

ПЕРСПЕКТИВЫ ПРИМЕНЕНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ СКРИНИНГА ЗЛОКАЧЕСТВЕННЫХ НОВООБРАЗОВАНИЙ

Д.Г. ЗАРИДЗЕ¹

¹ ФГБУ «НМИЦ онкологии им. Н.Н. Блохина» Министерства здравоохранения Российской Федерации, г. Москва, Россия.

УДК 614.2

DOI: 10.21045/2782-1676-2024-4-4-24-42

Аннотация

Введение. Эффективность скрининга как одной из наиболее действенных стратегий контроля злокачественных опухолей не вызывает сомнений. Скрининг снижает риск диагностики рака на поздней стадии и выявляет предраковые патологии, чем предотвращает и его развитие. Потенциальные ограничения и опасность скрининга заключаются в высокой вероятности ложноположительных, ложноотрицательных результатов и гипердиагностики. Последствия – дополнительные обследования и ненужное и, часто, чрезмерное лечение. В то же время, при скрининге часто не попадают в поле зрения интервальные раки, которые характеризуются агрессивным течением. **Цель исследования:** изучить эффективность искусственного интеллекта (ИИ) для повышения чувствительности и специфичности скрининга злокачественных новообразований (ЗНО) и снижения частоты ложноотрицательных, ложноположительных результатов и гипердиагностики. **Материалы и методы.** Обзор и анализ опубликованных научных данных, посвященных: а) скринингу рака молочной железы (РМЖ), рака легкого (РЛ), рака предстательной железы (РПЖ), рака шейки матки (РШМ) и рака толстой кишки (РТК); б) разработке и применению ИИ для улучшения эффективности скрининговых программ. Поиск соответствующих публикаций был произведен в базах данных PubMed и Cochrane Library. **Результаты.** При маммографическом скрининге ИИ снижает количество неправильной интерпретации маммограм, количество повторных вызовов, количество биопсий с отрицательным результатом, повышает эффективность интерпретации маммограм независимо от характеристик органа (плотная молочная железа, кальцификаты). Применение ИИ совместно с низкодозовой компьютерной томографией (НДКТ) для скрининга РЛ не только улучшает диагностику его различных форм, но и предсказывает риск развития рака на несколько лет вперед. Систематический обзор и мета-анализ 12 работ по оценке эффективности ИИ в тандеме с мультипараметрической магнитнорезонансной томографией (мпМРТ) предстательной железы показал высокую суммарную эффективность в диагностике клинически значимого РПЖ, что способствовало статистически достоверному снижению количества дополнительных приглашений и ненужных биопсий. Вопрос эффективности ИИ в сочетании с колоноскопией, несмотря на применение его самых продвинутых систем (система глубокого обучения, основанная на сверточной нейронной сети), остается спорным. Решение этой проблемы зависит от того, какую цель мы преследуем, разрабатывая и обучая систему: повышение «выявляемости» аденом и их удаление независимо от их размеров или идентификацию и удаление только больших аденом, из которых, с высокой вероятностью, может развиваться рак. Успешное применение ИИ для цитологической диагностики патологии шейки матки, включая все стадии цервикальных интраэпителиальных неоплазий (ЦИН), вызывает оптимизм. Внедрение систем ИИ, обученных взаимодействию с цитопатологом в прочтении и оценке цитологического материала и диагностике ЦИН и РШМ, снизит нагрузку на цитологов и на другой медицинский персонал. **Заключение.** Представленные исследования указывают на перспективность применения ИИ для диагностики ЗНО, особенно в контексте популяционного скрининга, при котором исследование проходят многие тысячи человек. Применение ИИ достоверно повышает эффективность диагностических методов, улучшает показатели чувствительности и специфичности, снижает вероятность ложноотрицательных, ложноположительных результатов и гипердиагностики. Эффективность ИИ для прогнозирования риска развития рака на несколько лет вперед может способствовать удлинению интервалов между раундами скрининга и, соответственно, снижению нагрузки на систему здравоохранения и сокращению затрат. Решение о внедрении в программу популяционного скрининга любой из систем ИИ, с доказанной эффективностью в рамках клинических исследований, должно быть принято только после ее апробирования на популяционном уровне. Необходимо разработать формы «информированного согласия» для пациентов, в которых подробно и объективно описаны все преимущества и недостатки применения ИИ по сравнению с существующей принятой практикой.

Ключевые слова: скрининг, искусственный интеллект, ИИ, рак молочной железы, рак предстательной железы, рак легкого, рак толстой кишки, рак шейки матки.

Для цитирования: Заридзе Д.Г. Перспективы применения искусственного интеллекта для повышения эффективности скрининга злокачественных новообразований. Общественное здоровье. 2024; 4(4):24–42, DOI: 10.21045/2782-1676-2024-4-4-24-42

Контактная информация: Заридзе Давид Георгиевич; e-mail: dgzaridze@rcs-pror.org

Финансирование: исследование не имело спонсорской поддержки.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Статья поступила в редакцию: 14.08.2024. **Статья принята к печати:** 25.10.2024. **Дата публикации:** 10.12.2024.

UDC 614.2

DOI: 10.21045/2782-1676-2024-4-4-24-42

PROSPECTS OF USING ARTIFICIAL INTELLIGENCE FOR IMPROVING CANCER SCREENING EFFICACY

D.G. Zaridze¹¹ National Medical Research Center of Oncology named after N.N. Blokhin, Moscow, Russia.**Abstract**

Introduction. The effectiveness of screening as one of the strategies for cancer control is beyond doubt. Screening reduces the risk of diagnosing cancer at a late stage and identifies precancerous pathologies, thereby preventing the development of cancer. Potential limitations of screening include the high probability of false positives, false negatives, and overdiagnosis. The consequences are additional examinations and unnecessary and, often, excessive treatment. At the same time, interval cancers, which are characterized by an aggressive course, often do not come into view. *The purpose of the study:* to explore the data on effectiveness of artificial intelligence (AI) for improving the sensitivity and specificity of cancer screening and reducing the probability of false negative and false positive results, and overdiagnosis. *Materials and methods.* Review and analysis of published data on a) screening of breast cancer (BC), lung cancer (LC), prostate cancer (PC), cervical cancer (CC) and large bowel cancer (LBC); b) development and application of AI systems to improve the effectiveness of screening. The PubMed and Cochrane Library databases were searched for relevant publications. *Results.* In mammography screening, AI reduces the number of abnormal interpretations of mammograms, the number of recalls, the number of biopsies with a negative result, and increases the efficacy of mammogram interpretation regardless of the characteristics of the breast (dense breast, calcifications). The use of AI in conjunction with low-dose computed tomography (LDCT) for LC screening not only improves the diagnosis of various types of LC, but also predicts the risk of developing cancer several years in advance. A systematic review and meta-analysis of 12 studies evaluating the effectiveness of AI in tandem with multiparametric magnetic resonance imaging (mpMRI) of the prostate showed high overall effectiveness in the diagnosis of clinically significant PC. The performance of the AI system – based on the multimodal data including demographics, clinical characteristics, laboratory tests and ultrasound reports of patients with PC, was better than the effectiveness of PSA tests in diagnosing clinically significant PC. The effectiveness of AI in tandem with colonoscopy, despite the use of the most advanced AI systems (deep learning system based on a convolutional neural network), remains controversial. The solution to this problem depends on what goal we are pursuing when developing and training the system? Increasing “detection rate” of adenomas, regardless of their size, and removing them, or identifying and removing only large adenomas? The successful use of AI for cytological diagnosis of cervical pathology, including all stages of cervical intraepithelial neoplasia (CIN), is encouraging. The introduction of AI systems developed and trained to interact with a cytopathologist in reading and evaluating cytological material and diagnosing CIN and CC into general practice will reduce the burden on cytopathologists and other medical personnel. *Conclusion.* The analysis of published data has shown the promising results concerning the use of AI for cancer diagnostics, especially in the setting of population screening programs, which cover many thousands of people. The use of AI significantly increases the effectiveness of diagnostic tool, improves its sensitivity and specificity, and reduces the probability of false negative, false positive results and overdiagnosis. The decision to introduce into practice any of the AIs with proven effectiveness in clinical trials should be made only after its testing in a real world, at the population level. The “informed consent” forms that objectively describe all the advantages and disadvantages of the use of AI compared to current practice has to be developed.

Keywords: screening, artificial intelligence, AI, cancer of the breast, lung, prostate, large bowel, cervix.

For citation: Zaridze D.G. Prospects of using artificial intelligence for improving cancer screening efficacy. Public health. 2024; 4(4):24–42, DOI: 10.21045/2782-1676-2024-4-4-24-42

For correspondence: David G. Zaridze, e-mail: dgzaridze@rcs-pror.org

Funding: the study had no sponsorship.

Conflict of interests: the authors declare that there is no conflict of interests.

Аннотации на испанском и французском языках приводятся в конце статьи

ВВЕДЕНИЕ

Скрининг – важнейший компонент организации онкологической помощи. Методам скрининга и оценке их эффективности посвящены сотни научных статей, монографий и руководств, подготовленных и опубликованных ВОЗ, Международным агентством по изучению рака, Европейским Союзом [1–4].

Представленная статья не претендует на всеобъемлющий охват проблемы скрининга. Тем не менее, в ней рассматриваются наиболее

чувствительные этапы организации и оценки эффективности скрининговых программ, которые часто игнорируются. Автор объясняет это недостаточной информированностью организаторов скрининга, ввиду очевидного дефицита в России научно обоснованной информации по этой проблеме. Статей, посвященных скринингу, опубликованных в России, единицы [5–7], и, соответственно, автор считает необходимым восполнить этот пробел и определить возможные пути повышения его эффективности с помощью искусственного интеллекта (ИИ).

Цель скрининга – раннее активное выявление бессимптомного рака и его лечение. Скрининг не всегда эффективен. Это противоречит укоренившемуся среди врачей мнению, что любой скрининг должен дать положительный результат [5, 6]. Эффективность скрининговых программ определяется: а) степенью снижения смертности от той формы рака, для преклинического выявления которой он проводится; б) приемлемым соотношением его пользы и вреда.

Эффективность скрининга выше для тех форм опухолей, естественное развитие которых позволяет выявлять и лечить предраковые образования (например, аденоматозные полипы толстой кишки и цервикальные интраэпителиальные неоплазии – ЦИН), что приводит к снижению как смертности, так и заболеваемости.

Для оценки эффективности скрининга необходимо проведение рандомизированных клинических исследований (РКИ). Проведение РКИ требует определения целевой популяции с достаточным риском развития заболевания, т.е. с высокой заболеваемостью; тщательной рандомизации для создания группы, которая будет приглашаться на скрининг и группы, которая будет получать стандартное медицинское обслуживание; учета прохождения скрининга участниками контрольной группы вне рамок этого РКИ; достаточной длительности периода наблюдения и сбора данных об исходах для всех рандомизированных участников.

Эффективность скрининга рака молочной железы (РМЖ), рака толстой кишки (РТК) и рака легкого (РЛ) подтверждена в РКИ до начала широкого их применения в популяции. Эффективность цитологического (тест по Папаниколау) скрининга рака шейки матки (РШМ) была подтверждена ретроспективно на основании сравнения смертности от рака этого органа в Финляндии, где проводился массовый цитологический скрининг среди всего женского населения, и Норвегии, где организованной программы скрининга РШМ не было [5, 6]. Простатический специфический антиген (ПСА), являющийся онкомаркером рака предстательной железы (РПЖ), получил широкое распространение в США до подтверждения его эффективности, подразумевая снижение смертности от РПЖ. В результате, начался стремительный рост заболеваемости при сохранении стабильных значений смертности от РПЖ. Отказ от ПСА-скрининга, как и следовало ожидать, привел к снижению заболеваемости, которая быстро достигла исходных (до скрининговых) показателей. РКИ

по ПСА-скринингу РПЖ были проведены уже в XXI столетии.

Эффективность скрининга как одной из наиболее действенных стратегий контроля рака не вызывает сомнения, однако сам по себе он не является совершенной процедурой и связан не только с преимуществами, но и с ограничениями и негативными эффектами, что необходимо учитывать при разработке клинических рекомендаций и при обсуждении участия в программе скрининга с потенциальными её участниками – пациентами. Преимущества включают снижение риска выявления рака на поздней стадии и выявление предраковых патологий и их лечение, т.е. профилактику и, соответственно, снижение заболеваемости. Потенциальные ограничения и опасность заключаются в достаточно высокой вероятности ложноположительных и ложноотрицательных результатов, а также гипердиагностики.

