

DOI: <https://doi.org/10.17650/1726-9784-2025-24-1-34-45>

Анализ электронных медицинских записей при помощи технологий искусственного интеллекта для определения групп скрининга рака легкого: систематический обзор клинических исследований

И.В. Самойленко¹, В.В. Назарова¹, З.Р. Магомедова¹, П.В. Кононец^{1,2}, И.М. Боровков¹, Т.Г. Геворкян¹

¹ФГБУ «Национальный медицинский исследовательский центр онкологии им. Н.Н. Блохина» Минздрава России; Россия, 115522 Москва, Каширское шоссе, 24;

²ФГБОУ ВО «Российский университет медицины» Минздрава России; Россия, 127006 Москва, ул. Долгоруковская, 4

Контакты: Игорь Вячеславович Самойленко i.samoylenko@ronc.ru

Введение. Рак легкого (РЛ) занимает 1-е место в структуре онкологической заболеваемости и смертности в мире, а доля впервые выявленных локализованных стадий не превышает 20 % общего числа заболевших. Ранняя диагностика и скрининг могут значительно улучшить результаты лечения, однако современные подходы требуют новых решений для более эффективного отбора пациентов.

Цель. Провести систематический обзор исследований по применению искусственного интеллекта (ИИ) для анализа социально-демографических данных и рутинных лабораторных тестов для отбора пациентов на скрининг.

Материалы и методы. Проведен поиск публикаций (2014–2024 гг.) в базах PubMed, ResearchGate, Scopus и др. Включены исследования, анализирующие использование ИИ для прогнозирования риска РЛ на основе социально-демографических данных и медицинских записей.

Результаты. Анализ позволил идентифицировать 5 исследований моделей на основе ИИ, которые были применены для отбора кандидатов на скрининг РЛ. Наиболее часто используемыми факторами в моделях ИИ являлись возраст, стаж курения, хронические заболевания легких и индекс массы тела, а сами модели демонстрировали высокую чувствительность (до 92,7 %) и площадь под кривой рабочей характеристики приемника (до 0,9).

Заключение. Применение ИИ может улучшить точность отбора пациентов для скрининга по сравнению с традиционными методами. Использование ИИ для прогнозирования риска развития РЛ имеет значительный потенциал, дополнительно раскрывающийся при сочетании анализа социально-демографических данных и медицинских записей. Необходимы дальнейшие исследования для улучшения моделей и оценки их влияния на клиническую практику.

Ключевые слова: рак легкого, искусственный интеллект, ранняя диагностика, скрининг, медицинская запись

Для цитирования: Самойленко И.В., Назарова В.В., Магомедова З.Р. и др. Анализ электронных медицинских записей при помощи технологий искусственного интеллекта для определения групп скрининга рака легкого: систематический обзор клинических исследований. Российский биотерапевтический журнал 2025;24(1):34–45. DOI: <https://doi.org/10.17650/1726-9784-2025-24-1-34-45>

Analysis of electronic medical records using artificial intelligence technologies for lung cancer screening group identification: a systematic review of clinical studies

Igor V. Samoylenko¹, Valeria V. Nazarova¹, Zakhra R. Magomedova¹, Pavel V. Kononets^{1,2}, Ivan M. Borovkov¹, Tigran G. Gevorkyan¹

¹N.N. Blokhin National Medical Research Center of Oncology, Ministry of Health of Russia; 24 Kashirskoe Shosse, Moscow 115522, Russia;

²Russian University of Medicine, Ministry of Health of Russia; 4 Dolgorukovskaya St., Moscow 127006, Russia

Contacts: Igor Vyacheslavovich Samoylenko i.samoylenko@ronc.ru

Background. Lung cancer remains one of the leading causes of mortality, with early detection significantly improving prognosis. Modern approaches require new solutions for more effective screening patient selection.

Aim. To conduct a systematic review of studies applying artificial intelligence (AI) for analyzing socio-demographic data and routine laboratory tests to optimize patient selection for screening and pathology classification.

Materials and methods. A literature search (2014–2024) was conducted in databases including PubMed, ResearchGate, and Scopus. Included studies analyzed the use of AI for lung cancer risk prediction based on socio-demographic data and medical records.

Results. The analysis identified 5 studies of AI-based models that were applied to select candidates for lung cancer screening. Age, smoking, chronic lung disease, and BMI were the most frequently used factors in the AI models. The models demonstrated high sensitivity (up to 92.7 %) and area under the receiver operating characteristic (up to 0.90). The results confirmed that AI can improve the accuracy of patient selection for screening compared to traditional methods.

Conclusion. AI application for lung cancer risk prediction shows substantial potential, especially with combined use of socio-demographic and medical record data. Further studies are needed to improve models and evaluate their clinical impact.

Keywords: lung cancer, artificial intelligence, early diagnosis, screening, medical records

For citation: Samoylenko I.V., Nazarova V.V., Magomedova Z.R. et al. Analysis of electronic medical records using artificial intelligence technologies for lung cancer screening group identification: a systematic review of clinical studies. *Rossijskij bioterapevticeskij zurnal = Russian Journal of Biotherapy* 2025;24(1):34–45. (In Russ.). DOI: <https://doi.org/10.17650/1726-9784-2025-24-1-34-45>

Введение

Рак легкого (РЛ) остается одной из ведущих причин смерти от злокачественных новообразований (ЗНО) в Российской Федерации, США, Европе и Азии [1]. Несмотря на активную разработку методов профилактики, ранней диагностики и лечения РЛ, прогноз заболевания по-прежнему остается неудовлетворительным, а доля пациентов с I–II стадией заболевания не превышает 20 % [2, 3]. Ранняя диагностика РЛ может стать ключом к успешному лечению, так как повышает шансы на увеличение безрецидивной и общей выживаемости.

Попытки структуризации факторов риска и разработки новых методов первичной профилактики и скрининга РЛ активно предпринимались на протяжении второй половины XX в. При этом в большинстве стран данные подходы основаны преимущественно на проведении компьютерной томографии (КТ) или рентгенографии органов грудной клетки [4]. Организация массового и регулярного скрининга связана с рядом неизбежных сложностей, вынуждающих медицинское сообщество искать новые пути оптимизации диагностического процесса и сужать группы риска путем включения в прогностические модели большего числа легкодоступных параметров из рутинных медицинских записей или анализов. Вместе с тем, несмотря на указанные сложности, революционные достижения в области технологий искусственного интеллекта (ИИ) и появление новых способов обработки большого количества мультимодальных данных сделали подобную работу решаемой задачей [5].

