ISSN 1607-0763 (Print); ISSN 2408-9516 (Online)
https://doi.org/10.24835/1607-0763-1425

Обзор метаанализов о применении искусственного интеллекта в лучевой диагностике

© **Васильев Ю.А., Владзимирский А.В., Омелянская О.В., Решетников Р.В., Блохин И.А., Коденко М.Р., Нанова О.Г.***

ГБУЗ города Москвы "Научно-практический клинический центр диагностики и телемедицинских технологий ДЗ города Москвы"; 127051 Москва, ул. Петровка, д. 24, стр. 1, Российская Федерация

Введение. Искусственный интеллект (ИИ) – эффективный инструмент автоматизации рутинных процедур в лучевой диагностике. Вопрос диагностической точности алгоритмов ИИ в обнаружении различных патологий на данных лучевых исследований вызывает большой интерес у научного сообщества: число работ постоянно растет, появляются метаанализы, посвященные данной тематике. В связи с большим числом публикуемых результатов и их разнообразием возникает необходимость в систематизации данных.

Цель исследования: зонтичный систематический обзор современных метаанализов применения ИИ в лучевой диагностике.

Материал и методы. Поиск англоязычных статей осуществлялся в базе PubMed. Для полнотекстового анализа было отобрано 38 систематических обзоров с метаанализами 2021–2023 гг. Извлеченные данные включали: цель, дизайн, модальность лучевых исследований, объем выборок, показатели качества включенных работ, показатели диагностической точности ИИ, параметры референтной методики, показатели клинической эффективности внедрения алгоритма ИИ. Методологическое качество включенных систематических обзоров было оценено с помощью инструмента AMSTAR-2.

Результаты. Почти половина (47%) всех включенных метаанализов посвящена диагностике, стадированию и сегментации злокачественных новообразований. Четыре метаанализа выполнены в области стоматологии и посвящены детекции челюстно-лицевых структур, еще 4 метаанализа посвящены диагностике поражений мозга. По 3 метаанализа были посвящены диагностике COVID-19 и диагностике переломов. По одному метаанализу выполнено в области диагностики следующих патологий: колоректальные полипы, пневмоторакс, легочная эмболия, остеопороз, аневризмы, рассеянный склероз, острое нарушение мозгового кровообращения, внутричерепное кровоизлияние, ожоги, риск задержки внутриутробного развития. В 35 (92%) метаанализах проведена оценка риска систематической ошибки. Основной инструмент для оценки рисков – QUADAS-2, его использовали в 28 (80%) метаанализах. Из 28 метаанализов риск систематической ошибки оценен как низкий в 14 (50%) обзорах, как средний – в 4 (14%), как высокий – в 10 (36%). Основные риски обусловлены несбалансированностью выборок по объему и составу, недостаточно подробным описанием использованных методик, малым числом проспектовых исследований и исследований с внешней валидацией данных. Обобщенные результаты свидетельствуют о том, что диагностическая точность ИИ сопоставима с диагностической точностью врачей, а иногда превосходит ее. Средние значения чувствительности, специфичности и площадь под ROC-кривой для ИИ и врачей составляют 85,2, 89,5, 93,5 и 84,4, 90,0, 92,8% соответственно. Впрочем, во многих исследованиях, сопоставляющих диагностическую точность ИИ и врачей, наблюдается недостаток информации по числу и опыту врачей, которые выступили в роли референс-теста. Результаты внедрения ИИ в диагностическую практику приводятся только в одном метаанализе.

Обсуждение. ИИ способен сократить время рентгенологического описания при несрочных обследованиях. В качестве инструмента верификации первичной врачебной классификации исследования ИИ эффективно выявлял ложноотрицательные результаты рентгенологов, но при этом эффективность выявления ложноположительных результатов оказалась неудовлетворительной. По результатам оценки качества систематических обзоров по AMSTAR-2 показано, что методики поиска, отбора и анализа литературы нуждаются в стандартизации и улучшении. Также необходима разработка специализированного инструмента для оценки качества систематических обзоров в области использования ИИ. В настоящее время ИИ является многообещающим инструментом оптимизации описания исследований лучевой диагностики благодаря высоким показателям диагностической точности. Однако необходимы дополнительные исследования по внедрению ИИ в клиническую практику. Также необходимы повышение качества методологии исследований и ее стандартизация.

Ключевые слова: искусственный интеллект; диагностическая точность; метаанализ; лучевая диагностика

Авторы подтверждают отсутствие конфликтов интересов.