Ложноположительные результаты приводят к дополнительным диагностическим тестам и дополнительной нагрузке на медицинский персонал, отрицательным эмоциям, волнению и стрессу пациентов, гипердиагностика¹ – к необоснованным дополнительным диагностическим процедурам и ненужному лечению. Скрининг может обнаружить медленно прогрессирующую опухоль у пациента, который, скорее всего, умрет от других причин до того, как она клинически себя проявит. В каждом из этих случаев лечение является бесполезным и может рассматриваться как наносящее существенный вред.

Скрининг и сопровождающая его гипердиагностика – основной драйвер астрономического роста заболеваемости некоторыми формами рака. Рост заболеваемости при стабильной или снижающейся смертности характерен для РПЖ и РМЖ, а также других форм рака, не включенных в программы скрининга: рака почки (РП), рака щитовидной железы (РЩЖ) и меланомы. Этот феномен был описан в 1984 г. Zaridze и соавт. [8] на примере США: «Рост заболеваемости раком простаты, последовавший за широким распространением ПСА-скрининга, можно объяснить выявлением латентного клинически незначимого рака, который в отсутствие скрининга клинически не проявляется, не дает симптомов и никогда не прогрессирует. На продолжительность и качество жизни мужчин этот рак

¹ Гипердиагностика – это выявление с помощью скрининга опухоли, которая, в противном случае, осталась бы незамеченной и никак не проявила бы себя на протяжении всей жизни пациента.

не влияет». Welch и Black [9] продемонстрировали рост заболеваемости РМЖ, РПЖ, РП, РЩЖ и меланомы в США, при стабильной или снижающейся смертности (рис. 1). В России подобный тренд, характерный для гипердиагностики, описан Д. Г. Заридзе и соавт. в 2020 г. (рис. 2, 3) [10].

Рост количества пациентов с гипердиагностированными формами рака сопровождается растущей нагрузкой на систему здравоохранения, онкологическую помощь, как на ее профессиональный, так и финансовый потенциал.

Многие тысячи пациентов проходят скрининг, как популяционный, так и оппортунистический, при котором в 20–30% случаев имеет место гипердиагностика. Последствия – ненужные дополнительные обследования и ненужное и, часто, чрезмерное лечение. В то же время, при скрининге и профилактических осмотрах часто не попадают в поле зрения агрессивные формы, так называемые интервальные раки, что оказывает отрицательное влияние на прогноз, показатели выживаемости и смертность.

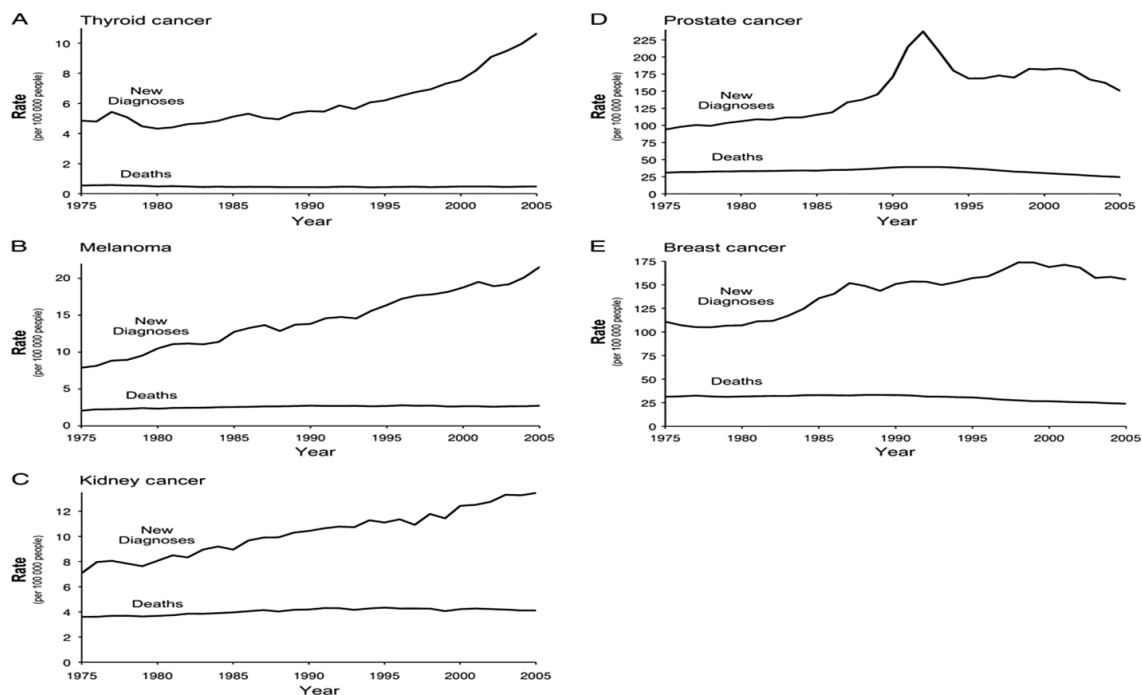


Рис. 1. Динамика заболеваемости (диагностирования) и смертности в США от: А) рака щитовидной железы, В) меланомы, С) рака почки, D) рака простаты, Е) рака молочной железы [9]

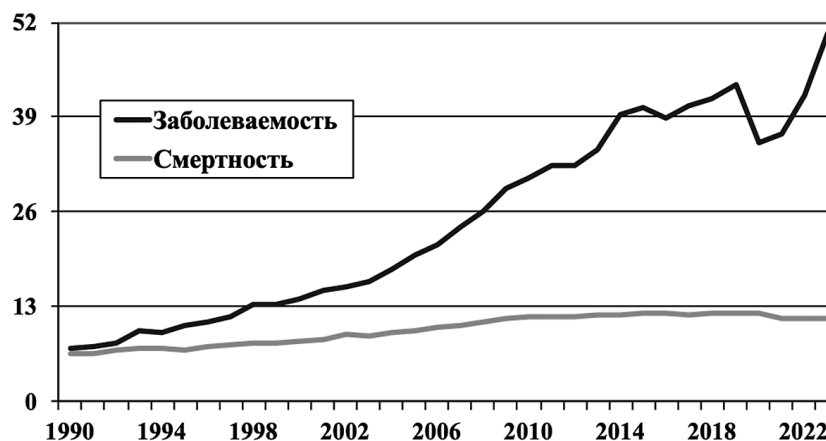


Рис. 2. Динамика заболеваемости и смертности от РПЖ в России
Источник: собственные данные.

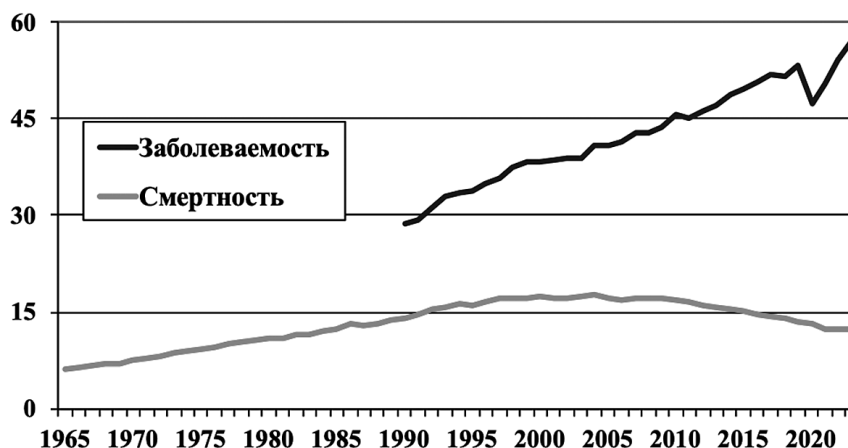


Рис. 3. Динамика заболеваемости и смертности от РМЖ в России

Источник: собственные данные.

Все вышесказанное указывает на необходимость повышения эффективности скрининга. Накоплен опыт успешного использования ИИ в повышении чувствительности и специфичности, применяемых при скрининге методов визуализации, таких, как маммография, низкодозовая компьютерная томография (НДКТ) и магниторезонансная томография (МРТ) [11]. В литературе также представлены исследования по успешному применению ИИ алгоритма при скрининге и диагностике РТК и РШМ.

Цель исследования:

1. Обзор и анализ опубликованных научных данных эффективности скрининга РМЖ, РЛ, РПЖ, РТК, РШМ в РКИ на практике, в популяции.
2. Анализ и оценка опубликованных данных, посвященных применению ИИ при скрининге злокачественных новообразований (ЗНО) для: а) повышения чувствительности и специфичности скрининга; б) снижения частоты ложноотрицательных результатов, т.е. вероятности пропуска клинически значимого образования и в) снижения частоты ложноположительных результатов и гипердиагностики.

МАТЕРИАЛЫ И МЕТОДЫ

Обзор и анализ опубликованных научных данных, посвященных: а) скринингу РМЖ, РЛ, РПЖ, РШМ и РТК; б) разработке и применению ИИ для улучшения эффективности скрининговых программ. Статьи, посвященные перечисленным проблемам, отбирались в международной базе данных Pubmed и Cochrane Library.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Скрининг РМЖ. Эффективность маммографического скрининга доказана в РКИ с длительностью наблюдения от 10 до 18 лет [12–15]. В общей сложности во всех программах участвовали более 600 тыс. женщин, 50% из которых ежегодно или 1 раз в 2 года проводилась маммография в двух проекциях, а в некоторых исследованиях и пальпация молочной железы. Мета-анализ 7 европейских РКИ, включавших данные о 500 тыс. женщин, которые получили приглашение принять участие в маммографическом скрининге, показал снижение смертности в группе скрининга, по сравнению с контрольной группой, на 25%. Среди женщин, которые фактически приняли участие в скрининге, смертность снизилась на 30–35% [13]. Во всех цитируемых работах для скрининга использовались маммографы с аналоговым изображением. Однако последующие исследования с использованием цифровых маммографов подтвердили предыдущие результаты [16].

Маммографический скрининг – наиболее распространённая форма скрининга в мире. Предложения по использованию других методов визуализации (УЗИ, томосинтез, МРТ) для скрининга РМЖ ничем не обоснованы, так как их преимущество, по сравнению с маммографией, не доказано [17]. В то же время, возможности маммографии, как метода скрининга, ограничены. Маммография пропускает от 10 до 30% случаев РМЖ. Чаще всего, ложноотрицательные результаты имеют место при прочтении маммограм у женщин с плотной молочной железой. При этом не исключены ошибки, связанные

с человеческим фактором, такие, как недостаточная квалификация рентгенолога и большая загруженность. Другая проблема – чрезмерное количество повторных приглашений для дополнительного обследования. Например, в США 43% рентгенологов превышает рекомендованное количество повторных приглашений, а диагноз рака подтверждается только у 28,6% пациентов с подозрением на рак, которым произведена биопсия. Вызывает озабоченность значительная вариабельность между рентгенологами в оценке маммограм [18–23]. Внедрение CAD (computer-aided detection) не только не способствовало улучшению интерпретации маммограм, но привело к росту ложноположительных результатов [24, 25].

В популяциях, в которых проводится массовый маммографический скрининг, растет заболеваемость неинвазивной формой внутрипротокового рака *in situ*, и за счет этого – общая заболеваемость РМЖ. До распространения маммографического скрининга, в начале 1980-х гг., внутрипротоковый рак *in situ* встречался крайне редко и составлял не более 2% всех случаев РМЖ. В настоящее время около 20% случаев РМЖ диагностируются на преинвазивной стадии [26].

Далее рассмотрим исследования, целью которых являлась разработка алгоритма искусственного интеллекта (ИИ) для повышения эффективности маммографической диагностики, снижения количества ложноположительных результатов и связанных с этим повторных приглашений. Обе работы основаны на хорошо охарактеризованных и прослеженных когортах пациентов-участниц диагностических и скрининговых программ. В южнокорейском исследовании [27] программа ИИ разработана на базе глубокой нейронной сети обработки изображения с использованием 170 230 маммографических исследований, проведенных в Южной Корее, США и Великобритании. Эффективность ИИ оценивалась на основании показателя кривой AUROC (Area Under ROC, AUC) – площадь под кривой ROC². Диагностическая эффективность при прочтении изображения только ИИ (0,940, 95% ДИ 0,91–0,96) была достоверно выше по сравнению с прочтением только рентгенологом (0,810, 95% ДИ 0,77–0,85) и рентгенологом + ИИ (0,881, 95% ДИ 0,850–0,911) ($p < 0,0001$). Чувствительность и специфичность оценки маммограм также достоверно выше при их прочтении толь-

ко ИИ по сравнению с прочтением рентгенологами. Преимущество ИИ сохраняется во всех подгруппах независимо от возраста, характеристик ткани молочной железы, наличия кальцификаций и гистологического типа. Например, при плотной молочной железе чувствительность ИИ на 16% выше, чем рентгенологов ($p < 0,0001$), а разница в специфичности составляет 14,7% ($p < 0,0001$).