Применение ИИ в медицине показывает в последние годы значительный потенциал по улучшению диагностических и прогностических алгоритмов. Однако его интеграция для раннего выявления рака требует учета социально-демографических и медицинских данных, что является сложной и недостаточно изученной задачей. Многочисленные исследования подтверждают, что применение ИИ и анализ больших данных могут выявлять скрытые закономерности и делать точные прогнозы, недоступные при традиционных методах исследования. В контексте онкологии, где время – критический фактор, ИИ может стать важным инструментом раннего выявления рисков и предотвращения прогрессирования заболевания. Программное обеспечение, созданное на базе ИИ, может стать ключевым инструментом в прогнозировании рисков и ранней диагностике, как уже демонстрируют примеры “IBM Watson for Oncology”, “Google DeepMind” и “PathAI”. Эти технологии показывают, что ИИ способен повысить точность диагностики, оптимизировать медицинские процессы и улучшить результаты лечения. Ожидается, что персонализированные планы скрининга, учитывающие индивидуальные факторы риска (возраст, пол, генетику, образ жизни и т.д.), помогут улучшить скрининг за счет оптимизации ресурсов для пациентов с высоким риском, сокращения ненужных процедур и затрат, повышения точности диагностики путем автоматизированного анализа данных, поддержки принятия решений врачами, улучшения доступности скрининговых программ и других параметров, что способствует раннему выявлению заболеваний,

снижению смертности и более рациональному использованию ресурсов [5].

Сегодня обширный пласт работ направлен на решение проблемы анализа изображений при помощи технологий ИИ (так называемых систем автоматизированного проектирования – computer-aided design), позволяющих значительно упростить дифференциацию рентгенологических изменений на высоко- и низкорисковые в отношении раннего РЛ [6–8]. В то же время настоящее исследование акцентирует внимание в большей мере на работах, изучающих несколько другой процесс: формирование групп риска развития РЛ. Действительно ли технологии ИИ помогут уточнить имеющиеся предиктивные модели?

В рамках данного обзора проведено детальное изучение работ, посвященных анализу социально-демографических данных и рутинных лабораторных тестов. Анализ таких исследований необходим для создания современных моделей, способных эффективно идентифицировать пациентов с высоким онкологическим риском и своевременно направлять их на скрининг. Подобное сочетание подходов помогает оптимизировать процессы диагностики, минимизировать ненужные обследования и увеличить эффективность раннего выявления РЛ.

Цель исследования – провести систематический обзор актуальных исследований, посвященных применению моделей ИИ, использующих социально-демографические данные, структурированные электронные медицинские записи и результаты рутинных лабораторных тестов, для улучшения ранней диагностики и скрининга РЛ.

Материалы и методы

Критерии включения/невключения исследований в анализ

В настоящий обзор включены исследования, посвященные использованию моделей ИИ для анализа медицинских изображений (таких как компьютерные томограммы и рентгенограммы) различной патологии легких. Будут также рассмотрены работы, применяющие ИИ для анализа социально-демографических данных и записей медицинских карт, включая сведения о ранее установленных диагнозах, симптомах и описаниях врачебных осмотров. Включению подлежат исследования, использующие данные о потенциальных факторах риска, таких как результаты анкетирования о стаже курения, а также рутинные лабораторные тесты для отбора пациентов на скрининг РЛ. Особое внимание будет уделено работам, которые объединяют анализ изображений или их описание и классификацию человеком, например, по системе LUNG-RADS, с социально-демографическими и медицинскими данными, а также сведениями о потенциальных факторах риска.

В обзор включены исследования, опубликованные в период с 2014 по 2024 г., представленные на любом языке и доступные в полном тексте.

Стратегия поиска литературы

С 01.05.2024 по 01.09.2024 проведен комплексный поиск публикаций в нескольких электронных базах данных, включая PubMed, ResearchGate, Scopus, Web of Science и IEEE Xplore. Целью поиска было выявление исследований, посвященных применению ИИ для ранней диагностики и скрининга РЛ. Поиск запросы содержали следующие ключевые слова: “application of artificial intelligence for lung cancer screening based on database information”, “artificial intelligence and medical records data and lung cancer and screening”, “artificial intelligence-based lung cancer risk prediction using electronic health records”, “selection criteria for lung cancer screening” (“big data” or “machine learning” or “deep learning” or “artificial intelligence” or “AI” or “radiomics”) and (“cancer screening” or “tumor screening” or “neoplasm screening” or “oncology screening” or “early diagnosis”) and (“risk stratification” or “population stratification” or “targeted screening”), “AI-based early detection of lung cancer”, “machine learning in lung cancer screening”, “predictive analytics for lung cancer”.

Производили также фильтрацию по годам (с 2014 по 2024 г.) и типу публикаций (только рецензируемые статьи). Первичный отбор проводили по заголовкам и аннотациям, после чего выполняли более детальный анализ полнотекстовых статей. Для повышения полноты обзора вручную просмотрены списки литературы в релевантных исследованиях, рекомендациях и обзорах. Этот этап проводили систематически для выявления дополнительных источников. Кроме того, включали публикации, в которых обсуждалось прогнозирование риска развития РЛ на основе данных электронных медицинских карт, социально-демографических и рутинных лабораторных данных и использование методов ИИ для их обработки. Работы, которые включали только анализ изображений (компьютерных томограмм, рентгенограмм) и классификацию изменений в легких, не включали в анализ. На этапе предварительного анализа все найденные публикации проверены на дублирование и оценены по установленным критериям качества (Preferred reporting items for systematic reviews and meta-analyses – «Предпочтительные элементы отчетности для систематических обзоров и метаанализов»). Оценку качества включаемых публикаций и соответствие критериям проводили два независимых исследователя (ВВН и ЗРМ). Дублирующиеся публикации исключали из дальнейшего анализа.

