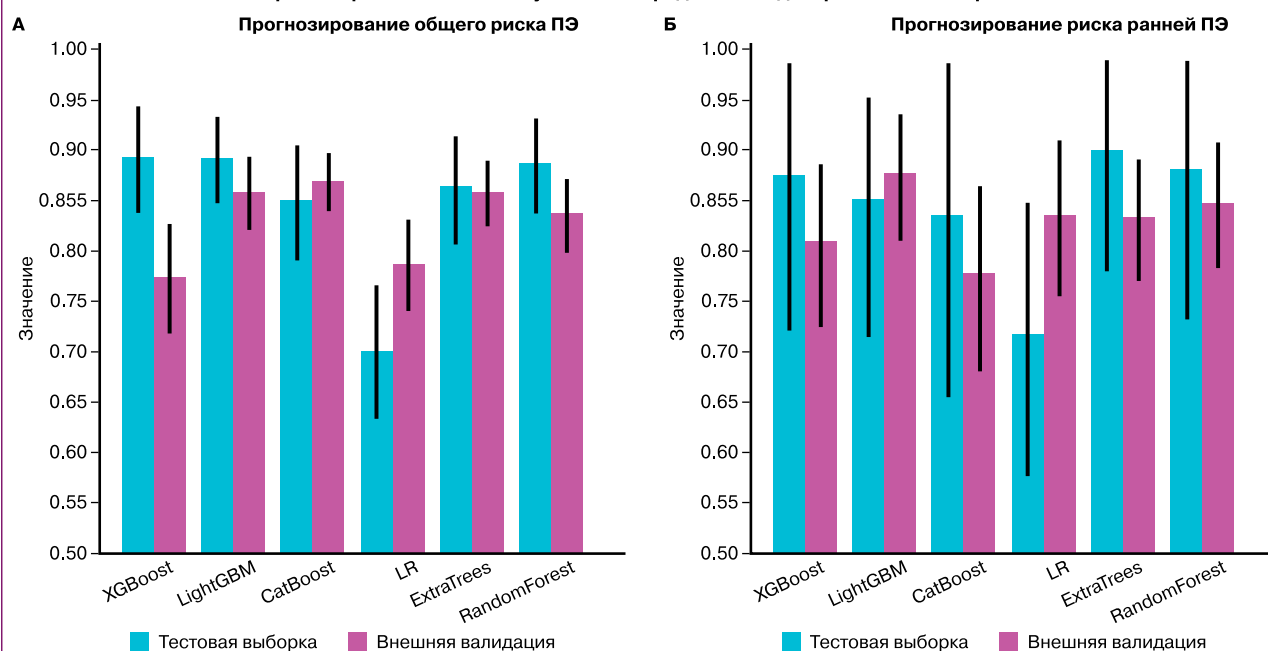


Рис. 4. ROC-кривые с 95% ДИ, полученные по результатам прогнозирования на наборах данных для внешней валидации.
А: прогнозирование общего риска преэклампсии, Б: прогнозирование риска ранней преэклампсии