В исследовании, проведенном в Швеции [28] на базе когорты 55 581 женщин, прошедших маммографический скрининг, использовалась ИИ система, разработанная в Южной Корее [27]. Исследование показало, что ИИ не только не уступает результатам стандартного прочтения маммограм двумя радиологами, но имеет ряд преимуществ. Количество неправильной интерпретации маммограм было на 49% ниже (2 916 vs 3 854) при прочтении маммограм только ИИ, чем при применении стандартного метода; повторных приглашений – на 47% (768 vs 1 629) ниже при прочтении маммограм только ИИ, по сравнению со стандартным методом. Соответственно, ниже число женщин, проходящих дополнительное обследование, включая томосинтез и УЗИ, не говоря уже о стрессе, который испытывают женщины в промежутке между первым и дополнительным обследованием.

Представленные исследования указывают на перспективность применения ИИ для маммографического скрининга, при котором исследование проходят многие тысячи женщин. В популяционных программах все ограничения, присущие маммографии, усугубляются и, соответственно, растет вероятность вреда, наносимого скринингом. Это, в первую очередь, обусловлено «человеческим фактором», а именно, усталостью рентгенолога и снижением качества его работы в результате его непомерной нагрузки. ИИ может быть инкорпорирован в диагностический процесс для второго прочтения маммограм или полностью заменить рентгенолога, при соответствующем контроле качества его диагностики.

Скрининг РЛ. Анализ крупных и наиболее часто цитируемых РКИ (Бельгийско-голландское – NELSON, Немецкое – LUSI, итальянское – MILD, США – NLST) скрининга РЛ с использованием НДКТ показал, что скрининг достоверно снижает смертность от РЛ в опытной группе, по сравнению с контрольной. Смертность от РЛ статистически достоверно снизилась в NLST на 20% и NELSON на 33% среди женщин и на 24% среди мужчин. Статистически достоверное снижение

² ROC – позволяет суммировать производительность модели одним числом, измеряя площадь под кривой ROC.

смертности от всех причин в группе скрининга было продемонстрировано только в NLST. Сроки прослеживания участников скрининга варьировали от 8,8 до 12 лет. Во всех этих исследованиях для участия в программах скрининга отбирались мужчины и женщины, в возрасте 55–74 лет, которые курили на момент их включения в программу или отказались от курения 10–15 лет назад, с показателем интенсивности курения более 20 пачек-лет (т.е. 1 пачка/день в течение 20 лет или 2 пачки/день в течение 10 лет) [29, 30].

Мета-анализ 7 опубликованных РКИ, который включил, в общей сложности, 84 558 участников, выявил статистически достоверное снижение смертности от РЛ на 17% (ОР = 0,83, 95% ДИ 0,76–0,91) и недостоверное снижение смертности от всех причин на 4% (ОР = 0,96, 95% ДИ 0,92–1,00) [31].

В Кокрейновский обзор и мета-анализ, цель которых состояла в оценке влияния НДКТ на смертность от РЛ и от всех причин, включены 8 исследований, проведенных в США и Европе, с общим числом участников, равным 91 122, в возрасте от 40 и более лет и с индексом истории курения ≥ 20 пачек-лет. В группе НДКТ скрининга отмечено снижение смертности от РЛ на 21% по сравнению с контрольной группой, в которой состояли пациенты, не прошедшие НДКТ (ОР 0,79, 95% ДИ 0,72–0,87; умеренная достоверность по Кохрановской классификации). В группе НДКТ снизилась и смертность от всех причин на 5% (ОР 0,95%, 95% ДИ 0,91–0,99; умеренная достоверность). Ложноположительные результаты и частота повторных вызовов чаще зарегистрированы в группе НДКТ скрининга, а расчетный показатель гипердиагностики, связанный с НДКТ скринингом, составил 18% (ОР 18%, 95% ДИ 0,0–0,36; низкая достоверность) [32].

Скрининг РЛ сталкивается с теми же проблемами, что и другие формы скрининга: с высокой частотой ложноположительных и ложноотрицательных результатов и гипердиагностикой. Прослеживания участников NLST (США) и анализ исходов (конечных точек) показал, что доля гипердиагностики среди случаев РЛ, выявленных НДКТ, составляет 18,5% (95% ДИ, 5,4–30,6) для РЛ в целом, 22,5% (95% ДИ, 9,7–34,3) для НМРЛ и 78,9% (95% ДИ, 62,2–93,5) для бронхоальвеолярного рака [15]. Предложенная в РКИ NELSON практика измерения объема, а не размера узла, и использования этого измерения для стратификации между образованием, которое нуждается в дальнейшей оценке (биопсия)

и хирургическом вмешательстве, или образованием, за которым нужно наблюдать, значительно снизило количество ложноположительных диагнозов и гипердиагностики [30]. Однако проблема остается, и для ее решения разрабатываются алгоритмы ИИ, которые будут представлены ниже.

Улучшение показателей чувствительности и специфичности – важная задача скрининговых программ в целом и РЛ, в частности. Ложноотрицательные и ложноположительные результаты сопряжены с необходимостью повторных приглашений, дополнительного обследования и биопсий, что, в свою очередь, повышает нагрузку на медицинский персонал и систему здравоохранения. Компьютерные программы CAD и рекомендации (LUNG-RADS) для интерпретации обнаруженных узлов с точки зрения злокачественности и определение клинической тактики на основании их размеров и плотности несколько улучшили эти показатели. Однако сохраняется необходимость разработки более совершенных систем, способных к комплексному прочтению изображения и составлению его целостного образа (холистическое изображение).

Ardila и соавт. [33] разработали систему глубокого обучения, в которой нейронная сеть обучалась для выполнения сложных задач на основании не аннотированных или не размеченных изображений (без указания локализации и характеристики патологических очагов), т.е. не обработанных «вручную» изображений НДКТ. Для этого была использована наиболее совершенная система – глубокая сверточная нейронная сеть (deep convolutional neural networks – CNN), нацеленная на эффективное распознавание образов. Этот процесс фактически имитирует прочтение НДКТ рентгенологом, включая просмотр полного объема изображений, концентрации внимания на «подозрительных» участках, сегментацию изображения, уточнение локализации и оценку риска наличия рака. Работа включала три компонента: 1. разработку трехмерной CNN модели, которая анализирует весь объем не обработанных «вручную», не аннотированных изображений НДКТ, гистологически подтвержденных случаев РЛ; 2. разработку (тренинг) CNN модели, направленную на идентификацию на полном объеме изображений НДКТ участков, представляющих интерес, т.е. подозрительных с точки зрения наличия рака; 3. разработку CNN модели для прогностической оценки вероятности развития рака, которая функционирует на основании (на выходе) предыдущих моделей.

Модель глубокого обучения для анализа НДКТ изображений для скрининга РЛ разработана на основе базы данных лиц, прошедших скрининг в рамках Национальной программы скрининга РЛ в США (NLST). Были проанализированы 42 290 НДКТ изображений 14 851 пациентов, из которых у 578 в течение 1 года был диагностирован и гистологически подтвержден РЛ. Методом случайной выборки пациенты разделены на 3 группы: 1. группа трейнинга (обучения) – 70%; 2. группа разработки (тунинга) – 15%; 3. группа тестирования – 15%.

При прочтении изображений 6 716 пациентов из группы 1, из которых у 86 был РЛ, эффективность ИИ алгоритма была выше ($AUC = 94,4\%$) эффективности рентгенологов, которые оценивали изображения на основании классификационной системы Американского общества радиологов. Группа 2, в которую входили 507 пациентов, из которых у 83 был РЛ, AUC ИИ модели была выше (95,9%), чем у 6 высококвалифицированных рентгенологов, независимо от того, к какой категории классификации LUNG-RADS относилась патология в легком. При валидации на 1 130 случаях, полученных из внешнего источника, эффективность также была высокой – 94,2%. Прочтение ИИ изображений НДКТ привело к снижению количества ложноположительных и ложноотрицательных результатов на 11% и 5%, соответственно.

Модель ИИ глубокого обучения (SYBIL) разработана с использованием НДКТ от 15 тыс. участников NLST, США [34]. SYBIL готова для практического использования и может «работать» рядом с рентгенологом на его рабочем месте. Она не нуждается в аннотации демографических, клинических и радиологических данных. Разработка модели проводилась в 3-х последовательных группах: обучения, разработки и тестирования. Тестирование модели проведено на 3 независимых базах данных. Модель предсказывает риск развития рака легкого через 1 год ($AUC = 0,92\%$, ДИ 0,86 и 0,95), 2 года ($AUC = 0,86\%$, ДИ 0,82, 0,87), 3 года ($AUC = 0,80\%$, ДИ 0,79, 0,87), 4 года ($AUC = 0,77\%$, 0,79, 0,71), 5 лет ($AUC = 0,75\%$, 0,78, 0,77) и 6 лет (0,75, 0,74). Это преимущество модели ИИ SYBIL, по сравнению с традиционными компьютерными моделями и LUNG-RADS, стала еще более очевидной при их сравнении. Авторы приводят примеры пациентов, у которых клиническая оценка, данная на основании LUNG-RADS, соответствовала низкому риску РЛ (low risk score 1 или 2), в то время, как клиническая оценка риска рака, данная Sybil, была достаточно высокой

(>60% процентиля). Среди 55 611 пациентов, которые ежегодно проходили скрининг НДКТ и у которых вероятность риска по LUNG-RADS была низкой (1 или 2), AUC прогноза риска по SYBIL на 1 и 2 года были высокими – 0,86 (95% ДИ, 0,76–1,0) и 0,79 (95% ДИ, 0,73–0,85), соответственно. На основании этих цифр делается вывод, что прогностическая эффективность ИИ модели SYBIL способствует «своевременной» диагностике интервальных раков, которые, как известно, не попадают в поле зрения при традиционном скрининге. Оценка ИИ SYBIL изображений НДКТ не обязательно соответствует тому, как человек (рентгенолог) подошел бы к их анализу. Модель правильно предсказывает локализацию будущей опухоли и вероятность ее развития на данной локализации, участке. Соответственно, ИИ модель использует «сигнал», основанный на локальных особенностях изображения, а не на изменениях во всем органе, отображаемых на полном объеме НДКТ. Кроме того, выяснилось, что SYBIL может определить статус курения на основании НДКТ изображений.

Проблема интервальных раков, т.е. раков, которые диагностируются на основании симптомов у пациентов в период между очередными раундами скрининга, а точнее, прошедших очередной тур скрининга, известна давно, однако в литературе обсуждается редко. Интервальные раки обычно диагностируются на поздних стадиях и имеют плохой прогноз, низкие показатели выживаемости и высокой смертности. Результаты исследований, указывающие на эффективность ИИ для прогнозирования риска развития рака на несколько лет вперед, могут способствовать решению этой проблемы. ИИ распознает на НДКТ изображениях участки, которые она (система), в отличие от рентгенолога, рассматривает как «предрак», а возможно и состоявшийся рак, на основании одной ей известных признаков. Это определяемая ИИ «патология», а точнее, отличная от нормы структура, может и быть предшественником интервального рака, который ускользает от внимания рентгенолога. Исследование этого феномена, т.е. прогнозирование ИИ риска рака на несколько лет вперед, имеет и другое приложение – сокращение сроков между раундами скрининга, что приведет к снижению вреда, наносимого пациенту радиацией, который, хоть и минимальный, но существует, и рассматривается как один из отрицательных эффектов скрининга.

Скрининг РПЖ. Два ключевых РКИ, цитируемых в научной литературе, проведены в США

(PLCO) и в Европе [35, 36]. В исследование, проведенном в США (PLCO), включены 76 685 мужчин в возрасте 55–74 лет. Потенциальные участники скрининга методом рандомизации были разделены на 2 группы: группу скрининга (38 340 мужчин), с ежегодным тестированием на ПСА в течение 6 лет + ежегодное пальцевое исследование в течение 4 лет, и контрольную группу (38 345 мужчин). В результате 13-летнего наблюдения за когортой в опытной группе зарегистрировано 4 250 новых случаев РПЖ (кумулятивная заболеваемость 108,4/10 000 человеко-лет), а в контрольной – 3 815 новых случаев РПЖ (кумулятивная заболеваемость 97,1/10 000 человеко-лет). В опытной группе от РПЖ умерли 158 мужчин (кумулятивная смертность 3,7/10 000 человеко-лет), в контрольной – 145 мужчин (кумулятивная смертность 3,4/10 000 человеко-лет). Таким образом, в группе, в которой проводилось тестирование на ПСА, заболеваемость РПЖ была статистически значимо выше, чем в контрольной группе ($OR = 1,12$ при 95%-ном ДИ 1,07–1,17). В то же время, скрининг-ПСА не привел к снижению смертности от РПЖ ($OR = 1,09$ при 95%-ном ДИ 0,87–1,36), что указывает на отрицательный результат [35].