Извлечение данных

Разработана стандартная форма для извлечения данных из публикаций для оценки эффективности

технологий ИИ в ранней диагностике или оптимизации скрининга РЛ:

- Автор/год: краткая библиографическая ссылка.
- Анализируемые данные: описание данных, которые используются в исследовании (например, КТ-сканы, медицинские записи, социально-демографические данные).
- Потенциальная доступность данных в российских когортах пациентов: оценка возможности использования аналогичных данных в исследуемых когортах в Российской Федерации.
- Тип ИИ: категория ИИ, использованного в исследовании (например, машинное обучение, глубокое обучение).
- Название ИИ-алгоритма: конкретное наименование алгоритма ИИ (например, CNN, Random Forest).
- Дизайн: дизайн исследования (например, ретроспективное, проспективное, рандомизированное, нерандомизированное).
- Особенности популяции: краткое описание характеристик исследуемой группы (например, возраст, пол, факторы риска).
- Число ЗНО в исследуемой группе: число случаев ЗНО в выборке.
- Число обследованных в группе: общее число участников исследования.
- Чувствительность: показатель чувствительности изучаемого алгоритма (в сравнении с показателем для «золотого стандарта», если таковой приводится).
- Специфичность: показатель специфичности изучаемого алгоритма (в сравнении с показателем для «золотого стандарта», если таковой приводится).
- AUC: площадь под кривой рабочей характеристики приемника «чувствительность—специфичность» для изучаемого алгоритма (в сравнении с показателем для «золотого стандарта», если таковой приводится).
- NPV (negative predictive value): отрицательная прогностическая ценность изучаемого алгоритма (в сравнении с показателем для «золотого стандарта», если таковой приводится).
- PPV (positive predictive value): положительная прогностическая ценность изучаемого алгоритма (в сравнении с показателем для «золотого стандарта», если таковой приводится).
- Комментарий: дополнительные примечания, важные для понимания результатов исследования. Вывод авторов о вкладе технологии ИИ по сравнению с группой контроля.
- Веб-ссылка: ссылка на исследование для быстрой проверки данных.
- Страна: страна, в которой было проведено исследование.

Анализ данных

Качественный анализ использовали для описательного анализа извлеченных данных. Он включает сравнение различных подходов и методов применения ИИ в анализе изображений (например, КТ-сканов и рентгенограмм) и данных, таких как социально-демографическая информация и результаты рутинных лабораторных тестов. Качественный анализ позволит выявить основные тенденции, сильные и слабые стороны каждого подхода, а также определить факторы, влияющие на успешность применения ИИ в скрининге и диагностике РЛ. Для структурирования данных будет использована тематическая классификация, позволяющая группировать подходы по общим признакам:

1. Тип данных для анализа:

- социально-демографические данные: исследования, использующие данные о возрасте, поле, семейной истории заболеваний, образе жизни и других социальных факторах для прогнозирования риска;
- лабораторные тесты: исследования, анализирующие результаты рутинных лабораторных тестов (например, биохимические показатели крови, маркеры воспаления) для предсказания риска;
- комбинированные данные: исследования, использующие сочетание визуальных данных и социально-демографических или лабораторных показателей для повышения точности прогнозирования.

2. Метод ИИ:

- машинное обучение: использование традиционных методов машинного обучения, таких как Random Forest, SVM (support vector machines);
- глубокое обучение: применение нейронных сетей, таких как CNN (convolutional neural networks), RNN (recurrent neural networks);
- гибридные подходы: сочетание методов машинного и глубокого обучения для достижения более точных результатов.

3. Цель исследования:

- прогнозирование риска и отбор пациентов для проведения скрининга: исследования, направленные на определение вероятности развития РЛ у пациентов;
- классификация патологий легкого: исследования, занимающиеся различением доброкачественных образований и ЗНО.

4. Дизайн исследования:

- ретроспективные: анализ данных, собранных в прошлом;
- проспективные: исследования, в которых данные собираются в процессе исследования;
- рандомизированные и нерандомизированные: с учетом особенностей отбора участников.

Результаты

Поиск литературы

При использовании ключевых слов были получены следующие результаты:

- “application of artificial intelligence for lung cancer screening based on database information” – найдено 14 публикаций;
- “artificial intelligence and medical records data and lung cancer and screening” – 9 публикаций;
- “artificial intelligence-based lung cancer risk prediction using electronic health records” – 7 публикаций;
- “selection criteria for lung cancer screening” – 6 публикаций;
- (“‘big data’ or ‘machine learning’ or ‘deep learning’ or ‘artificial intelligence’ or ‘AI’ or ‘radiomics’) and (‘cancer screening’ or ‘tumor screening’ or ‘neoplasm screening’ or ‘oncology screening’ or ‘early diagnosis’) and (‘risk stratification’ or ‘population stratification’ or ‘targeted screening’)” – 4 публикации.

По результатам поиска, идентификации и отбора исследований для обзора найдено 40 записей. Из них 5 записей удалены как дубликаты на этапе предварительного просмотра, после чего 35 записей прошли проверку. Из проверенных записей 18 были исключены на стадии скрининга. В результате для извлечения данных осталось 17 отчетов. Последние оценены на предмет приемлемости, и 12 из них были исключены по различным причинам – 5 систематических обзоров, не оценивающих точность диагностики, 4 отчета низкого методологического качества, 3 отчета с описанием моделей, не включавших технологию ИИ, и 3 дубликата. В итоге в систематический обзор включено 5 исследований (рис. 1). Следует отметить, что поиск в дополнительных к PubMed базах данных литературных источников не добавил новых публикаций к уже выявленным, потому результаты поисков по остальным базам данных не приводятся.

Включенные в обзор исследования были проведены в США, Норвегии, Тайване и Израиле, и их



Рис. 1. Методология поиска научных публикаций

Fig. 1. Scientific publications survey methodology

результаты опубликованы на английском языке. Оценку точности применения ИИ для скрининга проводили по следующим показателям: чувствительность, специфичность, PPV, NPV и AUC. Показатель NPV отсутствовал во всех клинических исследованиях, поэтому его не включили в дальнейшее рассмотрение.

Математические модели расчета индивидуального риска развития рака легкого, доступные онлайн для самостоятельной оценки риска

В настоящее время по данным хорошо спланированных когортных исследований разработано достаточное количество калькуляторов риска, валидированных на большой популяции пациентов (табл. 1) [9–14]. Такие калькуляторы доступны для самостоятельного заполнения любым пользователям (в том числе без медицинского образования) и активно используются как для демонстрации расчетной вероятности развития РЛ, так и для разработки потенциальных мер по модификации подобного риска. В большинстве уже разработанных моделей на основе многофакторного анализа принимаются во внимание очень похожие ключевые факторы риска, которые также включаются и в модели с ИИ.

В исследовании P.V. Bach и соавт. использована многофакторная модель оценки риска, основанная на данных из исследования CARET [9]. Были включены 18 172 участника, преимущественно курильщики с разным уровнем риска в зависимости от стажа курения, возраста и воздействия асбеста. Исследова-

ние носило ретроспективный характер с внутренней валидацией, сравнивающей прогнозируемые и фактические случаи РЛ. Модель показала хорошую калибровку, а риск развития РЛ в популяции «тяжелых» и «бывших легких» курильщиков составил 15 и 0,8 % соответственно в течение 10 лет наблюдений [9].