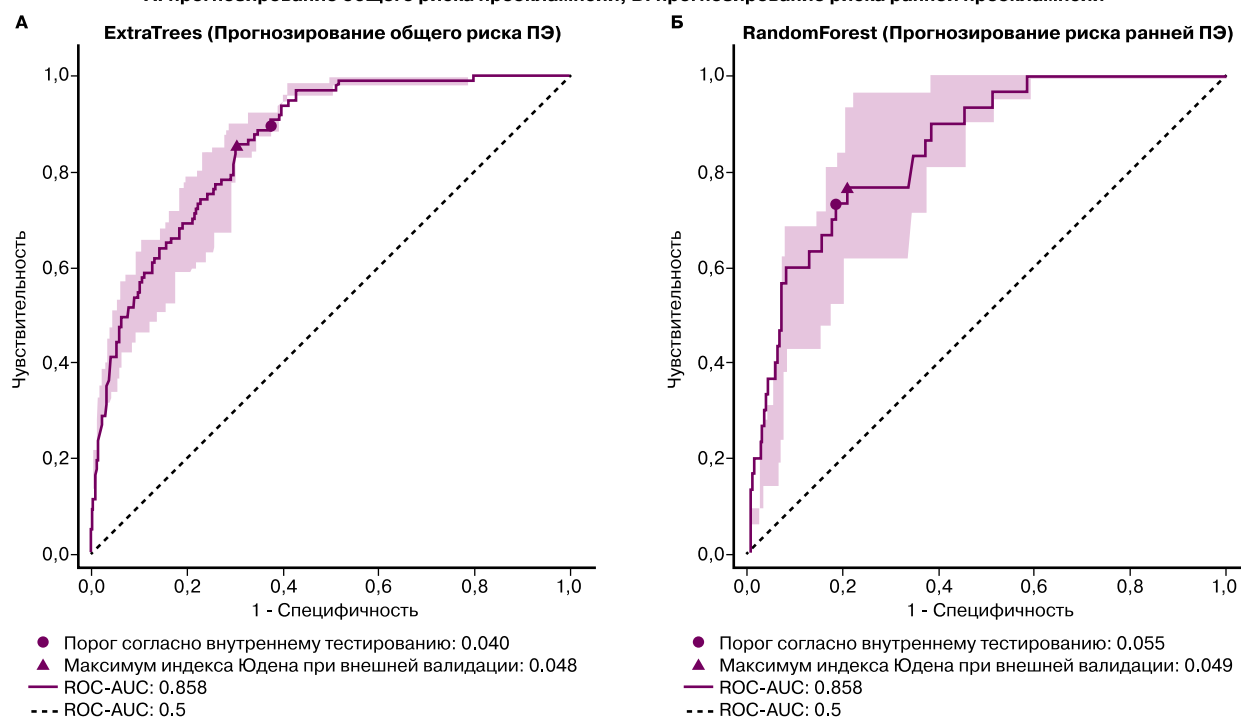


Таблица 3. Метрики качества разработанных моделей при проведении внутреннего тестирования (95% доверительные интервалы метрик указаны в квадратных скобках)

Модель	AUC	Точность	Чувствит.	Специф.	Прогн. ценность полож. класса	Прогн. ценность отриц. класса	F1 (полож.)	F1 (отриц.)
Общий риск развития преэклампсии								
XGBoost	0.892 [0.831–0.94]	0.877 [0.862–0.891]	0.821 [0.724–0.908]	0.879 [0.864–0.894]	0.204 [0.16–0.25]	0.992 [0.987–0.996]	0.326 [0.264–0.389]	0.932 [0.924–0.941]
LightGBM	0.892 [0.85–0.929]	0.879 [0.863–0.893]	0.731 [0.62–0.828]	0.884 [0.869–0.898]	0.192 [0.141–0.238]	0.989 [0.983–0.994]	0.304 [0.232–0.364]	0.934 [0.925–0.942]
CatBoost	0.85 [0.795–0.901]	0.82 [0.802–0.837]	0.761 [0.651–0.864]	0.822 [0.804–0.84]	0.139 [0.105–0.176]	0.989 [0.984–0.994]	0.234 [0.183–0.289]	0.898 [0.887–0.908]
LR	0.703 [0.64–0.764]	0.888 [0.874–0.902]	0.403 [0.293–0.522]	0.906 [0.893–0.92]	0.139 [0.097–0.194]	0.976 [0.968–0.982]	0.207 [0.147–0.276]	0.94 [0.932–0.948]
ExtraTrees	0.862 [0.8–0.914]	0.773 [0.754–0.793]	0.821 [0.717–0.913]	0.771 [0.751–0.791]	0.119 [0.091–0.15]	0.991 [0.986–0.996]	0.208 [0.161–0.257]	0.868 [0.855–0.88]
RandomForest	0.886 [0.84–0.928]	0.834 [0.817–0.851]	0.821 [0.722–0.906]	0.834 [0.817–0.851]	0.157 [0.12–0.197]	0.992 [0.987–0.996]	0.264 [0.207–0.318]	0.906 [0.896–0.917]
Риск развития ранней преэклампсии								
XGBoost	0.875 [0.722–0.99]	0.965 [0.957–0.974]	0.688 [0.428–0.923]	0.968 [0.96–0.976]	0.162 [0.077–0.255]	0.997 [0.994–0.999]	0.262 [0.132–0.379]	0.982 [0.978–0.987]
LightGBM	0.85 [0.729–0.954]	0.792 [0.773–0.81]	0.812 [0.609–1.0]	0.792 [0.772–0.81]	0.034 [0.017–0.052]	0.998 [0.995–1.0]	0.065 [0.034–0.098]	0.883 [0.871–0.894]
CatBoost	0.837 [0.673–0.986]	0.885 [0.87–0.899]	0.75 [0.529–0.938]	0.886 [0.871–0.9]	0.056 [0.026–0.088]	0.997 [0.995–0.999]	0.104 [0.05–0.16]	0.939 [0.93–0.947]
LR	0.721 [0.582–0.841]	0.795 [0.777–0.814]	0.5 [0.25–0.769]	0.798 [0.78–0.817]	0.022 [0.008–0.038]	0.994 [0.99–0.998]	0.042 [0.015–0.071]	0.885 [0.874–0.897]
ExtraTrees	0.90 [0.772–0.99]	0.891 [0.876–0.905]	0.875 [0.667–1.0]	0.892 [0.876–0.905]	0.068 [0.035–0.105]	0.999 [0.997–1.0]	0.126 [0.067–0.188]	0.942 [0.933–0.95]
RandomForest	0.88 [0.74–0.988]	0.909 [0.894–0.923]	0.812 [0.6–1.0]	0.909 [0.895–0.923]	0.075 [0.04–0.121]	0.998 [0.996–1.0]	0.137 [0.074–0.213]	0.952 [0.944–0.959]