В многоцентровое исследование, которое проводилось в нескольких странах Европы (EMCRCST- European multicenter randomized clinical screening trial), были включены 136 689 мужчин в возрасте 55–69 лет. Методом рандомизации участники исследования были разделены на 2 группы: опытную (72 891 человек), участникам которой тестирование на ПСА проводилось ежегодно в течение 4 лет, и контрольную (89 352). В результате 11-ти летнего наблюдения в опытной группе выявлено 6 963 новых случаев РПЖ (кумулятивная заболеваемость 9,7/1000 человеко-лет), а в контрольной – 5 396 (кумулятивная заболеваемость 6,0/1000 человеко-лет). Анализ результатов исследования показал статистически значимое повышение заболеваемости РПЖ в группе, в которой проводился скрининг, по сравнению с контрольной ($OR = 1,63$ при 95%-ном ДИ 1,57–1,69). В опытной группе от РПЖ умерли 299 человек (кумулятивная смертность 0,39/1000 человеко-лет), в контрольной – 462 (кумулятивная смертность 0,50/1000 человеко-лет). Анализ показал статистически значимое снижение (на 21%) смертности от РПЖ в группе скрининга по сравнению с контрольной группой ($OR = 0,79$ при 95%-ном ДИ 0,68–0,91). Различий в общей смертности между двумя группами не обнаружено. Абсолютное

снижение смертности от РПЖ в опытной группе составило 1,07 на 1000 человек, т.е., чтобы сохранить жизнь 1 больному РПЖ, нужно протестировать 1 055 человек и выявить 33 случая рака [36]. Во всех РКИ скрининга РПЖ отмечен феномен гипердиагностики. Процент гипердиагностики зависит от многих факторов и значительно варьирует (12%–68%).

Полученные в РКИ результаты, указывающие на отсутствие выгоды (PLCO) или весьма скромную выгоду в смертности от РПЖ (EMCRCST) у пациентов, прошедших ПСА-тестирование, отсутствие выгоды в смертности от всех причин, наносимый гипердиагностикой вред, привели к его переоценке. Многие профессиональные организации в США и других странах рекомендовали от него отказаться [37–39]. Однако, поиск новых подходов и новых методов скрининга способствовал его возвращению, но в новом качестве. Предложен был индивидуальный подход, получение от потенциальных участников скрининга информированного согласия, после того как врач проведет с ними беседу о пользе и вреде ПСА-скрининга. Далее ПСА-тестирование дополнили 4К тестом (4П score), который включает общий ПСА, свободный ПСА, интактный ПСА и hK2 (человеческий калликрейн). Этот тест дифференцирует между Glisson G6 и G7. При «положительном» результате у пациента рекомендуется брать биопсию, но только под контролем МРТ, а не УЗИ, как это было принято в течение многих десятилетий. МРТ и мультипараметрическая (мп)МРТ стали неотъемлемой частью скрининга и диагностики РПЖ [40, 41]. Для стандартизации результатов оценки и интерпретации изображений МРТ разработана PI-RADS v.1, v.2 (the prostate imaging and reporting and data system). Внедрение в практику PI-RADS привело к значительному (с 60 до 90%) улучшению точности диагностики [42]. Однако, субъективность и вариабельность оценки изображений, которые характерны для рентгенологов (человеческий фактор), присутствуют и при использовании PI-RADS [43].

Применение алгоритмов ИИ машинного обучения для прочтения изображений мпМРТ устраняет или снижает недостатки PI-RADS, характерные для человека, т.е. субъективизм, гетерогенность результатов и вариабельность в оценке изображений; ИИ эффективно дифференцирует между отсутствием и наличием рака; опухолями с разными уровнями шкалы Глисона [44]. В исследовании Chen и соавт. [45] мпМРТ в комбинации с ИИ машинного обучения

показала высокую чувствительность и специфичность в дифференциальной диагностике и оценке агрессивности РПЖ по сравнению с PI-RADS. При дифференциальной диагностике РПЖ (т.е. РПЖ+/РПЖ-) эффективность ИИ модели (AUC = 0,985, 0,982, 0,999), в зависимости от режима MPT изображения (T2WI и ADC), достоверно выше эффективности PI-RADS v2 (AUC = 0,867). Эффективность ИИ модели также была выше при дифференциальной диагностике между G6 и G7 (AUC ИИ модели = 0,888, 0,865, 0,930 и AUC PI-RADS v.2 = 0,763). Таким образом, модель ИИ показала высокую диагностическую эффективность и превосходит PI-RADS v2 в диагностике РПЖ и дифференциальной диагностике между Глисон 6 и Глисон 7.

Представляет интерес еще одна работа, выполненная в Китае [46]. В исследование были включены 4 747 пациентов с диагнозом РПЖ. Система ИИ диагностики рака простаты (Prostate Cancer Artificial Intelligence Diagnostic System (PCAIDS)) разработана на основании демографических, клинических, лабораторных данных и результатов УЗИ этих пациентов. Эффективность PCAIDS выше (AUC = 0,82%–0,85%), чем эффективность ПСА-тестов (общий ПСА и свободный ПСА) (0,58%–0,73%). При тестировании PCAIDS в 3 когортах количество ненужных биопсий снизилось на 32,2%, 17,6%, 26,3%; в то же время было пропущено 4,9%, 4,5%, 4,1% случаев клинически значимого РПЖ.

Систематический обзор и мета-анализ 12 работ по оценке эффективности ИИ машинного обучения в MPT диагностике клинически значимого РПЖ показал высокую суммарную эффективность (AUC = 0,86, 95% ДИ 0,81–0,91). Показатель AUC варьировал от 0,73 до 0,99% [47].

Таким образом, системы ИИ показали высокую эффективность, по сравнению с ПСА-тестированием, в дифференциальной диагностике клинически значимого РПЖ, что способствовало статистически достоверному снижению количества дополнительных приглашений и ненужных биопсий и, соответственно, снижению гипердиагностики.

Скрининг РТК. Цель скрининга РТК – выявление аденоматозных полипов и РТК на ранней стадии развития, и, в результате, снижение заболеваемости и смертности от РТК. Снижение заболеваемости РТК в результате скрининга достигается благодаря выявлению аденоматозных полипов, из которых развивается более 80% рака. Вероятность малигнизации аденоматозного полипа зависит от его размера [48]. В связи

с этим удалению подлежат аденоматозные полипы диаметром более 10 мм.

Для скрининга РТК применяются следующие методы: тест на скрытую кровь (gFOBT – проба с гваяковой смолой или и iFOBT – иммунохимический тест), сигмоидоскопия, колоноскопия и виртуальная колоноскопия. Эффективность скрининга с помощью теста на скрытую кровь (gFOBT) с последующей колоноскопией или сигмоидоскопией при положительных результатах подтверждена в нескольких РКИ. Суммарный анализ исследований показал, что скрининг раз в 2 года или ежегодно снижает смертность от РТК на 15–40%. Применение иммунохимического (iFOBT) теста значительно улучшает результаты скрининга. Чувствительность iFOBT для диагностики рака выше (61–91%), чем стандартного метода (Hemoccult II) gFOBT (25–38%). Специфичность iFOBT по данным разных авторов варьирует от 91 до 98% и слегка ниже, чем специфичность пероксидазных тестов (98–99%). Чувствительность iFOBT (HemeSelect) для диагностики аденоматозных полипов диаметром >10 мм в 2 раза выше, чем чувствительность Hemoccult II [49]. В 2010 г. опубликованы результаты первого РКИ по оценке эффективности сигмоидоскопии для скрининга РТК [50]. В группе скрининга, т.е. группе, участникам которой была проведена сигмоидоскопия, было 40 674 человека, в контрольной – 113 тыс. В результате 11-ти летнего наблюдения заболеваемость РТК в группе скрининга снизилась на 33% (ОР = 0,67 при 95%-ном ДИ 0,60–0,76), смертность – на 43% (ОР = 0,57 при 95%-ном ДИ 0,45–0,72). Заболеваемость раком дистальных отделов, т.е. сигмовидной и прямой кишки, снизилась на 50% (при 95%-ном ДИ 0,42–0,59). Однако, скрининг не повлиял на заболеваемость и смертность от рака проксимальных отделов толстой кишки. Авторы исследования пришли к заключению, что сигмоидоскопия – эффективный метод скрининга РТК, и что ее достаточно проводить раз в жизни после 55 лет.

В рандомизированном исследовании PLCO (Prostate Lung & Colorectal) наблюдение за участниками исследования в течение, в среднем, 11,5 лет (154 900 мужчин) показало снижение заболеваемости как дистальным (ОР = 0,71 при 95% ДИ от 0,64 до 0,80; $p < 0,001$), так и проксимальным РТК (ОР = 0,86 при 95% ДИ от 0,76 до 0,97; $p = 0,01$). Смертность от РТК в группе скрининга также была на 26% статистически достоверно ниже, чем в контрольной группе (ОР = 0,74 при 95% ДИ от 0,63 до 0,87; $p < 0,001$). Снижение

смертности от дистального рака, т.е. рака сигмовидной и прямой кишки, в группе, где проводилась сигмоидоскопия, была ниже на 50% ($OR = 0,50$ при 95%-ном ДИ $0,38-0,64$; $p < 0,001$), чем в контрольной группе. Различий в смертности от рака проксимального отдела между группами скрининга и контрольной не было. Авторы исследования пришли к заключению, что сигмоидоскопия – эффективный метод скрининга РТК и что ее достаточно проводить раз в жизни в возрасте 55–64 лет [51].

В США и других западных странах растет смертность от рака проксимальных отделов толстой кишки, в то время как заболеваемость и смертность от рака дистальных отделов толстой кишки снижается [52]. Это частично связано с неэффективностью скрининга для рака проксимальных отделов толстой кишки. В 2022 г. опубликованы результаты многонационального РКИ по оценке эффективности колоноскопии [53]. Потенциальные участники программы были разделены на 2 группы, в одной из которых проводилась колоноскопия (28 220 человек), в другой – колоноскопия в рамках скрининговой программы не проводилась (56 365 человек), что не исключает возможности того, что часть из них колоноскопию прошли как компонент рутинного профилактического осмотра. В течение 10-ти летнего (медиана) наблюдения за участниками исследования, в группе колоноскопии диагностировано 259 случаев колоректального рака, в контрольной группе – 622 случая. Риск колоректального рака составил в опытной группе 0,98%, а в контрольной – 1,20%, что указывает на снижение риска на 18% и статистически достоверному снижению отношения рисков ($OR = 0,82$; 95% ДИ $0,70-0,93$). Риск смерти от колоректального рака в группе скрининга – 0,28%, контрольной группе – 0,31%, что соответствует статистически недостоверному отношению рисков ($OR = 0,90$; 95% ДИ $0,64-1,16$). Разницы в смертности от всех причин не отмечено. Для того, чтобы предотвратить один случай колоректального рака, необходимо провести 455 колоноскопий. Полученные результаты, по мнению авторов, не лучше результатов РКИ, в которых изучалась эффективность сигмоидоскопии, и поэтому вряд ли будут способствовать распространению колоноскопического скрининга. Однако это не исключает поиска путей повышения эффективности этого метода, например, усиления эффективности эндоскопических методов с помощью ИИ технологий.

Накоплен достаточно большой опыт использования различных систем ИИ в сочетании с колоноскопией для улучшения ее результатов. В большинстве РКИ, в которых оценивается эффективность сочетания колоноскопии с ИИ, отмечено повышение выявляемости полипов, в основном за счет маленьких образований, не представляющих опасности малигнизации, и повышение количества вмешательств для удаления обнаруженных образований. Подобные вмешательства с точки зрения профилактики рака, т.е. предотвращения развития из них злокачественной опухоли, сомнительная мера, так как это вмешательство связано с риском, хоть и небольшим, кровотечения и повреждения стенки кишки, т.е. перфорации. [55].

Далее кратко представлены результаты РКИ, которые опубликованы недавно и не вошли ни в один из проведенных мета-анализов [56]. Основная цель исследования – оценить эффективность CAD (computer Aided Diagnostics) в тандеме с колоноскопией для снижения числа «пропущенных» при колоноскопии аденом, основана на гипотезе, что 50% РТК, который диагностируется у пациентов, прошедших скрининг, развивается из «пропущенных» при колоноскопии аденом. В исследовании участвовали 916 пациентов: 385 в группе колоноскопии совместно с ИИ и 398 – в группе обычной колоноскопии. Число выявленных аденом на одну колоноскопию было достоверно выше в группе с ИИ+колоноскопия, чем в группе колоноскопии без поддержки ИИ ($0,70$ vs $0,51$, $p = 0,015$; 314 аденом на 449 колоноскопий vs 238 аденом на 467 колоноскопий). Количество пропущенных аденом было достоверно ниже при использовании колоноскопии совместно с ИИ (19%), чем при обычной колоноскопии (36%), $p = 0,024$ [56].