В исследовании M.C. Tammemägi и соавт. [10] в 2013 г. модифицированная модель PLCO [15] для оценки риска развития РЛ адаптирована для данных исследования National Lung Screening Trial (NLST) [16] с участием 80 375 человек, куривших когда-либо. Проспективное исследование сравнивало модель PLCO(M2012) с критериями NLST, достигая AUC 0,803 в обучающей выборке и 0,797 – при валидации. Чувствительность составила 83 %, а положительная прогностическая ценность – 4 %, что на 41,3 % снизило число пропущенных случаев РЛ по сравнению с критериями NLST, демонстрируя преимущества модели PLCO(M2012) [10]. Указанная исследовательская группа позднее провела проспективное исследование в Канаде с участием 2537 курильщиков в возрасте 50–75 лет без предшествующего РЛ. Участники были отобраны на основе модели PanCan с 2 % риска развития рака за 6 лет. Хотя показатель AUC и не был детализирован, но доля раннего выявления была значительной (77 % на стадиях I–II). Частота выявленных случаев была выше, чем в NLST, что подчеркивает эффективность применяемой модели [17].

В 2016 г. H.A. Katki и соавт. провели собственную работу с использованием риск-адаптированных моделей на данных PLCO и NHIS [18] с участием активных

Таблица 1. Математические модели расчета индивидуального риска развития рака легкого, доступные онлайн для самостоятельной оценки риска

Table 1. Mathematical models for calculating individual risk of developing lung cancer, available online for self-assessment

Библиографическая ссылка Reference bibliographic	Веб-ссылка на онлайн калькулятор Web link to online calculator
P.V. Bach et al. (2003) [9]	https://www.evidencio.com/models/show/990
M.C. Tammemägi et al. (2013) [10]	https://shouldiscreen.com/English/lung-cancer-risk-calculator https://www.msmanuals.com/professional/multimedia/clinical-calculator/lung-cancer-risk-estimation-in-current-and-past-smokers-6-year
H.A. Katki et al. (2016) [11], L.C. Cheung et al. (2019) [12]	https://screenlc.com/dpp-vue/index.html
M. Markaki et al. (2018) [13]	https://www.omnicalculator.com/health/lung-cancer-risk-calculator-for-smokers
J.K. Field et al. (2021) [14]	https://liverpoollungproject.org.uk/MLRV3/MLRCalculation.html
Не приводится Not cited	https://www.mycanceriq.ca/Assessment/LUN#section3

курильщиков 50–80 лет [11]. Исследование было ретроспективным с внешней валидацией на данных NHIS. AUC для модели оценки риска смерти составил 0,78. Риск-адаптированный подход снизил число пациентов, которых нужно было просканировать для предотвращения одной смерти, на 17 % по сравнению с рекомендациями USPSTF [19], поддерживая целесообразность целевого скрининга КТ [11].

В 2019 г. L.C. Cheung и соавт. исследовали когорту с участием 130 964 человек из NHIS, который включает данные 60 млн курильщиков в США. Используя калиброванную и валидированную модели смертности среди постоянно курящих людей в США в возрасте от 40 до 84 лет и выбрав 8,3 млн постоянно курящих людей, чтобы соответствовать числу, выбранному по критериям USPSTF в 2013–2015 гг., авторам удалось показать, что отбор пациентов на скрининг на основе представлений о приобретенных годах жизни увеличит общую ожидаемую продолжительность жизни при выполнении КТ (633 400 против 607 800 лет), но предотвратит меньше смертей от РЛ (52 600 против 55 000) по сравнению с отбором на основе риска смерти. Таким образом, отбор пациентов для прохождения скрининга на основе модели о приобретенных годах жизни может максимизировать преимущества скрининга РЛ среди населения США за счет включения постоянно курящих людей, которые имеют как высокий риск РЛ, так и большую ожидаемую продолжительность жизни [12].

М. Markaki и соавт. в 2018 г. выполнили проспективное исследование в Норвегии с участием 65 237 человек старше 20 лет, со средним периодом наблюдения 15,2 года. Модель была протестирована на внешней выборке из 45 341 человека и показала AUC 0,87. Было выявлено, что 81,85 % случаев РЛ можно обнаружить, скринируя только 22 % курильщиков. Модель продемонстрировала высокую точность и практичность для долгосрочного прогнозирования [13].

В 2021 г. J.K. Field и соавт. опубликовали результаты тестирования модели LLPv3. Данная модель разработана на основе результатов исследования «случай–контроль» в Ливерпуле и протестирована на выборке из 75 958 человек с 5-летним периодом наблюдения. Сравнение версий LLPv2 и LLPv3 показало улучшенную калибровку последней с пропорциональной разницей в 22 %. Обе версии достигли показателя AUC 0,81. Модель LLPv3 обеспечила более точное прогнозирование абсолютного риска РЛ, что полезно для запуска программ скрининга в Великобритании [14]. При этом подобные модели включали такие факторы повышенного риска, как возраст, стаж курения, пол, наличие хронических заболеваний легких, индекс массы тела, социально-экономический статус, уровень образования, расу и этническую принадлежность. Некоторые модели включали симптомы,

такие как кашель, боль в груди и одышка, лабораторные данные, в том числе уровень С-реактивного белка. Данные из медицинских записей, содержащие информацию о предыдущих диагнозах, результатах обследований и приеме лекарств, также были важными компонентами изучаемых моделей.

Извлечение данных

Основные показатели эффективности технологий искусственного интеллекта для ранней диагностики или скрининга рака легкого по данным структурированных электронных медицинских записей

Мы извлекли данные из полных текстов опубликованных работ (табл. 2) и сделали их краткую характеристику. К сожалению, ни в одном из исследований не оценивали влияние изучаемых технологий на такие показатели, как снижение смертности и экономическое влияние на систему здравоохранения. В большинстве исследований приводятся данные аналитической точности новых методов, такие как чувствительность, специфичность и показатели AUC чувствительность/специфичность. Не все исследования имели непосредственную проспективную группу сравнения.

В исследовании X. Wang и соавт., опубликованном в 2019 г., модель на основе алгоритма XGBoost использовала данные из электронных медицинских карт и достигла показателя AUC 0,881. Согласно полученным результатам, модель продемонстрировала наибольший онкологический риск у активных курильщиков и пациентов с выраженной сопутствующей патологией. Таким образом, ИИ показал высокую точность в прогнозировании риска развития РЛ и был особенно полезен для оценки риска у пациентов со множественными хроническими заболеваниями [20].