Для цитирования: Васильев Ю.А., Владзимирский А.В., Омелянская О.В., Решетников Р.В., Блохин И.А., Коденко М.Р., Нанова О.Г. Обзор метаанализов о применении искусственного интеллекта в лучевой диагностике. *Медицинская визуализация.* 2024; 28 (3): 22–41. https://doi.org/10.24835/1607-0763-1425

Поступила в редакцию: 28.12.2023. **Принята к печати:** 04.04.2024. **Опубликована online:** 08.07.2024.

Review of meta-analyses on the use of artificial intelligence in radiology

© Yuri A. Vasiliev, Anton V. Vladzymyrskyy, Olga V. Omelyanskaya, Roman V. Reshetnikov, Ivan A. Blokhin, Maria R. Kodenko, Olga G. Nanova*

Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of Moscow Healthcare Department; 24, bld. 1, Petrovka str., Moscow 127051, Russian Federation

Introduction. Artificial intelligence (AI) is an effective tool for automating routine tasks in radiology. The diagnostic accuracy of AI in detecting various pathologies on medical images has generated considerable interest in the scientific community: the number of studies and meta-analyses has been constantly growing. The abundance of published evidence and the diversity of outcomes necessitate the need to systematize the available publications. The aim of this paper is to conduct an umbrella systematic review of contemporary meta-analyses on the use of AI in radiology.

Materials and methods. PubMed was searched for studies published in the English language. Thirty-eight systematic reviews with meta-analyses published between 2021 and 2023 were selected for full-text analysis. The extracted data included the goal, study design, imaging modality, sample size, quality assessment of the included studies, AI diagnostic accuracy estimates, reference method parameters, and clinical efficacy metrics of AI implementation. The methodological quality of included systematic reviews was assessed using the AMSTAR-2 tool.

Results. Nearly half (47%) of the included meta-analyses focused on the diagnosis, staging and segmentation of malignancies. Four meta-analyses were related to detection of maxillofacial structures in dentistry, while another four meta-analyses addressed the diagnosis of brain lesions. The diagnosis of COVID-19 and the diagnosis of bone fractures were each covered in three meta-analyses. One meta-analysis was reviewed for each of the following fields: colorectal polyps, pneumothorax, pulmonary embolism, osteoporosis, aneurysms, multiple sclerosis, acute cerebrovascular accident, intracranial hemorrhage, burns, and the risk of intrauterine growth restriction. Thirty-five (92%) meta-analyses assessed the risk of bias. Twenty-eight (80%) meta-analyses utilized QUADAS-2 to assess the risk of bias. 14 out of 28 papers reported low risk of bias (50%); 4 (14%) – moderate; 10 (36%) – high. The major risks were associated with samples that were unbalanced in terms of size and composition, a lack of details about the methods, a low number of prospective studies, and a lack of external validation of the outcomes. The overall results indicate that the diagnostic accuracy of AI is comparable to or even greater than that of radiologists. The mean sensitivity, specificity and area under the ROC curve for AI and radiologists were 85.2%, 89.5%, 93.5% and 84.4%, 90.0%, 92.8%, respectively. However, many studies that compared the diagnostic accuracy of AI and radiologists lack the data on the number and experience of the latter. Only one paper presented results of implementing AI into routine clinical diagnosis.

Discussion. AI is capable of reducing the turnaround time for non-urgent examinations. When used to verify the primary interpretation, AI was effective in detecting false-negative results from radiologists. However, the efficacy of detecting false-positive results was inadequate. Our assessment of the quality of systematic reviews with AMSTAR-2 show that the methods of searching, selecting and analyzing literature must be improved and brought to a common standard. The development of a specialized tool for assessing the quality of systematic reviews in the AI implementation is also necessary. Due to high diagnostic accuracy, AI is currently considered a promising tool for optimizing the turnaround time. However, more evidence is needed to study the AI outcomes in routine clinical practice. Furthermore, it is necessary to standardize and improve the quality of research methodology.

Keywords: artificial intelligence; diagnostic accuracy; meta-analysis; medical imaging

Conflict of interest. The authors declare no conflict of interest. The study had no sponsorship.

For citation: Vasilev Y.A., Vladzimirskyy A.V., Omelyanskaya O.V., Reshetnikov R.V., Blokhin I.A., Kodenko M.R., Nanova O.G. Review of meta-analyses on the use of artificial intelligence in radiology. *Medical Visualization.* 2024; 28 (3): 22–41. https://doi.org/10.24835/1607-0763-1425

Received: 28.12.2023. **Accepted for publication:** 04.04.2024. **Published online:** 08.07.2024.