ИИ система, которая была разработана и применена в данном исследовании, основана на сверточной нейронной сети для обработки в реальном времени цифрового видео сигнала, полученного с колоноскопа. Она накладывает изображение выявленного образования (полипа, рака) на изображение, получаемое эндоскопистом непосредственно с камеры колоноскопа. Эти ИИ системы являются дополнением к колоноскопу. Они обнаруживают «подозрительные» участки слизистой оболочки кишки и выделяют их, тем самым обращая на них внимание эндоскописта. Эта система тестирована разработчиками на 172 полных колоноскопических видео, в которых были представлены 263 полипа с гистологически подтвержденным

диагнозом (чувствительность – 99,6%, специфичность – 98%).

Применение ИИ при колоноскопическом исследовании можно считать успешным только в случае, если придерживаться мнения (гипотезы), что выявление и удаление аденоматозных полипов всех размеров, а не только больших, снижает риск развития рака. Однако далеко не все с этим согласны и считают, что повышение «выявляемости» аденом при использовании ИИ происходит за счет маленьких полипов, при том, что риск развития рака высок при больших аденомах (>10 мм). В связи с этим удаление маленьких аденом нецелесообразно как с точки зрения риска их малигнизации, так и в связи с риском повреждения кишки (кровотечение, перфорации). Так что вопрос эффективности ИИ в tandem с колоноскопией, несмотря на применение самых продвинутых систем ИИ, остается спорным, и зависит от того, какую цель мы преследуем, разрабатывая и обучая систему. А определение цели применения ИИ в tandem с колоноскопией требует проведения РКИ, в котором будут сравниваться 2 тактики профилактики РТК: повышение «выявляемости» аденом, независимо от их размеров и их удаление, и идентификацию, и удаление только больших аденом.

Скрининг РШМ. Скрининг РШМ направлен на выявление не только ранних форм рака, но и, в первую очередь, предраковых изменений, в частности – интраэпителиальных неоплазий (CIN I, II и III) различной степени дисплазии и интраэпителиального рака *in situ*. Особенностью скрининга РШМ является тот факт, что его эффективность определяется снижением заболеваемости и смертности, а не только смертности. Наиболее широко используемым типом скрининга РШМ является цитологическое исследование (тест по Папаниколау, *Pap smear*), который получил распространение в 1960-х гг. Его эффективность подтверждена в результате длительного наблюдения за динамикой заболеваемости и смертности в скандинавских странах, Исландии и Финляндии [57]. В результате разработки новых технологий для выявления ДНК вирусов папилломы человека (ВПЧ) появилась возможность использования этих технологий для скрининга РШМ [58].

Чувствительность ВПЧ-тестирования по данным, полученным в США и некоторых Европейских странах, равна 96%, в то время как чувствительность цитологического исследования почти в 2 раза ниже (56%). При этом специфичность методов практически одинаковая

и равна 91% и 96%, соответственно. Анализ РКИ показал статистически достоверное преимущество ВПЧ-тестирования, по сравнению с цитологическим исследованием в диагностике ЦИН 3+. Учитывая высокую эффективность ВПЧ-тестирования, в идеале, страны должны перейти к тестированию на ДНК выявление ВПЧ в качестве основного метода скрининга РШМ с последующим цитологическим исследованием мазков у женщин с положительным ВПЧ-тестом (ВОЗ) [59]. Ввиду высокой прогностичности отрицательного ВПЧ-теста женщинам с отрицательным тестом проходить следующее обследование в течение 7–10 лет не требуется, т.е. скрининг достаточно проводить 2 раза в течение жизни женщины, в 35 и 45 лет [59].

Для повышения качества цитологической диагностики мазков с использованием жидкостной технологии китайскими учеными разработана система ИИ глубокого обучения (Artificial Intelligence Cervical Cancer Screening-AICCS) [61]. AICCS с высокой эффективностью, чувствительностью и специфичностью диагностирует все изменения, характерные для слизистой оболочки шейки матки. Эффективность AICCS оценивалась в трех валидационных базах данных женщин, прошедших цитологический скрининг: основной базе данных, в которой проводилось обучение AICCS, и в двух внешних базах данных. Эффективность диагностики LSIL (low grade squamous intraepithelial lesion) соответствовала AUC0,950, 0,927 и 0,946 в основной базе данных и в 2 внешних базах данных, соответственно. Эффективность (AUC) диагностики HSIL (high grade squamous intraepithelial neoplasia) была соответственно 0,960, 0,896 и 0,930 в 3 валидационных базах данных. Диагностическая эффективность ASC-US (atypical squamous cell of undetermined significance) также высока с AUC0,923, 0,879 и 0,929. Чувствительность AICCS превышала 0,9%, а специфичность – 0,8% при диагностике всех форм патологии шейки матки во всех 3 валидационных когортах. Чувствительность, специфичность и эффективность (AUC) диагностики всех типов цитопатологии шейки матки была достоверно выше при применении AICCS совместно с цитологом по сравнению с эффективностью цитолога без поддержки ИИ ($p = 0,024, 0,002, 0,001$). Статистически достоверное преимущество в пользу прочтения мазков цитологом + AICCS по сравнению с цитологом без поддержки сохранилось при диагностике практически всех форм патологии слизистой оболочки шейки матки.

Внедрение ВПЧ-тестирования приведет к значительному снижению количества цитологических исследований и значительному снижению нагрузки на цитопатологов и медицинские учреждения, в которых проводится скрининг РШМ. Тем не менее, сохраняются проблемы, связанные с несовершенством цитологического метода, которые в значительной степени зависят от «человеческого фактора» и, в первую очередь, от квалификации специалиста, с вытекающими из этого последствиями.

Применение ИИ для диагностики патологии шейки матки можно считать успешным и дает основания для оптимизма. Внедрение апробированных систем ИИ, разработанных и обученных эффективному взаимодействию с цитопатологом в прочтении и оценке мазков, взятых с шейки матки, и диагностике ЦИН и рака шейки матки, снизит нагрузку на цитологов и медицинский персонал в целом.

Коротко о российском опыте применения и оценки эффективности ИИ для диагностики ЗНО. Диагностическая эффективность отечественных систем ИИ в онкологии детально изучена на примере маммографии [62–65]. В статье Солодкий и соавт. [63] представляют довольно подробный анализ международного опыта по использованию ИИ в маммографии, включая анализ эффективности отечественных систем ИИ (платформа Botkin.AI и Цельс). В исследовании дана положительная оценка диагностической эффективности этих систем.

По данным Арзамасова и соавт. [64] диагностическая точность врачей-рентгенологов (AUC) статистически значимо превосходила показатели 2 из 5 ИИ-сервисов. Ни один из представленных в исследовании ИИ-сервисов статистически значимо не превосшел по значению AUC «среднего» врача-рентгенолога. Метрики диагностической точности для «среднего» врача-рентгенолога составили: AUC 0,928 (95% ДИ 0,883–0,976), чувствительность 0,792 (95% ДИ 0,677–0,907), специфичность 0,940 (95% ДИ 0,874–1,000). Тем не менее, в статье подчеркивается роль ИИ в поддержке принятия врачебных решений, позволяющая снизить нагрузку на рентгенологов, повысить эффективность их работы, а также избежать пропуск заболеваний.

В статье Павловой и соавт. [65] представлены результаты исследования по оценке маммографических снимков у женщин, прошедших профилактическое исследование. Маммограммы 8 030 пациенток были проанализированы с помощью сервиса «Цельс» для просмотра

медицинских изображений. РМЖ выявлен у 13 (1,2%) женщин с использованием платформ ИИ и только у 0,2% без применения ИИ, при этом наибольшая доля РМЖ была выявлена в группе с маммографической плотностью D [65].

В российской и зарубежной печати опубликованы несколько статей по оценке эффективности систем ИИ в тандеме с флюорографией, рентгенографией и КТ органов грудной клетки [66, 67]. Arzamasov и соавт. [66] изучили эффективность 5 платформ ИИ и показали, что из них только у трех (Celsus, Lunit INSIGHT CXR, and qXR) результаты выявления узлов в легких соответствовали заявленным в спецификации (AUC = 0,956; 95% ДИ 0,918–0,994). По таким показателям как сегментация и классификация узлов эффективность ИИ была ниже рентгенологов и ниже заявленной в спецификации (AUC = 0,812; 95% ДИ 0,744–0,879). Как и следовало ожидать, применение ИИ не оказало значительного влияния на эффективность флюорографии грудной клетки, очень высокий процент ложноотрицательных результатов не претерпел изменений [67]. Пиллус и соавт. [68] проанализировали возможность применения технологий ИИ в качестве системы поддержки принятия решения врачей-рентгенологов для выявления узлов в легких при КТ органов грудной клетки на примере системы Botkin.AI и пришли к заключению, что платформа обладает высокой точностью обнаружения легочных узлов и что ее использование может быть полезным для предотвращения пропусков легочной патологии в условиях повышенной нагрузки на врачей-рентгенологов.

ЗАКЛЮЧЕНИЕ

Представленные исследования указывают на перспективность применения ИИ для диагностики ЗНО, особенно в контексте популяционного скрининга, при котором исследование проходят многие тысячи человек. В популяционных программах все ограничения, присущие диагностическим методам, усугубляются и, соответственно, растет вероятность вреда, наносимого скринингом. Это, в первую очередь, высокая частота ложноотрицательных, ложноположительных результатов и гипердиагностики, что сопряжено с необходимостью повторных приглашений, дополнительного обследования и биопсий, а также может приводить к чрезмерному, а и иногда и ненужному лечению.

Применение ИИ достоверно повышает эффективность диагностических методов, улучшает показатели чувствительности и специфичности, что является важнейшей задачей скрининговых программ. Эффективность ИИ для прогнозирования риска развития рака на несколько лет вперед может способствовать удлинению интервалов между раундами скрининга и, соответственно, снижению нагрузки на систему здравоохранения и экономии средств.

В заключение, важно подчеркнуть, что решение о внедрении в практику популяционного скрининга любой системы ИИ, с доказанной эффективностью в рамках клинических исследований, должно быть принято после ее апробирования на популяционном уровне. Необходимо разработать формы «информированного согласия», в которых подробно и объективно описаны все преимущества и недостатки применения ИИ по сравнению с принятой практикой.