В исследовании M.C. Yeh и соавт., опубликованном в 2021 г., построена прогностическая модель скрининга РЛ на основе нейронных сетей с использованием данных медицинских карт пациентов за последние 3 года. Были достигнуты показатели AUC 0,9 для общей популяции и 0,87 – для пациентов старше 55 лет, что свидетельствует о высокой точности рассматриваемой модели. Положительное прогностическое значение (PPV) достигло 14,3 % по отношению к пациентам с сопутствующими заболеваниями легких. Это свидетельствует о высокой предиктивной точности ИИ, особенно в группах с высоким риском, и подчеркивает эффективность его добавления к традиционным методам скрининга [21].

U. Chandran и соавт. в 2023 г. исследовали модель машинного обучения для прогнозирования риска развития РЛ на основе данных из электронных медицинских записей. Модель продемонстрировала значение AUC 0,76, что указывает на удовлетворительную точность. В качестве наиболее важных

Таблица 2. Извлеченные данные об эффективности технологий искусственного интеллекта (ИИ) для скрининга или ранней диагностики по данным электронных медицинских записей (ЭМЗ)
Table 2. Extracted data on the effectiveness of artificial intelligence (AI) technologies for screening or early diagnosis based on electronic medical records (EMR)

Автор/год/ссылка Author/year/reference	Источники данных Data sources	Тип ИИ Type of AI	Название ИИ алгоритма Name of AI algorithm	Дизайн Design features	Особенности популяции Features of the population	Чувствительность Sensitivity	Специфичность Specificity	Площадь под кривой рабочей характеристики приемника Area under the receiver operating characteristic characteristic	Отрицательная прогностическая ценность Negative predictive value	Положительная прогностическая ценность Positive predictive value	Комментарий Comment	Страна Country
X. Wang et al. (2019) [20]	Структурированные ЭМЗ Structured EMR	Машинное обучение Machine learning	XGBoost	Проспективное исследование Prospective study	873 598 ретроспективно, 836 659 проспективно retrospectively, 836 659 prospectively	Не указано Not specified	Не указано Not specified	0,881	Не указано Not specified	Не указано Not specified	Риск был выше у людей с хроническими заболеваниями и курящими The risk was higher in people with chronic diseases and smokers	Китай China
M.C. Yeh et al. (2021) [21]	Структурированные ЭМЗ Structured EMR	Глубокое обучение Deep learning	Convolutional Neural Networks	Проспективное исследование Prospective study	11 617 пациентов с РЛ, 1 423 154 контрольных 11 617 lung cancer patients, 1 423 154 controls	Не указано Not specified	Не указано Not specified	0,9 (0,87 для возраста 55 лет и старше) 0,9 (0,87 for ages 55 years and older)	Не указано Not specified	14,3 %	Высокая точность, лучшая для пациентов старше 55 лет High accuracy, best for patients over 55 years of age	Тайвань Taiwan
U. Chandran et al. (2023) [22]	Структурированные ЭМЗ Structured EMR	Машинное обучение Machine learning	LASSO regression	Ретроспективное исследование отобранных когорт Retrospective study of selected cohorts	4,7 млн человек, возраст 45–65 лет 4,7 million people, age 45–65 years	Не указано Not specified	Не указано Not specified	0,76	Не указано Not specified	Не указано Not specified	Модель продемонстрировала удовлетворительную точность The model showed satisfactory accuracy	США USA
M. Levi et al. (2024) [23]	Структурированные ЭМЗ Structured EMR	Машинное обучение Machine learning	Tree-based Pipeline Optimization Tool (Random Forest и XGBoost)	Ретроспективное исследование отобранных когорт Retrospective study of selected cohorts	4,076 обследуемых, возраст 35 лет и старше 4 076 subjects, age 35 years and older	69 %	Не указано Not specified	Не указано Not specified	Не указано Not specified	74 %	Модель показала высокую точность для курящих и некурящих The model showed high accuracy for smokers and non-smokers	Израиль Israel
S. Zhang et al. (2024) [24]	Структурированные ЭМЗ Structured EMR	Машинное обучение Machine learning	XGBoost, логистическая регрессия, наивный Байес, случайный лес XGBoost, logistic regression, naive bayes, random forest	Проспективное исследование Prospective study	467 888 — у 4299 из них развился РЛ 467,888 — 4299 of whom ultimately developed LC	Не указано Not specified	Не указано Not specified	0,998	Не указано Not specified	Не указано Not specified	Высокая точность, тест на нейротизм — один из важнейших факторов в модели. Выявлено большее влияние окружения белки, чем индекса массы тела High accuracy, Neuroticism Score is one of the most important factors in the model. A greater influence of hip circumference than body mass index was found	Китай China (Великобритания, UK)

неблагоприятных факторов ИИ выделил стаж курения и возраст пациентов. В то же время было выявлено снижение эффективности методики при анализе данных в азиатской и латиноамериканской популяциях, в связи с чем можно сделать вывод о точности ИИ в прогнозировании риска развития РЛ в общей популяции и наличии необходимости доработки алгоритма ИИ для обеспечения универсальности среди различных этнических групп [22].

M. Levi и соавт. в 2024 г. описали собственную модель ИИ, разработанную на основе машинного обучения с использованием TROT – Tree-based Pipeline Optimization Tool (Random Forest и XGBoost). ИИ оценивал факторы риска с помощью данных из электронных медицинских карт. Модель продемонстрировала точность 71,2 % и PPV 74 % для всей популяции. Результаты показали, что наиболее важными предикторами были наличие хронической обструктивной болезни легких (ХОБЛ), эмфиземы и хронического бронхита. Авторы пришли к выводу о том, что ИИ может служить полезным инструментом для идентификации пациентов с высоким риском развития РЛ (среди курящих и некурящих) и является перспективным направлением для дальнейшего изучения [23].

S. Zhang и соавт. в 2024 г. провели анализ данных UK Biobank (биобанка Соединенного Королевства Великобритании), большой биомедицинской базы данных, созданной в 2006 г., где были собраны образцы крови, мочи и слюны более 500 тыс. участников по всей стране, а также полная демографическая, социально-экономическая информация, информация об образе жизни и состоянии здоровья участников. Авторы использовали модели на основе XGBoost, логистической регрессии и naïve Bayes, random forest. Модель на основе XGBoost продемонстрировала наилучшие результаты (AUC = 0,998). Показана значимость таких факторов, как обхват бедра, Neuroticism score (стандартизированный психологический тест, включающий вопросы, касающиеся эмоциональных реакций, поведения в стрессовых ситуациях и общего психологического состояния), обхват талии, количество выкуриваемых сигарет и MET (метаболический эквивалент – metabolic equivalent of task) [24].