Введение

Искусственный интеллект (ИИ) – современный и мощный инструмент автоматизированной обработки лучевых исследований [1, 2]. Число исследований в области применения ИИ для решения диагностических задач стремительно растет, причем во многих исследованиях отмечаются высокая диагностическая точность ИИ и потенциал для его внедрения в клиническую практику [1, 2]. Впрочем, качество данных и используемые методики оценки диагностической точности зачастую вызывают сомнения. Кроме того, возникает вопрос, насколько в действительности ИИ может повысить качество и скорость диагностики по сравнению с врачами: подобные данные, хотя и представлены в литературе, крайне разнородны [3].

Стремительное развитие данной области сопровождается ростом публикаций и, как следствие, попытками систематизации данных [4, 5]: накопилось значительное число систематических обзоров и метаанализов диагностической точности ИИ в самых разных областях медицины [6]. Впрочем, унификация подходов при анализе данных остается актуальной задачей, что объясняет необходимость одновременного систематического рассмотрении области применения и разнообразия подходов, а также выявления основных проблемных точек с целью дальнейшей унификации проведения и представления результатов исследований в области ИИ.

Цель исследования: зонтичный систематический обзор существующих метаанализов применения ИИ в лучевой диагностике.

Методы

Исследование проводилось в соответствии с методологией PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) [7]. Дизайн исследования – зонтичный систематический обзор литературы.

Стратегия литературного поиска

Поиск осуществлялся по англоязычным статьям в базе PubMed и был ограничен временны́м интервалом 2021–2023 гг.

Поисковый запрос выглядел следующим образом:

“systematic review * OR meta-analysis * AND Neoplasm* OR Neoplasm, Lung* OR Radiotherapy * OR Cancer* OR Cancer, Lung* OR Pneumothorax* OR Fracture Detection* OR Fractures* OR bone* OR liver lesion* OR Thyroid nodule* OR Brain* OR Aneurysm* OR Lymph node* OR Metastasis OR Osteoporosis OR Cervical Cancer OR dentistry” AND “Intelligence, Artificial OR Computational Intelligence OR Intelligence, Computational OR Machine Intelligence OR Intelligence, Machine OR Computer Reasoning OR Reasoning, Computer OR AI (Artificial Intelligence) OR Computer Vision Systems OR Computer Vision System OR System, Computer Vision OR Systems, Computer Vision OR Vision System, Computer OR Vision Systems, Computer OR Knowledge Acquisition (Computer) OR Neural network”.

Критерии включения: метаанализы диагностической точности ИИ в лучевой диагностике.

Названия и аннотации найденных статей были независимо проанализированы двумя экспертами, по результатам была сформирована выборка для полнотекстового анализа. В качестве экспертов выступали научные сотрудники с опытом работы в лучевой диагностике и медицинской информатике более 5 лет. В случае расхождений относительно включения статьи в анализ привлекался третий эксперт – научный сотрудник с опытом работы в лучевой диагностике и медицинской информатике более 10 лет.

Извлечение информации и оценка методологического качества исследований

Из полных текстов отобранных статей была извлечена следующая информация:

1) библиометрические данные (имя первого автора, название статьи, год выхода, DOI, название журнала, импакт-фактор журнала, страна проведения исследования);

2) параметры включенных исследований (модальность, объем выборки, доля проспективных исследований, наличие тестирования на внешних данных);

3) метрики качества включенных исследований (риск систематической ошибки, неоднородность данных, объективность критериев включения статей в обзор – ”publication bias”);

4) значения диагностической точности ИИ; результаты сравнения диагностической точности ИИ с диагностической точностью врачей; информация по числу врачей и уровню их квалификации;

5) результаты внедрения.

Для оценки качества включенных систематических обзоров был использован инструмент AMSTAR-2 (A MeaSurement Tool to Assess systematic Reviews) [8].

Результаты

Поиск литературы и отбор работ

Результаты поискового запроса включали 2855 источников, для которых был проведен первичный анализ названий и аннотаций. В результа-

те применения критериев включения было исключено 2817 работ, которые не являлись обзорами либо принадлежали к иным предметным областям: медицина животных и гуманитарные науки. В итоговый анализ включено 38 метаанализов (табл. 1).