Таблица 4. Метрики качества разработанных моделей при проведении внешней валидации (95% доверительные интервалы метрик указаны в квадратных скобках)

Модель	AUC	Точность	Чувствит.	Специф.	Прогн. ценность полож. класса	Прогн. ценность отриц. класса	F1 (полож.)	F1 (отриц.)
Общий риск развития преэклампсии								
XGBoost	0.775 [0.72–0.825]	0.808 [0.792–0.822]	0.608 [0.51–0.703]	0.815 [0.8–0.83]	0.112 [0.085–0.142]	0.982 [0.976–0.987]	0.189 [0.147–0.233]	0.891 [0.881–0.9]
LightGBM	0.858 [0.821–0.89]	0.8 [0.785–0.815]	0.742 [0.655–0.826]	0.802 [0.787–0.818]	0.126 [0.101–0.154]	0.988 [0.983–0.992]	0.215 [0.176–0.257]	0.885 [0.876–0.895]
CatBoost	0.87 [0.837–0.897]	0.693 [0.676–0.711]	0.897 [0.831–0.951]	0.685 [0.667–0.704]	0.098 [0.079–0.118]	0.994 [0.99–0.997]	0.177 [0.145–0.21]	0.811 [0.799–0.825]
LR	0.788 [0.741–0.834]	0.844 [0.829–0.857]	0.495 [0.39–0.6]	0.857 [0.844–0.871]	0.117 [0.086–0.15]	0.978 [0.972–0.984]	0.189 [0.142–0.236]	0.914 [0.905–0.922]
ExtraTrees	0.858 [0.827–0.89]	0.634 [0.616–0.652]	0.897 [0.837–0.953]	0.624 [0.606–0.643]	0.084 [0.067–0.102]	0.994 [0.99–0.997]	0.153 [0.125–0.183]	0.767 [0.753–0.781]
RandomForest	0.836 [0.798–0.872]	0.704 [0.686–0.721]	0.753 [0.667–0.838]	0.702 [0.684–0.72]	0.088 [0.069–0.106]	0.987 [0.981–0.992]	0.158 [0.126–0.188]	0.82 [0.807–0.832]
Риск развития ранней преэклампсии								
XGBoost	0.806 [0.709–0.886]	0.931 [0.922–0.941]	0.4 [0.227–0.593]	0.938 [0.929–0.948]	0.071 [0.034–0.111]	0.992 [0.989–0.996]	0.12 [0.06–0.183]	0.964 [0.959–0.97]
LightGBM	0.878 [0.806–0.935]	0.704 [0.686–0.721]	0.9 [0.781–1.0]	0.702 [0.684–0.719]	0.034 [0.023–0.048]	0.998 [0.996–1.0]	0.066 [0.045–0.0]	0.824 [0.812–0.836]
CatBoost	0.777 [0.673–0.867]	0.835 [0.821–0.849]	0.6 [0.407–0.786]	0.838 [0.824–0.851]	0.042 [0.024–0.062]	0.994 [0.991–0.998]	0.079 [0.045–0.115]	0.91 [0.901–0.918]
LR	0.834 [0.762–0.9]	0.713 [0.696–0.73]	0.733 [0.567–0.881]	0.713 [0.696–0.73]	0.029 [0.018–0.042]	0.996 [0.992–0.998]	0.056 [0.035–0.079]	0.831 [0.819–0.843]
ExtraTrees	0.833 [0.771–0.887]	0.863 [0.849–0.876]	0.533 [0.351–0.714]	0.867 [0.853–0.88]	0.045 [0.024–0.07]	0.994 [0.99–0.997]	0.083 [0.046–0.126]	0.926 [0.918–0.933]
RandomForest	0.848 [0.785–0.904]	0.813 [0.798–0.828]	0.733 [0.565–0.885]	0.814 [0.799–0.828]	0.045 [0.027–0.063]	0.996 [0.993–0.999]	0.084 [0.052–0.117]	0.896 [0.887–0.905]