Таким образом, во всех исследованиях применение ИИ продемонстрировало значительное улучшение в прогнозировании риска РЛ по сравнению с традиционными методами оценки. Особенно это касается чувствительности и способности идентифицировать пациентов с высоким риском на ранних стадиях развития заболевания. Однако некоторые модели требуют доработки для обеспечения одинаково высокой точности среди всех популяций и этнических групп. Сведения о потенциальном вкладе данных технологий в снижение смертности от РЛ не представлены.

Основные показатели эффективности технологий искусственного интеллекта для ранней диагностики или скрининга рака легкого по данным структурированных электронных медицинских записей

Мы также проанализировали опубликованный состав атрибутов, включенных в модели ИИ (табл. 3), и опишем некоторые из них подробнее.

Возраст. Возраст был включен во все 5 анализируемых моделей. Этот фактор не только часто используется, но и в некоторых моделях рассматривается с дополнительными деталями, например, возрастными порогами (от 50 до 75 лет или старше 55 лет), которые значительно влияют на риск возникновения заболевания. В исследованиях отмечается, что риск РЛ увеличивается с возрастом, что делает этот фактор основным в моделях предсказания заболеваемости.

Курение. Стаж курения – ключевой фактор, включаемый в большинство моделей. В частности, учитываются следующие аспекты: статус курильщика (активный курильщик, бывший курильщик, никогда не куривший), длительность курения (количество лет), количество выкуриваемых сигарет и расчет пачко-лет (один из наиболее используемых показателей, определяющих кумулятивное воздействие курения). В ряде работ также оценивали влияние времени, прошедшего с момента отказа от курения, на риск развития РЛ. Эти параметры позволяют моделям более точно оценивать экологический риск для различных групп пациентов с учетом их анамнеза курения.

Хронические заболевания легких. Наличие хронических заболеваний легких, таких как ХОБЛ, хронический бронхит и эмфизема, включают в большинство моделей. Исследования показали, что наличие этих состояний существенно увеличивает риск развития РЛ независимо от статуса курения. У пациентов с такими заболеваниями модели ИИ часто показывают высокую точность предсказаний.

Индекс массы тела. Влияние данного индекса на риск развития РЛ неоднозначно, однако в нескольких моделях была выявлена прямая корреляция между его низким значением и повышенным риском развития РЛ, что может быть связано с кахексией и повышенной иммунологической толерантностью организма. Высокий индекс массы тела, напротив, ассоциировался со сниженным онкологическим риском. В исследовании S. Zhang и соавт. большее влияние продемонстрировали другие антропометрические факторы – обхват бедер и талии [24].

Пол и раса / этническая принадлежность. Эти факторы учитывают в нескольких моделях, подчеркивая их значимость в контексте предиктивной способности моделей. Например, некоторые исследования показали различия в рисках между мужчинами и

Таблица 3. Состав характеристик, включенных в модели искусственного интеллекта для прогнозирования риска развития рака легких по данным электронных медицинских записей, рутинных анализов крови и специализированных анкет

Table 3. Composition of attributes included in artificial intelligence models for predicting the risk of developing lung cancer based on electronic medical records, routine blood tests and specialized questionnaires

Автор/год/ссылка Author/year/reference	Оцениваемые показатели Estimated attributes
X. Wang et al. (2019) [20]	<p>Возраст, пол, курение, онкологический анамнез (любое злокачественное новообразование), заболевания легких в анамнезе (хроническая обструктивная болезнь легких – ХОБЛ, хронический бронхит, эмфизема и пневмония), другие хронические заболевания в анамнезе (сахарный диабет, хроническая болезнь почек (расчет скорости клубочковой фильтрации на основании уровня креатинина, веса, роста, возраста и пола), заболевания периферических сосудов, аритмия, аневризма брюшной аорты), симптомы (кровохарканье, кашель, боль в груди, одышка, периферические сосуды, аномальная потеря веса) и аномальные результаты лабораторных анализов (С-реактивный белок, фибриноген, глюкоза, гемоглобин, уровень лейкоцитов, тромбоцитов)</p> <p>Age, gender, smoking, oncological history (any malignant neoplasm), a history of lung diseases (chronic obstructive pulmonary disease – COPD, chronic bronchitis, emphysema and pneumonia), other chronic diseases in the anamnesis: coronary heart disease, peripheral vascular disease, arrhythmia, abdominal aortic aneurysm), symptoms (hemoptysis, cough, chest pain, shortness of breath, pleural effusion, abnormal weight loss) and abnormal laboratory test results (C-reactive protein, fibrinogen, glucose, hemoglobin, leukocyte, platelet levels)</p>
M.C. Yeh et al. (2021) [21]	<p>Возраст (от 20 до 90 лет, возраст старше 55 лет оказывает влияние), пол, история болезни обследуемых за 3 года (с/без предшествующими заболеваниями легких: асбестоз, бронхоэктазы, хронический бронхит, ХОБЛ, эмфизема, фиброз, пневмония, саркоидоз, силикоз, туберкулез и история приема лекарственных препаратов: противоядия, метронидазол, буфломедил, инсулин и др.)</p> <p>Age (from 20 to 90 years old, age over 55 years has an impact), gender, medical history of the subjects for 3 years (with/ without previous lung diseases: asbestos, bronchiectasis, chronic bronchitis, COPD, emphysema, fibrosis, pneumonia, sarcoidosis, silicosis, tuberculosis and history of taking medications: antidotes, metronidazole, buflomedil, insulin, etc.)</p>
U. Chandran et al. (2023) [22]	<p>Возраст, пол, раса (белая, черная, азиатская, неизвестная), этническая принадлежность (латиноамериканская, нелатиноамериканская, неизвестная), курение (нынешний курильщик, бывший курильщик, никогда не куривший) и установленный диагноз ХОБЛ за 365 дней до включения в исследование</p> <p>Age, gender, race (white, black, Asian, unknown), ethnicity (Hispanic, non-Hispanic, unknown), smoking (current smoker, former smoker, never smoked) and a diagnosis of COPD 365 days before inclusion in the study</p>
M. Levi et al. (2024) [23]	<p>Возраст, пол, индекс массы тела (риск больше с увеличением индекса массы тела на каждый 1 кг/м²), история курения, социально-экономический статус, наличие в анамнезе ХОБЛ, эмфиземы, хронического бронхита, интерстициального заболевания легких, фиброза легких и наличие в семейном анамнезе рака легкого.</p> <p>Age, gender, body mass index (the risk is higher with an increase in body mass index for every 1 kg/m²), smoking history, socio-economic status, a history of COPD, emphysema, chronic bronchitis, interstitial lung disease, pulmonary fibrosis and a family history of lung cancer</p>
S. Zhang et al. (2024) [24]	<p>Демографические характеристики (например, пол, возраст, образование), образ жизни и информация о здоровье (курение, бессонница, продолжительность сна, общего состояния здоровья, заболеваний легких, за исключением рака легкого), лабораторные и инструментальные данные (С-реактивный белок, объем форсированного выдоха за 1-ю секунду), профессиональные вредности (рабочее место, полное химические или других паров) и антропометрия (индекс массы тела, окружность талии, окружность бедер)</p> <p>Demographic characteristics (e.g., gender, age, education), lifestyle and health information (e.g., smoking, sleeplessness, overall health, lung diseases excluding lung cancer), laboratory data (e.g., c-reactive protein, forced expiratory volume in 1st second), work environment (e.g., workplace full of chemical or other fumes), and anthropometry (e.g., body mass index, waist circumference, hip circumference)</p>