Базовые характеристики включенных исследований

Базовые характеристики отобранных для обзора статей приведены в Приложении 1: https://medvis.vidar.ru/jour/article/view/1425 – дополнительные файлы.

Из отобранных 38 систематических обзоров 10 выполнены в Китае, 5 – в Великобритании, 4 – в Европе (Германия + Греция, Италия, Румыния, Дания), 3 – в Иране; по 2 из Тайваня, Бразилии и Австралии, по одной работе из Японии, Сингапура, Тайланда, Кореи, Израиля, Саудовской Аравии и Канады. Одна работа – результат сотрудничества Кореи, Европы (Швеция + Австрия), Великобритании; 1 – сотрудничество Великобритании и Европы (Австрия); 1 – сотрудничество Канады и Китая. В Азии выполнено 55% всех систематических обзоров, в Европе – 26%, в Америке – 8%, в Австралии – 5%. В сотрудничестве между учеными Азии и Европы выполнено 3% обзоров, между Азией и Америкой – 3%.

Самые высокорейтинговые публикации были в журналах:

1) *"Radiology"* с импакт-фактором 19,7: [9] – работа выполнена в сотрудничестве ученых Азии и Европы, [10, 11] – обе работы выполнены в Европе;

2) *eClinicalMedicine* с импакт-фактором 15,1: [12] – работа выполнена в Азии;

3) *JAMA Netw Open* с импакт-фактором 13,8: [13] – работа выполнена в Америке.

Параметры включенных исследований

18 (47%) метаанализов из 38 посвящены онкологическим исследованиям (рис. 1): из них (рис. 2) злокачественным опухолям пищеварительной системы – 5 (28%) работ, дыхательной системы – 4 (22%), репродуктивной системы – 4 (22%), нервной системы – 2 (11%), эндокринной системы – 2 (11%), злокачественным опухолям забрюшинного пространства – 1 (6%).

4 (11%) метаанализа посвящены диагностике поражений мозга: 1 – диагностике рассеянного склероза; 1 – обнаружению аневризм; 1 – раннему обнаружению инсульта; 1 – обнаружению внутричерепного кровоизлияния. 4 (11%) метаанализа выполнены в области стоматологии, из них 3 посвящены обнаружению и сегментации структур и 1 – принятию решений об удалении зубов. 3 (8%) метаанализа посвящены диагностике поражений легких, вызванных COVID-19, 1 (3%) – обнаружению легочной эмболии и 1 (3%) – обнаружению пневмоторакса, 3 (8%) – диагностике переломов, где большая часть исследований выполнена на конечностях, 1 (3%) – диагностике остеопорозов, 1 (3%) – диагностике колоректальных полипов, 1 (3%) – диагностике ожогов (их глубине, сегментации ожогов и предсказанию смертности), 1 (3%) – оценке риска задержки внутриутробного развития.

В большинство метаанализов – 20 (53%) из 38 – включены исследования нескольких разных модальностей (рис. 3), в остальные 18 (43%) – исследования только одной модальности.

Компьютерная томография (КТ) использована в 17 (45%) метаанализах из 38. Из них как единственный метод КТ использована в 3 метаанализах. В остальных 14 метаанализах наряду с КТ включены и другие методы: рентгенография [11, 14–16]; рентгенография грудной клетки [17, 18]; цифровой томосинтез [18]; ПЭТ/КТ [19–21]; магнитно-резонансная томография (МРТ) [20–26]; эндоскопия [25]; низкодозная КТ [14]; ультразвуковое исследование (УЗИ) [12].

МРТ использована в 11 (29%) метаанализах из 38. Из них как единственный метод МРТ рассматривается в 3 метаанализах. В остальных 8 метаанализах наряду с МРТ включены и другие методы.

Рентгенография использована в 6 (16%) метаанализах [11, 13–16, 27], при этом в одном из них – как единственный метод [13].

Рентгенография грудной клетки [17, 18, 28, 29] использована в 4 (11%) метаанализах, при этом в двух из них – как единственный метод [28, 29].

ПЭТ/КТ [19–21] использовали в 2 (5%) метаанализах совместно с КТ.

Для диагностики злокачественных опухолей молочной железы использовали цифровую маммографию в 2 (5%) метаанализах [9, 10]. В одном из этих метаанлизов [9] наряду с цифровой маммографией использовали также томосинтез.

Томосинтез использовали в 2 (5%) метаанализах совместно с маммографией [9] и КТ [18].