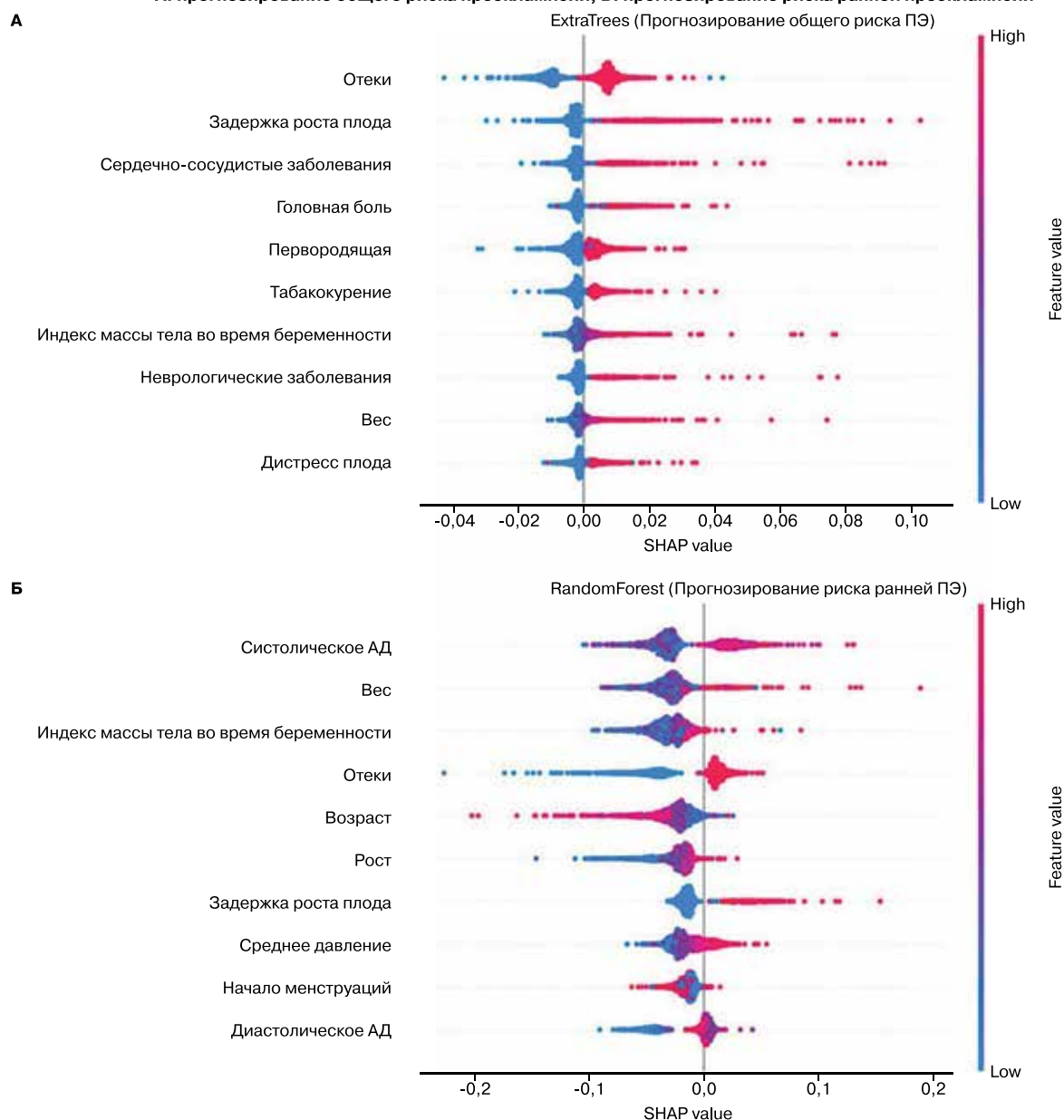
этой модели с порогом классификации 0,04 составила 0,634 (95% ДИ 0,616–0,652), чувствительность – 0,897 (95% ДИ 0,837–0,953), специфичность – 0,624 (95% ДИ 0,606–0,643).

Среди моделей прогнозирования развития ранней ПЭ наилучшие результаты показала модель на основе алгоритма RandomForest. По результатам внешней валидации при пороге бинаризации 0,055 были получены следующие метрики качества: AUC 0,848 (95% ДИ 0,785–0,904), точность 0,813 (95% ДИ 0,798–0,828), чувствительность 0,733 (95% ДИ 0,565–0,885), специфичность 0,814 (95% ДИ 0,799–0,828). Значения AUC, полученные на наборах для внешней и внутренней валидации для всех отобранных моделей, представлены на рисунке 3. ROC-кривые выбранных финальных моделей отражены на рисунке 4.