женщинами, а также между представителями разных расовых и этнических групп. Это важно для разработки моделей, которые можно адаптировать для специфических популяций.

Социально-экономический статус и уровень образования. Социально-экономический статус и уровень образования влияют на доступ к медицинской помощи, а также на образ жизни, включая большую распространенность активного табакокурения. Включение этих показателей в модели позволяет учесть влияние факторов среды и поведения на риск развития РЛ.

Симптомы и сопутствующие хронические заболевания. Некоторые модели включают такие симптомы, как кашель, боль в груди и одышку, что помогает эффективно идентифицировать пациентов с высоким риском за счет своевременного выявления усиления клинической симптоматики. Сопутствующая патология (в частности, сердечно-сосудистые заболевания и сахарный диабет) также может быть включена в модели для повышения их прогностической точности.

Таким образом, факторы возраста, стажа курения и наличия хронических заболеваний легких являются основой для построения эффективных моделей ИИ, демонстрирующих высокую чувствительность и специфичность, а также способных стать полезным дополнением к традиционным методам скрининга. Однако в ряде случаев, например при учете расы и этнической принадлежности, модели требуют дополнительных адаптаций для увеличения точности в различных популяциях. Использование расширенных данных, таких как социально-экономический статус и подробный анамнез хронических заболеваний, также позволяет повысить эффективность прогнозирования развития РЛ.

Заключение

Несмотря на принимаемые в разных странах меры первичной профилактики, очевидно способствующие снижению заболеваемости и смертности, РЛ остается среди лидирующих причин смертности от ЗНО в большинстве стран Европы, Северной Америки и Азии, включая Российскую Федерацию. Вместе с тем, если судить по ранее проведенным исследованиям, скрининг РЛ в группах риска может снижать опухолеспецифическую и общую смертность в популяциях, которые были подвергнуты скринингу. Основным методом для раннего выявления злокачественных опухолей легкого остается низкодозная мультиспиральная КТ, и впечатляющее количество современных работ посвящено вопросам правильной интерпретации и классификации непосредственно изображений (компьютерных томограмм или рентгенограмм) легких для разделения пациентов на группы высокого и низкого риска заболеваемости РЛ. Некоторые работы при этом также учитывают данные других модальностей (например, социально-демографические данные, стаж курения и т.д.), что повышает точность классификации изображений.

Безусловно, разрабатываемые модели на основе ИИ обладают рядом важных преимуществ, способствуют улучшению показателей предиктивной точности программ скрининга и обеспечивают дополнительный анализ данных. Подобный подход позволяет выявлять наличие закономерности даже при оценке факторов, динамически изменяющихся с течением времени или не обладающих, на первый взгляд, четкой патогенетической связью с развитием заболевания. В то же время прямое сравнение различных моделей ИИ сегодня отсутствует, что подчеркивает необходимость проведения дальнейших проспективных исследований и актуализирует направление нашей научной деятельности.

ЛИТЕРАТУРА / REFERENCES

1. Ferlay J., Ervik M., Lam F. et al. Global Cancer Observatory: Cancer Today. Lyon: International Agency for Research on Cancer, 2024. URL: <https://gco.iarc.who.int/today>.
2. Состояние онкологической помощи населению России в 2023 году. Под ред. А.Д. Каприна, В.В. Старинского, А.О. Шахзадовой. М.: МНИОИ им. П.А. Герцена – филиал ФГБУ «НМИЦ радиологии» Минздрава России, 2024. 262 с. State of oncological care for the population of Russia in 2023. Kaprin A.D., Starinskiy V.V., Shakhzadova A.O., eds. Moscow: MNIOI im. P.A. Gertsena – filial FGBU “NMITS radiologii” Minzdrava Rossii, 2024. 262 p. (In Russ.).
3. Злокачественные новообразования в России в 2023 году (заболеваемость и смертность). Под ред. А.Д. Каприна и др. М.: МНИОИ им. П.А. Герцена – филиал ФГБУ «НМИЦ радиологии» Минздрава России, 2024. 276 с. Malignant neoplasms in Russia in 2023 (incidence and mortality). Kaprin A.D. et al., eds. Moscow: MNIOI im. P.A. Gertsena – filial FGBU “NMITS radiologii” Minzdrava Rossii, 2024. 276 p. (In Russ.).
4. Toumazis I., Bastani M., Han S.S. et al. Risk-based lung cancer screening: a systematic review. *Lung Cancer* 2020;147:154–86. DOI: 10.1016/j.lungcan.2020.07.007
5. Koski E., Murphy J. AI in healthcare. *Stud Health Technol Inform* 2021;284:295–9. DOI: 10.3233/SHTI210619
6. Thong L.T., Chou H.S., Chew H.S.J. et al. Diagnostic test accuracy of artificial intelligence-based imaging for lung cancer screening: a systematic review and meta-analysis. *Lung Cancer* 2023;176:4–13. DOI: 10.1016/j.lungcan.2022.12.002
7. Hsu Y.C., Tsai Y.H., Weng H.H. et al. Artificial neural networks improve LDCT lung cancer screening: a comparative validation study. *BMC Cancer* 2020;20(1):1023. DOI: 10.1186/s12885-020-07465-1
8. Chauvie S., De Maggi A., Baralis I. et al. Artificial intelligence and radiomics enhance the positive predictive value of digital chest tomosynthesis for lung cancer detection within SOS clinical