УЗИ использовано в 6 (16%) метаанализах [12, 26, 30–33], а как единственный метод – в 4 метаанализах [12, 30, 32].

Конусно-лучевая компьютерная томография (КЛКТ) использована в 3 (8%) стоматологических метаанализах [27, 34, 35], из них как единственный метод – в 2 метаанализах [34, 35].

В 1 (3%) метаанализе [36], посвященном обнаружению легочной эмболии, использовали исключительно КТ-ангиографию.

Таблица 1. Список включенных в обзор метаанализов

Table 1. List of meta-analyses are included in the review

№	Первый автор, год First author, year	Название Title	Журнал Journal	Импакт-фактор Impact Factor
1	J.H. Yoon et al., 2023 [9]	Standalone AI for breast cancer detection at screening digital mammography and digital breast tomosynthesis: a systematic review and meta-analysis	*Radiology*	19.70
2	S.E. Hickman et al., 2023 [10]	Machine learning for workflow applications in screening mammography: systematic review and meta-analysis	*Radiology*	19.70
3	L.T. Thong et al., 2023 [18]	Diagnostic test accuracy of artificial intelligence-based imaging for lung cancer screening: a systematic review and meta-analysis	*Lung Cancer*	9.5
4	X. Zheng et al., 2022 [21]	Diagnostic accuracy of deep learning and radiomics in lung cancer staging: a systematic review and meta-analysis	*Front. Public Health*	6.45
5	M. Liu et al., 2023 [55]	The value of artificial intelligence in the diagnosis of lung cancer: a systematic review and meta-analysis	*PLoS ONE*	3.75
6	T. Sugibayashi et al., 2023 [17]	Deep learning for pneumothorax diagnosis: a systematic review and meta-analysis	*Eur. Respirat. Rev.*	9.55
7	R.Y.L.Kuo et al., 2022 [11]	Artificial intelligence in fracture detection: a systematic review and meta-analysis	*Radiology*	19.70
8	X. Zhang et al., 2022 [16]	Diagnostic accuracy and potential covariates of artificial intelligence for diagnosing orthopedic fractures: a systematic literature review and meta-analysis	*Eur. Radiol.*	7.03
9	J.R. Lex et al., 2023 [13]	Artificial intelligence for hip fracture detection and outcome prediction	*JAMA Network Open*	13.80
10	P. Potipimpanon et al., 2022 [32]	A comparison of artificial intelligence versus radiologists in the diagnosis of thyroid nodules using ultrasonography: a systematic review and meta analysis	*Eur. Arch. of Oto-Rhino-Laryngol.*	3.24
11	C.A. Campello et al., 2023 [30]	Machine learning for malignant versus benign focal liver lesions on US and CEUS: a meta analysis	*Abdom. Radiol.*	2.89
12	F. Nabizadeh et al., 2023 [46]	Diagnostic performance of artificial intelligence in multiple sclerosis: a systematic review and meta analysis	*Neurol. Sci.*	3.83
13	M.Din et al., 2023 [25]	Detection of cerebral aneurysms using artificial intelligence: a systematic review and meta-analysis	*J. NeuroIntervent. Surg.*	8.57
14	M.D. Jørgensen et al., 2023 [56]	Convolutional neural network performance compared to radiologists in detecting intracranial hemorrhage from brain computed tomography: A systematic review and meta-analysis	*Eur. J. Radiol.*	4.53
15	H.Y. Kim et al., 2021 [43]	Classification of true progression after radiotherapy of brain metastasis on MRI using artificial intelligence: a systematic review and meta-analysis	*Neuro-Oncology Advances.*	2.80
16	A. Adamou et al., 2023 [45]	Artificial intelligence-driven ASPECTS for the detection of early stroke changes in non-contrast CT: a systematic review and meta-analysis	*J. NeuroIntervent. Surg.*	8.57
17	S. Agarwal et al., 2023 [22]	Systematic review of Artificial Intelligence for abnormality detection in high-volume neuroimaging and subgroup meta-analysis for intracranial hemorrhage detection	*Clin. Neuroradiol.*	3.16
18	S. Bedrikovetski et al., 2021a [23]	Artificial intelligence for pre-operative lymph node staging in colorectal cancer: a systematic review and meta-analysis	*BMC Cancer*	4.40
19	S. Soffer et al., 2023 [36]	Deep learning for pulmonary embolism detection on computed tomography pulmonary angiogram: a systematic review and meta analysis	*Sci. Reports*	5.00
20	C. Yang et al., 2022 [40]	Deep learning in CT image segmentation of cervical cancer: a systematic review and meta analysis	*Radiation Oncology*	3.48

Таблица 1 *(окончание)*.
Table 1 *(end)*.