Значимость 10 ключевых признаков, вошедших в выбранные модели ExtraTrees и RandomForest показана на рисунке 5. Средняя абсолютная ошибка калибровочной кривой для выбранных финальных моделей прогнозирования ПЭ и ранней ПЭ равнялась 24,5% и 22,4%, соответственно.

Таким образом, разработанные модели МО позволяют автоматически анализировать обозначенные выше признаки из ЭМК пациенток, при условии, что срок беременности находится в диапазоне 7–16 недель, и выявлять высокий риск развития ПЭ, если результаты анализа превышают установленные пороги активации, а предусмотренная система «красных флагов» способна оповещать врача о высоком риске ПЭ и необходимости рассмотреть возможность профилактики.

Рис. 5. ТОП-10 значимых признаков финальных моделей.
А: прогнозирование общего риска преэклампсии, Б: прогнозирование риска ранней преэклампсии



Обсуждение

ПЭ является тяжелым осложнением беременных и угрожает здоровью и жизни матери и ребенка. В настоящее время патогенез ПЭ до конца неясен. Известно, что в развитии данного осложнения беременности участвуют материнские, фетальные, плацентарные и другие факторы, но ни один из них не может объяснить патогенез ПЭ и предсказать ее развитие [7, 16]. Во время ПЭ происходит нарушение плацентарной трофобластической инвазии, что приводит к недостатку кислорода в плаценте, высвобождению факторов воспаления эндотелия сосудов и их повреждению. Это вызывает системный спазм мелких сосудов и снижение перфузии в органах, приводя к их дисфункции, что в конечном случае чревато развитием эклампсии, отслойки плаценты и смерти матери и плода [2, 17]. Отдельного внимания заслуживает понятие ранней ПЭ с началом заболевания до 34 недели беременности ввиду наличия существенных клинико-патогенетических отличий от поздней ПЭ. Ранний дебют ПЭ является основным фактором, приводящим к перинатальной смертности [15]. Таким образом, раннее прогнозирование общего риска развития ПЭ и риска развития ранней ПЭ, профилактика и своевременное медицинское вмешательство особенно важны для снижения частоты данного осложнения и улучшения исходов беременности.

Созданы несколько многопараметрических моделей МО и проверена их точность. Максимальные показатели целевой метрики качества на собранных нами данных показали методы, основанные на деревьях решений. Важнейшее значение для итоговых моделей имеют факт выявления отеков, головной боли во время беременности, а также уровень АД. Значимыми признаками являлись конституциональные и антропометрические данные пациенток, такие как возраст, вес, рост и ИМТ во время беременности. Из анамнестических данных большую значимость для модели имели ССЗ и неврологические заболевания, задержка роста плода, а также возраст менархе.

Прогностическая ценность положительного класса у полученной модели оценки общего риска ПЭ при внутреннем тестировании варьировалась от 9% до 15%, а для прогнозирования ранней ПЭ — от 4 до 11%. По результатам внешней валидации было отмечено снижение этих показателей: 6–10% и 3–7%, соответственно. При этом разработанные нами инструменты обладали высокой чувствительностью и прогностической ценностью отрицательного результата (>99% у обеих моделей по результатам внешней валидации), что крайне важно для исключения в процессе диагностического поиска таких осложнений периода беременности и говорит о возможности эффективного использования моделей на этапе обследования в условиях женской консультации.

Полученные нами результаты сопоставимы с аналогичными разработками зарубежных авторов [18, 19]. В качестве предикторов в упомянутых работах использовался расширенный спектр материнских факторов, включавший такие специфические параметры, как плацентарный фактор роста, растворимая fms-подобная тирозинкиназа-1, ассоциированный с беременностью протеин-А плазмы и пульсационный индекс маточных артерий. Важной особенностью, которую необходимо учитывать при сравнении результатов моделей, является характеристика исследуемой когорты по сроку беременности. Так, в работе Marić I. et al. [18], в исследуемую группу вошли беременные на тех же сроках, что и в нашей работе, а в работе Tan M.Y. et al. [19] сроки беременности были ограничены 13⁺6 неделями. Совокупность упомянутых фактов говорит о наличии у разработанных нами моделей большого потенциала в качестве дополнительного инструмента раннего скрининга беременных, и с учетом продемонстрированной устойчивости моделей в рамках внешней валидации, данные модели могут быть допущены к практическому использованию в рамках проспективных исследований. Для дальнейшего повышения метрик точности необходимо увеличение размера обучающей выборки, в частности данных о прогнозируемом классе [20].