- trial. *Eur Radiol* 2020;30(7):4134–40. DOI: 10.1007/s00330-020-06783-z
9. Bach P.B., Kattan M.W., Thornquist M.D. et al. Variations in lung cancer risk among smokers. *J Natl Cancer Inst* 2003;95(6):470–8. DOI: 10.1093/jnci/95.6.470
 10. Tammemägi M.C., Katki H.A., Hocking W.G. et al. Selection criteria for lung-cancer screening. *N Engl J Med* 2013;368(8):728–36. DOI:10.1056/NEJMoa1211776
 11. Katki H.A., Kovalchik S.A., Berg C.D. et al. Development and validation of risk models to select ever-smokers for CT lung cancer screening. *JAMA* 2016;315(21):2300–11. DOI:10.1001/jama.2016.6255
 12. Cheung L.C., Berg C.D., Castle P.E. et al. Life-gained-based versus risk-based selection of smokers for lung cancer screening. *Ann Intern Med* 2019;171(9):623–32. DOI:10.7326/M19-1263
 13. Markaki M., Tsamardinos I., Langhammer A. et al. A validated clinical risk prediction model for lung cancer in smokers of all ages and exposure types: a HUNT study. *EBioMedicine* 2018;31:36–46. DOI: 10.1016/j.ebiom.2022.104187
 14. Field J.K., Vulkan D., Davies M.P.A. et al. Liverpool Lung Project lung cancer risk stratification model: calibration and prospective validation. *Thorax* 2021;76(2):161–8. DOI: 10.1136/thoraxjnl-2020-215158
 15. PLCO Lung dataset(s). In: Book PLCO Lung dataset(s). Ed. NIH NCI, 2024. URL: <https://cdas.cancer.gov/datasets/plco/21/>
 16. NLST dataset(s). In: Book NLST dataset(s). Ed. NIH NCI, 2024. URL: <https://cdas.cancer.gov/datasets/nlst/>
 17. Tammemägi M.C., Schmidt H., Martel S. et al. Participant selection for lung cancer screening by risk modelling (the Pan-Canadian Early Detection of Lung Cancer [PanCan] study): a single-arm, prospective study. *Lancet Oncol* 2017;18(11):1523–31. DOI: 10.1016/S1470-2045(17)30597-1
 18. NHIS Data, Questionnaires and Related Documentation. In: Book NHIS Data, Questionnaires and Related Documentation. Ed. CDC, 2024.
 19. Marshall R.C., Tiglaos S.M., Thiel D. Updated USPSTF screening guidelines may reduce lung cancer deaths. *J Fam Pract* 2021;70(7):347–9. DOI: 10.12788/jfp.0257
 20. Wang X., Zhang Y., Hao S. et al. Prediction of the 1-year risk of incident lung cancer: prospective study using electronic health records from the State of Maine. *J Med Internet Res* 2019;21(5):e13260. DOI: 10.2196/23586
 21. Yeh M.C., Wang Y.H., Yang H.C. et al. Artificial intelligence-based prediction of lung cancer risk using nonimaging electronic medical records: deep learning approach. *J Med Internet Res* 2021;23(8):e26256. DOI: 10.2196/33519
 22. Chandran U., Reys J., Yang R. et al. Machine learning and real-world data to predict lung cancer risk in routine care. *Cancer Epidemiol Biomarkers Prev* 2023;32(3):337–43. DOI: 10.1158/1055-9965.EPI-22-0873
 23. Levi M., Lazebnik T., Kushnir S. et al. Machine learning computational model to predict lung cancer using electronic medical records. *Cancer Epidemiol* 2024;92:102631. DOI: 10.1016/j.canep.2024.102649
 24. Zhang S., Yang L., Xu W. et al. Predicting the risk of lung cancer using machine learning: A large study based on UK Biobank. *Medicine (Baltimore)* 2024;103(16):e37879. DOI: 10.1097/MD.00000000000037879

Вклад авторов

И.В. Самойленко: научное руководство, обоснование концепции исследования, создание модели исследования;
 В.В. Назарова: аккумуляция исследовательских данных, написание текста рукописи;
 З.Р. Магомедова: обзор публикаций по теме статьи, анализ и систематизация полученных данных, написание текста рукописи;
 П.В. Кононец: редактирование рукописи, формулировка выводов;
 И.М. Боровков: редактирование рукописи, оформление рукописи;
 Т.Г. Геворкян: разработка дизайна исследования, редактирование рукописи.

Authors' contribution

I.V. Samoylenko: scientific guidance, validation of the research concept, creation of a research model;
 V.V. Nazarova: accumulation of research data, writing of the manuscript text;
 Z.R. Magomedova: review of publications on the topic of the article, analysis and systematization of the data obtained, writing the text of the manuscript;
 P.V. Kononets: manuscript editing, wording of conclusions;
 I.M. Borovkov: manuscript editing, designing the manuscript;
 T.G. Gevorkyan: research design development, manuscript editing.

ORCID авторов / ORCID of authors

И.В. Самойленко / I.V. Samoylenko: <https://orcid.org/0000-0001-7150-5071>
 В.В. Назарова / V.V. Nazarova: <https://orcid.org/0000-0003-0532-6061>
 З.Р. Магомедова / Z.R. Magomedova: <https://orcid.org/0009-0006-9145-0905>
 П.В. Кононец / P.V. Kononets: <https://orcid.org/0000-0003-4744-6141>
 И.М. Боровков / I.M. Borovkov: <https://orcid.org/0000-0002-2017-8047>
 Т.Г. Геворкян / T.G. Gevorkyan: <https://orcid.org/0009-0008-3486-302X>

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Conflict of interests. Authors declare no conflict of interests.

Финансирование. Исследование выполнено при поддержке НИР 124052700085-6 «Разработка предиктивных моделей на основе искусственного интеллекта для раннего выявления онкологических заболеваний по мультимодальным медицинским и социально-демографическим данным с формированием предложений по последующей оптимизации национальных программ скрининга» (АНО «Аналитический центр при Правительстве Российской Федерации»).

Funding. The work was supported by research fund 124052700085-6 “Development of predictive models based on artificial intelligence for early detection of oncological diseases based on multimodal medical and socio-demographic data with the formation of proposals for subsequent optimization of national screening programs” (Autonomous Non-commercial Organization “Analytical Center under the Government of the Russian Federation”).

Статья поступила: 18.11.2024. **Принята в печать:** 25.02.2025. **Опубликована онлайн:** 00.00.2025.

Article received: 18.11.2024. **Accepted for publication:** 25.02.2025. **Published online:** 00.00.2025.