№	Первый автор, год First author, year	Название Title	Журнал Journal	Импакт-фактор Impact Factor
21	L. Gao et al., 2021 [14]	Application of artificial intelligence in diagnosis of osteoporosis using medical images: a systematic review and meta-analysis	*Osteoporosis International*	4.6
22	R. Rescinito et al., 2023 [33]	Prediction models for Intrauterine growth restriction using Artificial Intelligence and Machine Learning: a systematic review and meta-analysis	*Healthcare*	2.01
23	Q. Wang et al., 2022 [29]	Diagnostic performance of corona virus disease 2019 chest computer tomography image recognition based on deep learning Systematic review and meta-analysis	*Medicine*	1.60
24	T.N. Poly et al., 2021 [15]	Application of Artificial Intelligence for screening COVID-19 patients using digital images: meta-analysis	*JMIR Med. Inform*	3.23
25	T.E. Komolafe et al., 2023 [28]	Diagnostic test accuracy of deep learning detection of COVID-19: a systematic review and meta-analysis	*Acad. Radiol.*	5.48
26	F. Abesi et al., 2023 [34]	Accuracy of artificial intelligence in the detection and segmentation of oral and maxillofacial structures using cone-beam computed tomography images: a systematic review and meta-analysis	*Polish J. Radiol.*	1.20
27	F.F. Badr, F.M. Jadu, 2022 [35]	Performance of artificial intelligence using oral and maxillofacial CBCT images: a systematic review and meta-analysis	*Nigerian J. Clin. Pract.*	1.12
28	W. Xing et al., 2023 [20]	Artificial intelligence predicts lung cancer radiotherapy response: a meta-analysis	*Artif. Intell. Med.*	7.01
29	N. Menon et al., 2023 [19]	Performance of radiomics-based artificial intelligence systems in the diagnosis and prediction of treatment response and survival in esophageal cancer: a systematic review and meta-analysis of diagnostic accuracy	*Dis. Esophagus*	2.60
30	B.G. Taib et al., 2023 [39]	Artificial intelligence in the management and treatment of burns: A systematic review and meta-analyses	*J. Plast. Reconstructive & Aesthetic Surg.*	2.90
31	S. Sadr. et al., 2023 [27]	Deep Learning for detection of periapical radiolucent lesions: a systematic review and meta-analysis of diagnostic test accuracy	*J. Endodont.*	4.42
32	M. Islam et al., 2022 [37]	Deep Learning for the diagnosis of esophageal cancer in endoscopic images: a systematic review and meta-analysis	*Cancers*	5.20
33	K. Evangelista et al., 2022 [57]	Accuracy of artificial intelligence for tooth extraction decision making in orthodontics: a systematic review and meta analysis	*Clin. Oral Invest.*	3.61
34	Y. Xue et al., 2022 [12]	Accuracy of ultrasound diagnosis of thyroid nodules based on artificial intelligence-assisted diagnostic technology: a systematic review and meta-analysis	*Int J Endocrinol.*	2.80
35	H.-L. Xu et al., 2022 [26]	Artificial intelligence performance in image-based ovarian cancer identification: a systematic review and meta-analysis	*eClinicalMedicine*	15.1
36	E.A. Dumitrescu et al., 2022 [31]	Diagnostic value of artificial intelligence-assisted endoscopic ultrasound for pancreatic cancer: a systematic review and meta-analysis	*Diagnostics*	3.61
37	S. Bedrikovetski et al., 2021b [24]	Artificial intelligence for the diagnosis of lymph node metastases in patients with abdominopelvic malignancy: a systematic review and meta-analysis	*Artif. Intell. Med.*	7.01
38	Y. Xu et al., 2021 [38]	Comparison of diagnostic performance between convolutional neural networks and human endoscopists for diagnosis of colorectal polyp: a systematic review and meta-analysis	*PLoS ONE*	3.75