Методы МО являются эффективными инструментами в разработке моделей для прогнозирования и диагностики редких и многофакторных патологических процессов, таких, как ПЭ. При этом в качестве факторов используются рутинные клинико-лабораторные параметры, которые можно легко отслеживать и контролировать во время течения беременности. Создание эффективных медицинских прогностических моделей — комплексный и многоступенчатый процесс, связанный со сбором, обработкой и анализом большого количества информации, активным взаимодействием между врачами, статистиками и специалистами по работе с данными.

Заключение

Метрики полученных итоговых моделей соответствовали опубликованным ранее аналогам. Результаты внешней валидации показали относительную устойчивость моделей к новым данным, что в совокупности с показателями качества предполагает возможность их использования в реальной клинической практике. Это наш первый опыт прогнозирования сложного по механизму развития осложнения беременности на основе данных реальной клинической практики. Качество прогностической модели напрямую зависит от используемых данных и статистических алгоритмов, которые нам предстоит улучшить в будущих исследованиях.

ПРИЛОЖЕНИЕ

Итоговый список предикторов, использованных в разработке финальных версий моделей (в алфавитном порядке), и их допустимые значения

Название предиктора	Прогноз ПЭ (N=34)	Прогноз ранней ПЭ (N=36)	Допустимые значения [min, max]
Анурия/Олигурия		+	
Бесплодие в анамнезе	+		
Болезни мочевыделительной системы	+	+	
Боль в животе	+	+	
Вес, кг	+	+	[20, 200]
Вес до беременности, кг		+	[20, 200]
Возраст, лет	+	+	[10, 60]
Гемолиз	+		
Гестационный сахарный диабет	+	+	
Гибель плода в анамнезе		+	
Головная боль	+	+	
Диастолическое АД, мм рт.ст.	+	+	[30, 200]
Дистресс плода в анамнезе	+		
Задержка роста плода в анамнезе	+	+	
ИМТ во время беременности, кг/м ²	+	+	[15,50]
ИМТ до беременности, кг/м ²		+	[15,50]
Интергенетический интервал	+	+	
Креатинин крови, мкмоль/л	+	+	[10, 500]
Маловодие	+		
Мертворождение в анамнезе		+	
Многоплодие	+	+	
Нарушение зрения		+	
Начало менструаций, возраст, лет	+	+	[7, 16]
Неврологические заболевания	+	+	
Отеки	+	+	
Первородящая	+	+	
Плацентарная недостаточность в анамнезе	+	+	
Преэклампсия в анамнезе	+	+	
Рвота	+		
Рост, см	+	+	[100, 250]
Сахарный диабет	+	+	
Сердечно-сосудистые заболевания	+	+	
Систолическое АД, мм рт. ст.	+	+	[50, 250]
Среднее давление	+	+	
Срок беременности, недель	+	+	[7, 16]
Судороги		+	
Табакокурение	+	+	
Тошнота	+	+	
Тромбозы, тромбоземболии и тромбофлебиты	+	+	
Тромбоциты, ×10 ⁹ /л	+		[25, 1000]
Фибриноген, г/л		+	[1, 10]
Экстракорпоральное оплодотворение	+	+	