ЛИТЕРАТУРА

- Breast Cancer Screening. IARC handbook of Cancer Prevention, International Agency for Research on Cancer, WHO, IARC Pres. 2002.
- European Commission, European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis, 2006.
- Cervix cancer screening. IARC Handbooks of cancer Prevention, International Agency for Research on Cancer, World Health Organization, IARC Press, 2005.
- European Commission. European Guidelines for Quality Assurance in Cervical Cancer Screening, 2nd ed. – Luxembourg: Office for Official Publications of the European Communities, 2008.
- Заридзе Д.Г. Профилактика рака. Руководство для врачей. – М.: ИМА-ПРЕСС, 2009. – 224 с. – 9 ил. – 63 табл.
- Давыдов М.И., Заридзе Д.Г. Скрининг злокачественных опухолей: современное состояние и перспективы. Вестник Московского Онкологического Общества. 2014; 3(606):2–6.
- Барчук А.А., Раскина Ю.В., Смирнова О.В., Беляев А.М., Багненко С.Ф. Скрининг онкологических заболеваний на уровне государственных программ: обзор, рекомендации и управление. Общественное здоровье. 2021; 1(1):19–31.
- Zaridze D.G., Boyle P, Smans M. International trends in prostatic cancer. Int J Cancer. 1984; 33(2): 223–230.
- Black W.C., Haggstrom D.A., Welch H. G. All-cause mortality in randomized trials of cancer screening. J Natl Cancer Inst. 2002; 94(3):167–173.
- Заридзе Д.Г., Максимович Д.М., Стилиди И.С. Новая парадигма скрининга и ранней диагностики: оценка пользы и вреда. Вопросы онкологии. 2020; 66(6):589–602.
- Заридзе Д.Г. Искусственный интеллект повышает эффективность скрининга и диагностики злокачественных опухолей. Национальная онкологическая программа (2030). 2024; 2:32–34.
- Tabár L., Vitak B., Chen H.H. et al. Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinoma mortality. Cancer. 2001; 91(9):1724–1731.
- Tabár L., Yen M.F., Vitak B. et al. Mammography service screening and mortality in breast cancer patients: 20-year follow-up before and after introduction of screening. Lancet. 2003; 361(9367):1405–1410.
- Nelson H.D., Tyne K., Naik A. et al. Screening for Breast Cancer: Systematic Evidence Review Update for the US Preventive Services Task Force. Report 10–05142-EF-1. U. S. Preventive Services Task Force Evidence Syntheses, formerly Systematic Evidence Reviews. Rockville, MD: Agency for Health Care Research and Quality (US). 2009.
- Swedish Organised Service Screening Evaluation Group. Reduction in breast cancer mortality from organized service screening with mammography: 1. Further confirmation with extended data. Cancer Epidemiol Biomarkers Prev. 2006; 15(1):45–51.
- Pisano E.D., Gatsonis C., Hendrick E. et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med. 2005; 353(17): 1773–1783.
- Eastern Cooperative Oncology Group – American College of Radiology Imaging Network Cancer Research Group. Digital Tomosynthesis Mammography and Digital Mammography in Screening Patients for Breast Cancer (NCT03233191). Philadelphia, PA: Eastern Cooperative Oncology Group – American College of Radiology Imaging Network Cancer Research Group; 2017. Clinicaltrials.gov/ct2/show/NCT03233191. Accessed December 18, 2017.
- Majid A.S., de Paredes E.S., Doherty R.D., Sharma N.R., Salvador X. Missed breast carcinoma: pitfalls and pearls. Radiographics. 2003; 23(4):881–895.
- Miglioretti D.L., Gard C.C., Carney P.A. et al. When radiologists perform best: the learning curve in screening mammogram interpretation. Radiology. 2009; 253(3): 632–640.
- Independent UK Panel on Breast cancer Screening. The benefits and harms of breast cancer screening: an independent review. Lancet. 2012; 380(9855):1778–1786.
- Lehman C.D., Arao R.F., Sprague B.L. et al. National Performance Benchmarks for Modern Screening Digital Mammography: Update from the Breast Cancer Surveillance Consortium. Radiology. 2017; 283(1):49–58.
- Salim M., Dembrower K., Eklund M., Lindholm P, Strand F. Range of Radiologist Performance in a Population-based Screening Cohort of 1 Million Digital Mammography Examinations. Radiology. 2020; 297(1):33–39.
- Kwee T.C., Kwee R.M. Workload of diagnostic radiologists in the foreseeable future based on recent scientific advances: growth expectations and role of artificial intelligence. Insights Imaging. 2021; 12(1):88.
- Fenton J.J., Abraham L., Taplin S.H. et al. Effectiveness of computer-aided detection in community mammography practice. J Natl Cancer Inst. 2011; 103(15):1152–1161.
- Lehman C.D., Wellman R.D., Buist D.S. et al. Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection. JAMA Intern Med. 2015; 175(11):1828–1837.
- Ernst V.L., Ballard-Barbash R., Barlow W.E. et al. Detection of ductal carcinoma in situ in women undergoing screening mammography. J Natl Cancer Inst. 2002; 94(20):1546–1554.

27. Kim H.-E., Kim H.H., Han B.-K. et al. Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study. *Lancet Digit Health*. 2020; 2(3): e138-e148.
28. Dembrower K., Crippa A., Colón E., Eklund M., Strand F.; ScreenTrustCAD Trial Consortium. Artificial intelligence for breast cancer detection in screening mammography in Sweden: a prospective, population-based, paired-reader, non-inferiority study. *Lancet Digit Health*. 2023; 5(10): e703-e711.
29. National Lung Screening Trial Research Team; Aberle D.R., Adams A.M., Berg C.D. et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011; 365(5):395–409.
30. De Koning H.J., van der Aalst C.M., de Jong P.A. et al. Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. *N Engl J Med*. 2020; 382(6):503–513.
31. Bach P.B., Mirkin J.N., Oliver T.K. et al. Benefits and harms of CT screening for lung cancer: a systematic review. *JAMA*. 2012; 307(22):2418–2429.
32. Bonney A., Malouf R., Marchal C. et al. Impact of low-dose computed tomography (LDCT) screening on lung cancer-related mortality. *Cochrane Database Syst Rev*. 2022; 8(8): CD013829.
33. Ardila D., Kiraly A.P., Bharadwaj S. et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med*. 2019; 25(6):954–961.
34. Mikhael PG, Wohlwend J., Yala A. et al. Sybil: A Validated Deep Learning Model to Predict Future Lung Cancer Risk From a Single Low-Dose Chest Computed Tomography. *J Clin Oncol*. 2023; 41(12):2191–2200.
35. Andriole G.L., Crawford E.D., Grubb R.L. 3rd et al. Prostate cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: mortality results after 13 years of follow-up. *J Natl Cancer Inst*. 2012; 104(2):125–132.
36. Schröder F.H., Hugosson J., Roobol M.J. et al. Prostate-cancer mortality at 11 years of follow-up. *N Engl J Med*. 2012; 366(11):981–990.
37. Moyer V.A.; U.S. Preventive Services Task Force. Screening for Prostate Cancer: US Preventive Services Task Force Recommendation Statement. *Ann Intern Med*. 2012; 157(2):120–134.
38. Carter H.B., Albertsen P.C., Barry M.J. et al. Early Detection of Prostate Cancer: AUA Guideline. *J Urol*. 2013; 190(2):419–426.
39. Qaseem A., Barry M.J., Denberg T.D. et al. Screening for prostate cancer: a guidance statement from the Clinical Guidelines Committee of the American College of Physicians. *Ann Intern Med*. 2013; 158(10):761–769.
40. Yakar D., Debats O.A., Bormers J.G. et al. Predictive value of MRI in the localization, staging, volume estimation, assessment of aggressiveness, and guidance of radiotherapy and biopsies in prostate cancer. *J Magn Reson Imaging*. 2012; 35(1):20–31.
41. Moradi M., Salcudean S.E., Chang S.D. et al. Multiparametric MRI maps for detection and grading of dominant prostate tumors. *J Magn Reson Imaging*. 2012; 35(6):1403–1413.
42. Zhao C., Gao G., Fang D. et al. The efficiency of multiparametric magnetic resonance imaging (mpMRI) using PI-RADS Version 2 in the diagnosis of clinically significant prostate cancer. *Clin Imaging*. 2016; 40(5):885–888.
43. Kasel-Seibert M., Lehmann T., Aschenbach R. et al. Assessment of PI-RADS v2 for the Detection of Prostate Cancer. *Eur J Radiol*. 2016; 85(4):726–731.
44. Fehr D., Veeraraghavan H., Wibmer A. et al. Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images. *Proc Natl Acad Sci U S A*. 2015; 112(46):6265–6273.
45. Chen T., Li M., Gu Y. et al. Prostate Cancer Differentiation and Aggressiveness: Assessment With a Radiomic-Based Model vs. PI-RADS v2. *J Magn Reson Imaging*. 2018; 49(3):875–884.
46. Zhang H., Ji J., Liu Z. et al. Artificial intelligence for the diagnosis of clinically significant prostate cancer based on multimodal data: a multicenter study. *BMC Med*. 2023; 21(1):270.
47. Cuocolo R., Cipullo M.B., Stanzione A. et al. Machine learning for the identification of clinically significant prostate cancer on MRI: a meta-analysis. *Eur Radiol*. 2020; 30(12):6877–6887.
48. Clark J.C., Collan Y., Zaridze D.G. et al. Prevalence of polyps in an autopsy series from areas with varying incidence of large-bowel cancer. *Int J Cancer*. 1985; 36(2):179–186.
49. Mandel J.S., Church T.R., Bond J.H. et al. The effect of fecal occult-blood screening on the incidence of colorectal cancer. *N Engl J Med*. 2000; 343(22):1603–1607.
50. Atkin W.S., Edwards R., Kralj-Hans I. et al.; UK Flexible Sigmoidoscopy Trial Investigators. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. *Lancet*. 2010; 375(9726):1624–1633.
51. Schoen R.E., Pinsky P.F., Weissfeld J.L. et al.; PLCO Project Team. Colorectal-cancer incidence and mortality with screening flexible sigmoidoscopy. *N Engl J Med*. 2012; 366(25):2345–2357.
52. Zhu C., Bassig B.A., Zaridze D. et al. A birth cohort analysis of the incidence of ascending and descending colon cancer in the United States, 1973–2008. *Cancer Causes Control*. 2013; 24(6):1147–1156.
53. Bretthauer M., Løberg M., Wieszczyn P. et al.; NordICC Study Group. Effect of Colonoscopy Screening on Risks of Colorectal Cancer and Related Death. *N Engl J Med*. 2022; 387(17):1547–1556.
54. Hassan C., Spadaccini M., Mori Y. et al. Real-Time Computer-Aided Detection of Colorectal Neoplasia During Colonoscopy: A Systematic Review and Meta-analysis. *Ann Intern Med*. 2023; 176(9):1209–1220.
55. Lou S., Du F., Song W. et al. Artificial intelligence for colorectal neoplasia detection during colonoscopy: a systematic review and meta-analysis of randomized clinical trials. *EClinicalMedicine*. 2023; 66:102341.
56. Maas M.H.J., Neumann H., Shirin H. et al. A computer-aided polyp detection system in screening and surveillance colonoscopy: an international, multicentre, randomised, tandem trial. *Lancet Digit Health*. 2024; 6(3): e157-e165.
57. International Agency for Research on Cancer; World Health Organization. Cervix cancer screening. IARC Handbooks of cancer Prevention. IARC. 2005.
58. Cox J.T., Castle P.E., Behrens C.M. et al.; Athena HPV study group. Comparison of cervical cancer screening strategies incorporating different combinations of cytology, HPV testing, and genotyping for HPV 16/18: results from the ATHENA HPV study. *Am J Obstet Gynecol*. 2013; 208(3):184.e1–184.e11.
59. WHO guideline for screening and treatment of cervical pre-cancer lesions for cervical cancer prevention, second edition. Geneva: World Health Organization. 2021.
60. Заридзе Д.Г., Стилиди И.С., Мукерия А.Ф. Научное обоснование эффективности первичной и вторичной (скрининга) профилактики рака шейки матки. *Общественное здоровье*. 2022; 2(4):15–23.

61. Wang J., Yu Y., Tan Y. et al. Artificial intelligence enables precision diagnosis of cervical cytology grades and cervical cancer. *Nat Commun.* 2024; 15(1):4369.
62. Солодкий В.А., Каприн А.Д., Нуднов Н.В., Харченко Н.В., Запиров Г.М., Дибирова Ш.М., Подольская М.В., Кунда М.А. Возможности искусственного интеллекта в оценке риска рака молочной железы на маммографических изображениях (клинические примеры). *Вестник российского научного центра рентгенорадиологии (вестник РНЦРР)*, 2023; 2023(1):25–32.
63. Солодкий В.А., Каприн А.Д., Нуднов Н.В., Харченко Н.В., Ходорович О.С., Запиров Г.М., Шерстнёва Т.В., Дибирова Ш.М., Канахина Л.Б. Современные системы поддержки принятия врачебных решений на базе искусственного интеллекта для анализа цифровых маммографических изображений. *Вестник рентгенологии и радиологии.* 2023; 104(2):151–162.
64. Арзамасов К.М., Васильев Ю.А., Владзимирский А.В., Омелянская О.В., Бобровская Т.М., Семенов С.С., Четвериков С.Ф., Кирпичев Ю.С., Павлов Н.А., Андрейченко А.Е. Применение компьютерного зрения для профилактических исследований на примере маммографии. *Профилактическая медицина.* 2023; 26(6):117–123.
65. Павлова В.И., Белая Ю.А., Воронцов А.Ю., Прищепов А.А., Князев С.М., Михайлов А.А., Ковалева А.В., Аревшатын Э.Г., Палтуев Р.М., Чёрная А.В., Захарова Н.А. Результаты научно-исследовательской работы Российского общества онкоммаммологов «Использование искусственного интеллекта для раннего выявления рака молочной железы». *Опухоли женской репродуктивной системы.* 2023; 19(2):54–60.
66. Arzamasov K., Vasilev Y., Zelenova M. et al. Independent evaluation of the accuracy of 5 artificial intelligence software for detecting lung nodules on chest X-rays. *Quant Imaging Med Surg.* 2024; 14(8):5288–5303.
67. Васильев Ю.А., Арзамасов К.М., Колсанов А.В., Владзимирский А.В., Омелянская О.В., Пестренин Л.Д., Нечаев Н.Б. Опыт применения программного обеспечения на основе технологий искусственного интеллекта на данных 800 тысяч флюорографических исследований. *Врач и информационные технологии.* 2023; 4:54–65.
68. Пилиус П.С., Дрокин И.С., Баженова Д.А., Маковская Л.А., Сеницын В.Е. Оценка перспектив использования технологий искусственного интеллекта для анализа КТ-изображений органов грудной клетки с целью выявления признаков злокачественных новообразований в легких. *Медицинская визуализация.* 2023; 27(2):138–146.