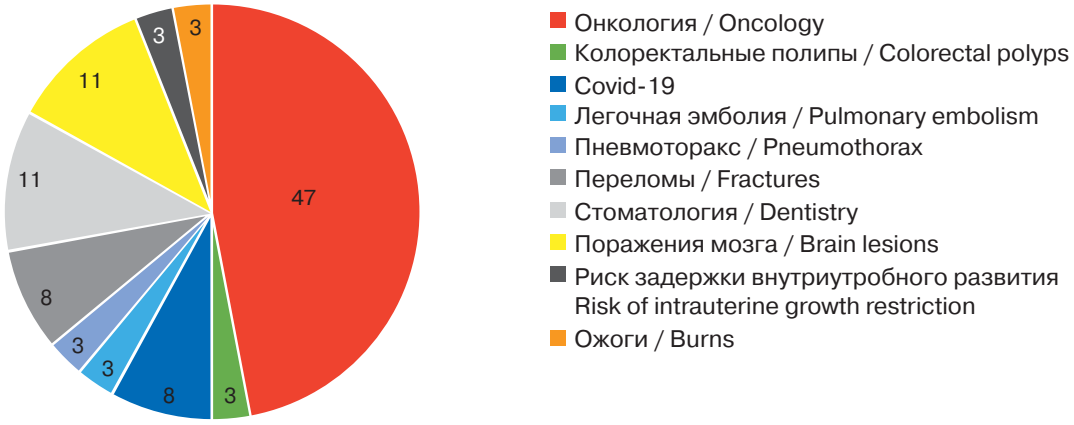


Рис. 1. Распределение (%) направлений исследований.
Fig. 1. Distribution (%) of research areas.

Рис. 2. Распределение (%) исследованных систем органов в метаанализах онкологической направленности.
Fig. 2. Distribution (%) of organ systems studied in oncology meta-analyses.

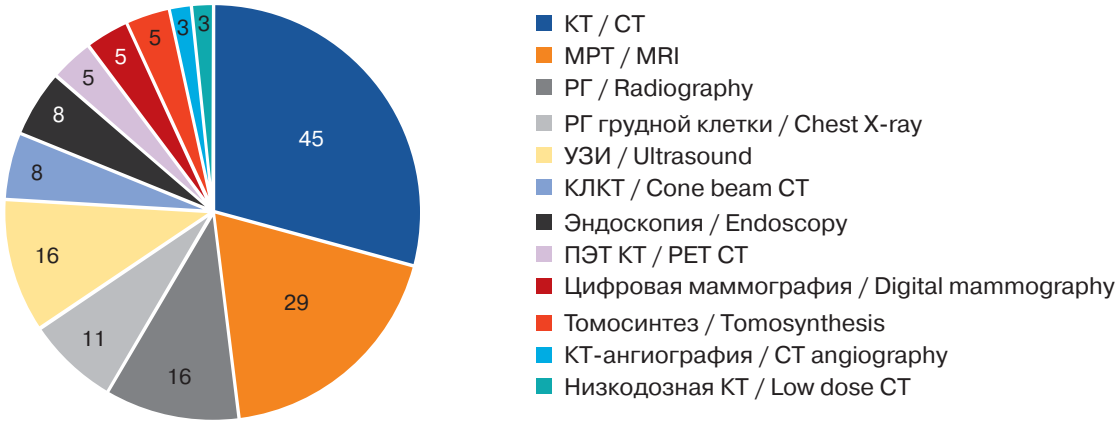


Рис. 3. Распределение (%) модальностей, использованных в метаанализах.
Fig. 3. Distribution (%) of modalities used in meta-analyses.

Методы эндоскопиии использовали в 3 (8%) метаанализах [25, 37, 38], в метаанализе [25] наряду с КТ и МРТ; в остальных двух метаанализах – как единственный метод.

В 2 (5%) обзорах нет данных по модальности изображений.

Оценка качества методологии

Число статей в метаанализах варьирует от 5 [36] до 63 [17], среднее значение 19,5, медиана 16,5, стандартное отклонение 12,7. Число пациентов в одном исследовании варьирует от 14 до 538 390 человек. Общее число пациентов, включенных в один метаанализ, варьирует от 443 [19] до 714 939 [13], среднее значение 128 762, медиана 17 429, стандартное отклонение 218 870. Из 9 (24%) метаанализов не удалось извлечь информацию об общем числе пациентов по следующим причинам. В двух метаанализах не указано общее число изученных пациентов, но указаны интервалы разброса числа пациентов: [17] – медиана 5288, интерквартильный интервал 516–30 805, интервал минимальных и максимальных значений 100–538 390; [11] – медиана 1169, интерквартильный интервал 425–2417, интервал минимальных и максимальных значений 65–21 456. В пяти метаанализах [28, 30, 35, 39, 40] приведено распределение числа пациентов в отдельных исследованиях по выборкам (например, обучающая, тестовая), но не ясно, перекрывались ли эти выборки или нет, неизвестно число пациентов в выборках, использованных для валидации. В обзоре [30] приведена информация по числу пациентов только для обучающих выборок, в обзоре [39] для большинства включенных исследований дано только соотношение долей числа пациентов между обучающей и тестовой выборками. В двух метаанализах, выполненных в области стоматологии [27, 34], не приведена информация по числу пациентов, а дана информация по числу использованных изображений: [34] – 102–2126 изображений, [27] – 1300–3900 изображений, либо по числу анатомических структур: [34] – 30–500 зубов. В двух метаанализах – [23] (опубликован в BMC Cancer с импакт-фактором 4,40) и [38] (опубликован в PlosOne с импакт-фактором 3,75) не приведены данные по объему выборок.