Литература/References

1. English F.A., Kenny L.C., McCarthy F.P. Risk factors and effective management of preeclampsia. *Integr. Blood Pressure Control*. 2015; 8: 7-12. <https://dx.doi.org/10.2147/IBPC.S50641>.
2. Jim B., Karumanchi S.A. Preeclampsia: pathogenesis, prevention, and long-term complications. *Semin. Nephrol.* 2017; 37(4): 386-97. <https://dx.doi.org/10.1016/j.semnephrol.2017.05.011>.
3. Bartsch E., Medcalf K.E., Park A.L., Ray J.G.; High Risk of Pre-eclampsia Identification Group. Clinical risk factors for pre-eclampsia determined in early pregnancy: systematic review and meta-analysis of large cohort studies. *BMJ*. 2016; 353: i1753. <https://dx.doi.org/10.1136/bmj.i1753>.
4. De Kat A.C., Hirst J., Woodward M., Kennedy S., Peters S.A. Prediction models for preeclampsia: a systematic review. *Pregnancy Hypertens.* 2019; 16: 48-66. <https://dx.doi.org/10.1016/j.preghy.2019.03.005>.
5. Thangaratinam S., Allotey J., Marlin N., Dodds J., Cheong-See F., von Dadelszen P. et al. Prediction of complications in early-onset preeclampsia (PREP): development and external multinational validation of prognostic models. *BMC Med.* 2017; 15(1): 68. <https://dx.doi.org/10.1186/s12916-017-0827-3>.
6. Zhang Y., Chen X.L., Chen W.M., Zhou H.B. Prognostic nomogram for the overall survival of patients with newly diagnosed multiple myeloma. *Biomed. Res. Int.* 2019; 2019: 5652935. <https://dx.doi.org/10.1155/2019/5652935>.
7. Chen W., Sun S. Clinical application of a multiparameter-based nomogram model in predicting preeclampsia. *Evid. Based Complement. Alternat. Med.* 2022; 2022: 7484112. <https://dx.doi.org/10.1155/2022/7484112>.
8. Moons K.G., Altman D.G., Reitsma J.B., Ioannidis J.P., Macaskill P., Steyerberg E. W. et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann. Intern. Med.* 2015; 162(1): W1-73. <https://dx.doi.org/10.7326/M14-0698>.
9. Ding Y., Simonoff J.S. An investigation of missing data methods for classification trees. *Econometrics: Data Collection & Data Estimation Methodology Journal*. Publ. 1 December 2006. <https://dx.doi.org/10.5555/1756006.1756012>.
10. Awais M., Shamsad F., Bae S. Towards an adversarially robust normalization approach. Cite as: arXiv:2006.11007 [cs.LG] (or arXiv:2006.11007v1 [cs.LG] for this version) . <https://dx.doi.org/10.48550/arXiv.2006.11007>.
11. Лучинин А.С. Искусственный интеллект в гематологии. Клиническая онкогематология. Фундаментальные исследования и клиническая практика. 2022; 15(1): 16-27. [Luchinin A.S. Artificial Intelligence in Hematology. *Clinical Oncohematology*. 2022; 15(1): 16-27. (in Russian)]. <https://dx.doi.org/10.21320/2500-2139-2022-15-1-16-27>.
12. Van Calster B., McLernon D.J., van Smeden M., Wynants L., Steyerberg E.W.; Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. *BMC Med.* 2019; 17(1): 230. <https://dx.doi.org/10.1186/s12916-019-1466-7>.
13. Zoubir A.M., Iskander D.R. Bootstrap methods and applications. *IEEE Signal Process. Mag.* 2007; 24(4): 10-9. <https://dx.doi.org/10.1109/MSP.2007.4286560>.
14. Lundberg S.M., Erion G., Chen H., DeGrave A., Prutkin J.M., Nair B. et al. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* 2020; 2(1): 56-67. <https://dx.doi.org/10.1038/s42256-019-0138-9>.
15. Министерство здравоохранения Российской Федерации. Преэклампсия. Эклампсия. Отеки, протеинурия и гипертензивные расстройства во время беременности, в родах и послеродовом периоде. Клинические рекомендации. М.; 2021. 79с. [Ministry of Health of the Russian Federation. Preeclampsia. Eclampsia. Edema, proteinuria and hypertensive disorders during pregnancy, childbirth and the postpartum period. Clinical guidelines. Moscow; 2021. 79p. (in Russian)].
16. Palomaki G.E., Haddow J.E., Haddow H.R., Salahuddin S., Geahchan C., Cerdeira A.S. et al. Modeling risk for severe adverse outcomes using angiogenic factor measurements in women with suspected preterm preeclampsia. *Prenat. Diagn.* 2015; 35(4): 386-93. <https://dx.doi.org/10.1002/pd.4554>.
17. March M.I., Geahchan C., Wenger J., Raghuraman N., Berg A., Haddow H. et al. Circulating angiogenic factors and the risk of adverse outcomes among haitian women with preeclampsia. *PloS One*. 2015; 10(5): e0126815. <https://dx.doi.org/10.1371/journal.pone.0126815>.
18. Marić I., Tsur A., Aghaeepour N., Montanari A., Stevenson D.K., Shaw G.M., Winn V.D. Early prediction of preeclampsia via machine learning. *Am. J. Obstet. Gynecol. MFM.* 2020; 2(2): 100100. <https://dx.doi.org/10.1016/j.ajogmf.2020.100100>.
19. Tan M.Y., Syngelaki A., Poon L.C., Rolnik D.L., O'Gorman N., Delgado J.L. et al. Screening for pre-eclampsia by maternal factors and biomarkers at 11-13 weeks' gestation. *Ultrasound Obstet. Gynecol.* 2018; 52(2): 186-95. <https://dx.doi.org/10.1002/uog.19112>.
20. Riley R.D., Debray T.P.A., Collins G.S., Archer L., Ensor J., van Smeden M., Snell K.I.E. Minimum sample size for external validation of a clinical prediction model with a binary outcome. *Stat. Med.* 2021; 40(19): 4230-51. <https://dx.doi.org/10.1002/sim.9025>.