REFERENCES

1. Breast Cancer Screening. IARC handbook of Cancer Prevention, International Agency for Research on Cancer, WHO, IARC Pres. 2002.
2. European Commission, European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis, 2006.
3. Cervix cancer screening. IARC Handbooks of cancer Prevention, International Agency for Research on Cancer, World Health Organization, IARC Press, 2005.
4. European Commission. European Guidelines for Quality Assurance in Cervical Cancer Screening, 2nd ed. – Luxembourg: Office for Official Publications of the European Communities, 2008.
5. Zaridze D.G. Cancer prevention. Manual for physicians. – M.: IMA-PRESS, 2009. – 224 p. – 9 ill. – 63 tables. (In Russ.)
6. Davydov M.I., Zaridze D.G. Cancer screening: current status and prospects. *Bulletin of the Moscow Oncological Society.* 2014; 3(606):2–6. (In Russ.)
7. Barchuk A.A., Raskina Yu.V., Smirnova O.V., Belyaev A.M., Bagnenko S.F. Cancer screening at the level of state programs: review, recommendations and management. *Public Health.* 2021; 1(1):19–31. (In Russ.)
8. Zaridze D.G., Boyle P., Smans M. International trends in prostatic cancer. *Int J Cancer.* 1984; 33(2): 223–230.
9. Black W.C., Haggstrom D.A., Welch H.G. All-cause mortality in randomized trials of cancer screening. *J Natl Cancer Inst.* 2002; 94(3):167–173.
10. Zaridze D.G., Maksimovich D.M., Stilidi I.S. New paradigm of screening and early diagnosis: assessment of benefits and harms. *Vopr. Onkol.* 2020; 66(6):589–602. (In Russ.)
11. Zaridze D.G. Artificial intelligence improves the efficiency of screening and diagnostics of cancer. *National oncology program (2030).* 2024; 2:32–34. (In Russ.)
12. Tabár L., Vitak B., Chen H.H. et al. Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinoma mortality. *Cancer.* 2001; 91(9):1724–1731.
13. Tabár L., Yen M.F., Vitak B. et al. Mammography service screening and mortality in breast cancer patients: 20-year follow-up before and after introduction of screening. *Lancet.* 2003; 361(9367):1405–1410.
14. Nelson H.D., Tyne K., Naik A. et al. Screening for Breast Cancer: Systematic Evidence Review Update for the US Preventive Services Task Force. Report 10–05142-EF-1. U.S. Preventive Services Task Force Evidence Syntheses, formerly Systematic Evidence Reviews. Rockville, MD: Agency for Health Care Research and Quality (US). 2009.
15. Swedish Organised Service Screening Evaluation Group. Reduction in breast cancer mortality from organized service screening with mammography: 1. Further confirmation with extended data. *Cancer Epidemiol Biomarkers Prev.* 2006; 15(1):45–51.
16. Pisano E.D., Gatsonis C., Hendrick E. et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. *N Engl J Med.* 2005; 353(17):1773–1783.
17. Eastern Cooperative Oncology Group – American College of Radiology Imaging Network Cancer Research Group. Digital Tomosynthesis Mammography and Digital Mammography in Screening Patients for Breast Cancer (NCT03233191). Philadelphia, PA: Eastern Cooperative Oncology Group – American College of Radiology Imaging Network Cancer Research Group; 2017. [Clinicaltrials.gov/ct2/show/NCT03233191](https://clinicaltrials.gov/ct2/show/NCT03233191). Accessed December 18, 2017.
18. Majid A.S., de Paredes E.S., Doherty R.D., Sharma N.R., Salvador X. Missed breast carcinoma: pitfalls and pearls. *Radiographics.* 2003; 23(4):881–895.
19. Miglioretti D.L., Gard C.C., Carney P.A. et al. When radiologists perform best: the learning curve in screening mammogram interpretation. *Radiology.* 2009; 253(3): 632–640.
20. Independent UK Panel on Breast cancer Screening. The benefits and harms of breast cancer screening: an independent review. *Lancet.* 2012; 380(9855):1778–1786.

21. Lehman C.D., Arao R.F., Sprague B.L. et al. National Performance Benchmarks for Modern Screening Digital Mammography: Update from the Breast Cancer Surveillance Consortium. *Radiology*. 2017; 283(1):49–58.
22. Salim M., Dembrower K., Eklund M., Lindholm P., Strand F. Range of Radiologist Performance in a Population-based Screening Cohort of 1 Million Digital Mammography Examinations. *Radiology*. 2020; 297(1):33–39.
23. Kwee T.C., Kwee R.M. Workload of diagnostic radiologists in the foreseeable future based on recent scientific advances: growth expectations and role of artificial intelligence. *Insights Imaging*. 2021; 12(1):88.
24. Fenton J.J., Abraham L., Taplin S.H. et al. Effectiveness of computer-aided detection in community mammography practice. *J Natl Cancer Inst*. 2011; 103(15):1152–1161.
25. Lehman C.D., Wellman R.D., Buist D.S. et al. Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection. *JAMA Intern Med*. 2015; 175(11):1828–1837.
26. Ernster V.L., Ballard-Barbash R., Barlow W.E. et al. Detection of ductal carcinoma in situ in women undergoing screening mammography. *J Natl Cancer Inst*. 2002; 94(20):1546–1554.
27. Kim H.-E., Kim H.H., Han B.-K. et al. Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study. *Lancet Digit Health*. 2020; 2(3): e138–e148.
28. Dembrower K., Crippa A., Colón E., Eklund M., Strand F; ScreenTrustCAD Trial Consortium. Artificial intelligence for breast cancer detection in screening mammography in Sweden: a prospective, population-based, paired-reader, non-inferiority study. *Lancet Digit Health*. 2023; 5(10): e703–e711.
29. National Lung Screening Trial Research Team; Aberle D.R., Adams A.M., Berg C.D. et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011; 365(5):395–409.
30. De Koning H.J., van der Aalst C.M., de Jong P.A. et al. Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. *N Engl J Med*. 2020; 382(6):503–513.
31. Bach P.B., Mirkin J.N., Oliver T.K. et al. Benefits and harms of CT screening for lung cancer: a systematic review. *JAMA*. 2012; 307(22):2418–2429.
32. Bonney A., Malouf R., Marchal C. et al. Impact of low-dose computed tomography (LDCT) screening on lung cancer-related mortality. *Cochrane Database Syst Rev*. 2022; 8(8): CD013829.
33. Ardila D., Kiraly A.P., Bharadwaj S. et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med*. 2019; 25(6):954–961.
34. Mikhael PG, Wohlwend J., Yala A. et al. Sybil: A Validated Deep Learning Model to Predict Future Lung Cancer Risk From a Single Low-Dose Chest Computed Tomography. *J Clin Oncol*. 2023; 41(12):2191–2200.
35. Andriole G.L., Crawford E.D., Grubb R.L. 3rd et al. Prostate cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: mortality results after 13 years of follow-up. *J Natl Cancer Inst*. 2012; 104(2):125–132.
36. Schröder F.H., Hugosson J., Roobol M.J. et al. Prostate-cancer mortality at 11 years of follow-up. *N Engl J Med*. 2012; 366(11):981–990.
37. Moyer V.A.; U.S. Preventive Services Task Force. Screening for Prostate Cancer: US Preventive Services Task Force Recommendation Statement. *Ann Intern Med*. 2012; 157(2):120–134.
38. Carter H.B., Albertsen P.C., Barry M.J. et al. Early Detection of Prostate Cancer: AUA Guideline. *J Urol*. 2013; 190(2):419–426.
39. Qaseem A., Barry M.J., Denberg T.D. et al. Screening for prostate cancer: a guidance statement from the Clinical Guidelines Committee of the American College of Physicians. *Ann Intern Med*. 2013; 158(10):761–769.
40. Yakar D., Debats O.A., Bomers J.G. et al. Predictive value of MRI in the localization, staging, volume estimation, assessment of aggressiveness, and guidance of radiotherapy and biopsies in prostate cancer. *J Magn Reson Imaging*. 2012; 35(1):20–31.
41. Moradi M., Salcudean S.E., Chang S.D. et al. Multiparametric MRI maps for detection and grading of dominant prostate tumors. *J Magn Reson Imaging*. 2012; 35(6):1403–1413.
42. Zhao C., Gao G., Fang D. et al. The efficiency of multiparametric magnetic resonance imaging (mpMRI) using PI-RADS Version 2 in the diagnosis of clinically significant prostate cancer. *Clin Imaging*. 2016; 40(5):885–888.
43. Kasel-Seibert M., Lehmann T., Aschenbach R. et al. Assessment of PI-RADS v2 for the Detection of Prostate Cancer. *Eur J Radiol*. 2016; 85(4):726–731.
44. Fehr D., Veeraraghavan H., Wibmer A. et al. Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images. *Proc Natl Acad Sci U S A*. 2015; 112(46):6265–6273.
45. Chen T., Li M., Gu Y. et al. Prostate Cancer Differentiation and Aggressiveness: Assessment With a Radiomic-Based Model vs. PI-RADS v2. *J Magn Reson Imaging*. 2018; 49(3):875–884.
46. Zhang H., Ji J., Liu Z. et al. Artificial intelligence for the diagnosis of clinically significant prostate cancer based on multimodal data: a multicenter study. *BMC Med*. 2023; 21(1):270.
47. Cuocolo R., Cipullo M.B., Stanzione A. et al. Machine learning for the identification of clinically significant prostate cancer on MRI: a meta-analysis. *Eur Radiol*. 2020; 30(12):6877–6887.
48. Clark J.C., Collan Y., Zaridze D.G. et al. Prevalence of polyps in an autopsy series from areas with varying incidence of large-bowel cancer. *Int J Cancer*. 1985; 36(2):179–186.
49. Mandel J.S., Church T.R., Bond J.H. et al. The effect of fecal occult-blood screening on the incidence of colorectal cancer. *N Engl J Med*. 2000; 343(22):1603–1607.
50. Atkin W.S., Edwards R., Kralj-Hans I. et al.; UK Flexible Sigmoidoscopy Trial Investigators. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. *Lancet*. 2010; 375(9726):1624–1633.
51. Schoen R.E., Pinsky P.F., Weissfeld J.L. et al.; PLCO Project Team. Colorectal-cancer incidence and mortality with screening flexible sigmoidoscopy. *N Engl J Med*. 2012; 366(25):2345–2357.
52. Zhu C., Bassig B.A., Zaridze D. et al. A birth cohort analysis of the incidence of ascending and descending colon cancer in the United States, 1973–2008. *Cancer Causes Control*. 2013; 24(6):1147–1156.
53. Bretthauer M., Løberg M., Wieszczyn P. et al.; NordICC Study Group. Effect of Colonoscopy Screening on Risks of Colorectal Cancer and Related Death. *N Engl J Med*. 2022; 387(17):1547–1556.
54. Hassan C., Spadaccini M., Mori Y. et al. Real-Time Computer-Aided Detection of Colorectal Neoplasia During Colonoscopy: A Systematic Review and Meta-analysis. *Ann Intern Med*. 2023; 176(9):1209–1220.
55. Lou S., Du F., Song W. et al. Artificial intelligence for colorectal neoplasia detection during colonoscopy: a