Поступила 18.04.2023

Принята в печать 29.09.2023

Received 18.04.2023

Accepted 29.09.2023

Сведения об авторах:

Андрейченко Анна Евгеньевна, к.ф.-м.н., руководитель направления искусственного интеллекта, ООО «К-Скай», +7(916)321-25-70, aandreychenko@webiomed.ru, <https://orcid.org/0000-0001-6359-0763>, 185910, Россия, Петрозаводск, набережная Варкауса, д. 17.

Лучинин Александр Сергеевич, к.м.н., с.н.с. отдела организации и сопровождения научных исследований, Кировский научно-исследовательский институт гематологии и переливания крови ФМБА России, +7(919)506-87-86, luchinin@niigpk.ru, <https://orcid.org/0000-0002-5016-210X>, 610027, Россия, Киров, ул. Дерендяева, д. 84.

Ившин Александр Анатольевич, к.м.н., доцент, заведующий кафедрой акушерства и гинекологии, дерматовенерологии, Петрозаводский государственный университет, +7(909)567-12-51, scipeople@mail.ru, <https://orcid.org/0000-0001-7834-096X>, 185001, Россия, Петрозаводск, ул. Красноармейская, д. 31.

Новицкий Роман Эдвардович, генеральный директор, ООО «К-Скай», +7(911)400-50-00, roman@webiomed.ru, <https://orcid.org/0000-0002-2350-977X>, 185910, Россия, Петрозаводск, набережная Варкауса, д. 17.

Ермак Андрей Дмитриевич, аналитик данных направления искусственного интеллекта, ООО «К-Скай», +7 (977) 563-52-72, aermak@webiomed.ru, <https://orcid.org/0000-0002-0513-8557>, 185910, Россия, Петрозаводск, набережная Варкауса, д. 17.

Гусев Александр Владимирович, к.т.н., с.н.с. отдела научных основ организации здравоохранения, Центральный научно-исследовательский институт организации и информатизации здравоохранения Минздрава России, +7(911)402-35-00, agusev@webiomed.ai, <https://orcid.org/0000-0002-7380-8460>, 127254, Россия, Москва, ул. Добролюбова, д. 11.

Автор, ответственный за переписку: Александр Анатольевич Ившин, scipeople@mail.ru

Authors' information:

Anna E. Andreychenko, PhD, Head of the Artificial Intelligence Department, K-SkAI, +7(916)321-25-70, aandreychenko@webiomed.ru, <https://orcid.org/0000-0001-6359-0763>, 17, Varkaus Embankment, Petrozavodsk, 185901, Russia.

Alexander S. Luchinin, PhD, Senior Researcher, Department of Organization and Support of Scientific Research, Kirov Research Institute of Hematology and Blood Transfusion, Federal Medical Biological Agency of Russia, +7(919)506-87-86, luchinin@niigpk.ru, <https://orcid.org/0000-0002-5016-210X>, 84, Derendyaeva str., Kirov, 610027, Russia.

Alexander A. Ivshin, PhD, Associate Professor, Head of the Department of Obstetrics and Gynecology, Dermatovenerology of the Medical Institute, Petrozavodsk State University, +7(909)567-12-51, scipeople@mail.ru, <https://orcid.org/0000-0001-7834-096X>, 31, Krasnoarmeyskaya str., Petrozavodsk, 185035, Russia.

Andrey D. Ermak, PhD, Data analyst, Artificial Intelligence Department, K-SkAI, +7(977)563-52-72, aermak@webiomed.ru, <https://orcid.org/0000-0002-0513-8557>, 17, Varkaus Embankment, Petrozavodsk, 185901, Russia.

Roman E. Novitskiy, CEO, K-SkAI, +7(911)400-50-00, roman@webiomed.ru, <https://orcid.org/0000-0002-2350-977X>, 17, Varkaus Embankment, Petrozavodsk, 185901, Russia.

Alexander V. Gusev, PhD, Senior Researcher, Department of Scientific Fundamentals of Health Organization, Russian Research Institute of Health, +7(911)402-35-00, agusev@webiomed.ai, <https://orcid.org/0000-0002-7380-8460>, 11, Dobrolyubova str., Moscow, 127254, Russia.

Corresponding author: Alexander A. Ivshin, scipeople@mail.ru