- systematic review and meta-analysis of randomized clinical trials. *EClinicalMedicine*. 2023; 66:102341.
56. Maas M.H.J., Neumann H., Shirin H. et al. A computer-aided polyp detection system in screening and surveillance colonoscopy: an international, multicentre, randomised, tandem trial. *Lancet Digit Health*. 2024; 6(3): e157-e165.
 57. International Agency for Research on Cancer, World Health Organization. Cervix cancer screening. IARC Handbooks of cancer Prevention. IARC. 2005.
 58. Cox J.T., Castle P.E., Behrens C.M. et al.; Athena HPV study group. Comparison of cervical cancer screening strategies incorporating different combinations of cytology, HPV testing, and genotyping for HPV 16/18: results from the ATHENA HPV study. *Am J Obstet Gynecol*. 2013; 208(3):184.e1–184.e11.
 59. WHO guideline for screening and treatment of cervical pre-cancer lesions for cervical cancer prevention, second edition. Geneva: World Health Organization. 2021.
 60. Zaridze D.G., Stilidi I.S., Mukeria A.F. Scientific evidence for the effectiveness of primary and secondary (screening) prevention of cervical cancer. *Public Health*. 2022; 2(4):15–23. (In Russ.)
 61. Wang J., Yu Y., Tan Y. et al. Artificial intelligence enables precision diagnosis of cervical cytology grades and cervical cancer. *Nat Commun*. 2024; 15(1):4369.
 62. Solodkiy V.A., Kaprin A.D., Nudnov N.V., Kharchenko N.V., Zapiro G.M., Dibirova Sh.M., Podolskaya M.V., Kunda M.A. Artificial intelligence capabilities in breast cancer risk assessment on mammographic images (clinical examples). *Vestnik of the Russian Scientific Center of Roentgenoradiology*. 2023.1. (In Russ.)
 63. Solodkiy V.A., Kaprin A.D., Nudnov N.V., Kharchenko N.V., Khodorovich O.S., Zapiro G.M., Sherstneva T.V., Dibirova Sh.M., Kanakhina L.B. Contemporary Medical Decision Support Systems Based on Artificial Intelligence for the Analysis of Digital Mammographic Images. *Journal of Radiology and Nuclear Medicine*. 2023; 104(2):151–162. (In Russ.)
 64. Arzamasov K.M., Vasiliev Yu.A., Vladzmyrskyy A.V., Omelyanskaya O.V., Bobrovskaya T.M., Semenov S.S., Chetverikov S.F., Kirpichev Yu.S., Pavlov N.A., Andreychenko A.E. The use of computer vision for the mammography preventive research. *The Russian Journal of Preventive Medicine*. 2023; 26(6):117–123. (In Russ.)
 65. Pavlova V.I., Belaya Yu.A., Vorontsov A.Yu., Prishchepov A.A., Knyazev S.M., Mikhailov A.A., Kovaleva A.V., Arevshatyan E.G., Paltuev R.M., Chernaya A.V., Zakharova N.A. Results of research work Russian Society of oncomammologists “The use artificial intelligence for early detection of breast cancer”. *Tumors of female reproductive system*. 2023; 19(2):54–60. (In Russ.)
 66. Arzamasov K., Vasilev Y., Zelenova M. et al. Independent evaluation of the accuracy of 5 artificial intelligence software for detecting lung nodules on chest X-rays. *Quant Imaging Med Surg*. 2024; 14(8):5288–5303.
 67. Vasiliev Yu.A., Arzamasov K.M., Kolsanov A.V., Vladzmyrskyy A.V., Omelyanskaya O.V., Pestrenin L.D., Nechaev N.B. Experience of application artificial intelligence software on 800 thousand fluorographic studies. *Medical doctor and information technology*. 2023; 4:54–65. (In Russ.)
 68. Pilius P.S., Drokin I.S., Bazhenova D.A., Makovskaya L.A., Sinitsyn V.E. Evaluation of the prospects for using artificial intelligence technologies to analyze CT scans of the chest organs in order to identify signs of malignant neoplasms in the lungs. *Medical Visualization*. 2023; 27(2):138–146. (In Russ.)

ES

Perspectivas del uso de inteligencia artificial para mejorar la eficiencia del cribado de neoplasias malignas

D.G. Zaridze

Anotación

Introducción. La eficacia del cribado como una de las estrategias más eficaces para controlar los tumores malignos está fuera de toda duda. El cribado reduce el riesgo de diagnosticar el cáncer en una etapa tardía e identifica patologías precancerosas, previniendo así su desarrollo. Las posibles limitaciones y peligros del cribado incluyen la alta probabilidad de resultados falsos positivos y falsos negativos y de sobrediagnóstico. Las consecuencias son exámenes adicionales y tratamientos innecesarios y, a menudo, excesivos. Al mismo tiempo, el cribado a menudo no incluye patologías oncológicas de intervalo que se caracterizan por un curso agresivo. **Objetivo del estudio:** estudiar la eficacia de la inteligencia artificial (IA) para aumentar la sensibilidad y especificidad de la detección del cáncer y reducir la incidencia de falsos negativos, falsos positivos y sobrediagnóstico. **Materiales y métodos.** Revisión y análisis de datos científicos publicados sobre: a) cribado de cáncer de mama (CM), cáncer de pulmón (CL), cáncer de próstata (CP), cáncer de cuello uterino (CC) y cáncer de colon (CC); b) desarrollo y aplicación de IA para mejorar la eficacia de los programas de detección. Se realizaron búsquedas de publicaciones relevantes en las bases de datos de PubMed y Cochrane Library. **Resultados.** Durante el cribado mamográfico, la IA reduce el número de interpretaciones incorrectas de las mamografías, el número de llamadas repetidas, el número de biopsias con resultado negativo y aumenta la eficiencia de la interpretación de las mamografías independientemente de las características del órgano (glándula mamaria densa, calcificaciones). El uso de IA

FR

Perspectives d'utilisation de l'intelligence artificielle pour améliorer l'efficacité du dépistage des tumeurs malignes

D.G. Zaridzé

Annotation

Introduction. L'efficacité du dépistage comme l'une des stratégies les plus efficaces pour contrôler les tumeurs malignes ne fait aucun doute. Le dépistage réduit le risque de diagnostiquer un cancer à un stade tardif et permet d'identifier les pathologies précancéreuses, empêchant ainsi son développement. Les limites potentielles et les dangers du dépistage comprennent la forte probabilité de résultats faussement positifs, faussement négatifs et de surdiagnostic. Les conséquences sont des examens complémentaires et des traitements inutiles et souvent excessifs. Dans le même temps, le dépistage n'inclut souvent pas les pathologies du cancer d'intervalle, caractérisées par une évolution agressive. **Objectif de l'étude:** étudier l'efficacité de l'intelligence artificielle (IA) pour augmenter la sensibilité et la spécificité du dépistage du cancer et réduire l'incidence des faux négatifs, des faux positifs et du surdiagnostic. **Matériels et méthodes.** Examen et analyse des données scientifiques publiées sur: a) le dépistage du cancer du sein (BC), du cancer du poumon (LC), du cancer de la prostate (PC), du cancer du col de l'utérus (CC) et du cancer du côlon (CC); b) développement et application de l'IA pour améliorer l'efficacité des programmes de dépistage. Les bases de données PubMed et Cochrane Library ont été consultées pour trouver les publications pertinentes. **Résultats.** Lors du dépistage par mammographie, l'IA réduit le nombre d'interprétations incorrectes des mammographies, le nombre d'appels répétés, le nombre de biopsies avec un résultat négatif, et augmente l'efficacité de l'interprétation des mammographies quelles que soient les caractéristiques de l'organe (glande mammaire dense, calcifications). L'utilisation de l'IA

junto con la tomografía computarizada de dosis baja (LDCT) para la detección del CP no solo mejora el diagnóstico de sus diversas formas, sino que también predice el riesgo de desarrollar cáncer con varios años de anticipación. Una revisión sistemática y un metanálisis de 12 estudios que evaluaron la eficacia de la IA junto con la resonancia magnética multiparamétrica (mpMRI) de la próstata mostraron una alta eficacia general en el diagnóstico del cáncer de próstata clínicamente significativo, lo que contribuyó a una reducción estadísticamente significativa en el número de invitaciones adicionales y biopsias innecesarias. La eficacia de la IA en combinación con la colonoscopia, a pesar del uso de sus sistemas más avanzados (un sistema de aprendizaje profundo basado en una red neuronal convolucional), sigue siendo controvertida. La solución a este problema depende del objetivo que persigamos al desarrollar y entrenar el sistema: aumentar la «detectabilidad» de los adenomas y su extirpación independientemente de su tamaño, o identificar y extirpar sólo los adenomas grandes, de los cuales existe una alta probabilidad de desarrollarse cáncer. Es alentador el uso exitoso de la IA para el diagnóstico citológico de la patología cervical, incluidas todas las etapas de la neoplasia intraepitelial cervical (NIC). La introducción de sistemas de inteligencia artificial capacitados para interactuar con un citopatólogo en la lectura y evaluación del material citológico y el diagnóstico de NIC y cáncer de cuello uterino reducirá la carga para los citólogos y otro personal médico. *Conclusión.* Los estudios presentados indican la promesa del uso de la IA para el diagnóstico de neoplasias malignas, especialmente en el contexto del cribado poblacional, en el que se examinan muchos miles de personas. El uso de IA aumenta de manera confiable la eficiencia de los métodos de diagnóstico, mejora la sensibilidad y la especificidad, reduce la probabilidad de resultados falsos negativos, falsos positivos y sobrediagnóstico. La eficacia de la IA para predecir el riesgo de cáncer con años de antelación podría ayudar a alargar los intervalos entre las rondas de detección, reduciendo así la carga para el sistema sanitario y los costes. La decisión de introducir cualquier sistema de IA con eficacia demostrada en ensayos clínicos en un programa de detección poblacional debe tomarse sólo después de haber sido probado a nivel poblacional. Es necesario desarrollar formularios de «consentimiento informado» para los pacientes que detallen y describan objetivamente todas las ventajas y desventajas del uso de la IA en comparación con la práctica aceptada actualmente.

Palabras clave: cribado, inteligencia artificial, IA, cáncer de mama, cáncer de próstata, cáncer de pulmón, cáncer de colon, cáncer de cuello uterino

en conjunción con la tomodensitometría a faible dose (LDCT) pour le dépistage de la LC améliore non seulement le diagnostic de ses diverses formes, mais prédit également le risque de développer un cancer plusieurs années à l'avance. Une revue systématique et une méta-analyse de 12 études évaluant l'efficacité de l'IA en tandem avec l'imagerie par résonance magnétique multiparamétrique (IRMm) de la prostate ont montré une efficacité globale élevée dans le diagnostic du cancer de la prostate cliniquement significatif, ce qui a contribué à une réduction statistiquement significative du nombre de invitations supplémentaires et biopsies inutiles. L'efficacité de l'IA en combinaison avec la coloscopie, malgré l'utilisation de ses systèmes les plus avancés (un système d'apprentissage profond basé sur un réseau neuronal convolutif), reste controversée. La solution à ce problème dépend de l'objectif que nous poursuivons lors du développement et de la formation du système: augmenter la «déteabilité» des adénomes et leur élimination quelle que soit leur taille, ou identifier et éliminer uniquement les gros adénomes, à partir desquels il existe une forte probabilité de développement du cancer. L'utilisation réussie de l'IA pour le diagnostic cytologique de la pathologie cervicale, y compris tous les stades de la néoplasie intraépithéliale cervicale (CIN), est encourageante. L'introduction de systèmes d'IA formés pour interagir avec un cytopathologiste lors de la lecture et de l'évaluation du matériel cytologique et du diagnostic des CIN et du cancer du col de l'utérus réduira la charge de travail des cytologistes et des autres personnels médicaux. *Conclusion.* Les études présentées indiquent la promesse de l'utilisation de l'IA pour le diagnostic des tumeurs malignes, en particulier dans le contexte du dépistage en population, dans lequel plusieurs milliers de personnes sont examinées. L'utilisation de l'IA augmente de manière fiable l'efficacité des méthodes de diagnostic, améliore la sensibilité et la spécificité, réduit le risque de résultats faussement négatifs, faussement positifs et de surdiagnostic. L'efficacité de l'IA dans la prévision du risque de cancer des années à l'avance pourrait contribuer à allonger les intervalles entre les cycles de dépistage, réduisant ainsi le fardeau du système de santé et les coûts. La décision d'introduire dans un programme de dépistage de la population tout système d'IA dont l'efficacité a été prouvée lors d'essais cliniques ne devrait être prise qu'après avoir été testé au niveau de la population. Il est nécessaire de développer des formulaires de «consentement éclairé» pour les patients qui détaillent et décrivent objectivement tous les avantages et inconvénients de l'utilisation de l'IA par rapport aux pratiques actuellement acceptées.

Mots clés: dépistage, intelligence artificielle, IA, cancer du sein, cancer de la prostate, cancer du poumon, cancer du côlon, cancer du col de l'utérus.

ИНФОРМАЦИЯ ОБ АВТОРАХ / ABOUT THE AUTHORS

Заридзе Давид Георгиевич – член-корреспондент РАН, доктор медицинских наук, профессор, заведующий отделом клинической эпидемиологии ФГБУ «НМИЦ онкологии им. Н.Н. Блохина» Министерства здравоохранения Российской Федерации, г. Москва, Россия.

David G. Zaridze – MD, Grand PhD in Medical sciences, Corresponding Member of the Russian Academy of Sciences, Head of the Department of Clinical Epidemiology, National Medical Research Center of Oncology named after N.N. Blokhin, Moscow, Russia.

E-mail: dgzaridze@rcs-pror.org, ORCID: 0000-0002-2824-3704, SPIN-код: 9739-1250