В 35 (92%) метаанализах из 38 приведены оценки риска систематической ошибки. В 3 метаанализах [13, 34, 37] эти оценки не приведены. В 13 метаанализах (37%) из 35 риск систематической ошибки оценен как высокий, как средний – в 4 (12%), как низкий – в 18 (51%) метаанализах.

Для оценки риска систематической ошибки инструмент QUADAS-2 (Приложение 2: https://medvis.vidar.ru/jour/article/view/1425 – дополнительные файлы) (https://www.bristol.ac.uk/population-health-sciences/projects/quadas/quadas-2/) [41] использовали в 28 (80%) метаанализах из 35. В одном из этих 28 метаанализов использовали модифицированный опросник QUADAS-AI [12]. Из 28 метаанализов риск систематической ошибки по QUADAS-2 оценен как низкий в 14 обзорах (50%), как средний – в 4 (14%), как высокий – в 10 (36%) обзорах. В 10 обзорах, где риск систематической ошибки оценен как высокий, риски анализа связаны с отбором пациентов (patient selection) – 8 метаанализов; индексным тестом (index test) – 4 метаанализа; референсным стандартом (reference standard) – 2 метаанализа; ходом и временем исследования (flow and timing) – 1 метаанализ. Вопросы применимости доказательств (applicability concerns) связаны с отбором пациентов – в двух метаанализах, индексным тестом – в одном метаанализе.

Еще в 2 (5%) метаанализах [11, 17] для оценки рисков систематической ошибки использовали инструмент PROBAST (Prediction Model Study Risk of Bias Assessment Tool). В обоих случаях риски систематической ошибки оценены как высокие из-за отсутствия как внешней, так и внутренней валидации на маленьких выборках [11, 17], неподходящих критериям включения и исключения [11, 17].

Другие инструменты оценки риска систематической ошибки использованы в четырех метаанализах: TRIPOD tool [40] – высокий риск систематической ошибки из-за отсутствия данных по валидации и применимости результатов сегментации ИИ в клинических исследованиях; JBI (Joanna Briggs Institute) critical appraisal tool and the CASP (Critical Appraisal Skills Programme) tool [33], Cochrane tool to assess the risk of bias [35], Newcastle-Ottawa Scale (NOS) [20] – низкий риск систематической ошибки. В обзоре [20] использовали два инструмента QUADAS-2 и NOS, оба инструмента показали низкий риск систематической ошибки.

Из 5 метаанализов, включавших и радиомические параметры [20, 21, 23, 24, 43], в одном [43] была проведена специализированная оценка качества радиомики RQS (https://www.radiomics.world/rqs2) [42]. В обзоре [43] оценили качество радиомики как низкое (4 балла из 36) из-за низкой детализации протоколов исследований, отсутствия валидации и закрытости данных.

Наличие либо отсутствие проспективных исследований отмечено в 17 (45%) метаанализах из 38. Из них в 7 (41%) метаанализах проспективные исследования отсутствуют. В остальных 10 метаанализах доля проспективных исследований варьи-

рует от 5 до 50% от общего числа включенных в обзоры исследовательских статей.

Наличие либо отсутствие исследований с использованием валидации результатов на внешних данных отмечено в 14 (37%) метаанализах из 38. Из них в 2 (14%) метаанализах исследования с внешней валидацией отсутствуют. В остальных 12 метаанализах доля исследований с валидацией результатов на внешних данных варьирует от 3 до 71% от общего числа включенных в метаанализы исследовательских статей.

Оценку предвзятости публикаций (publication bias risk) проводили в 22 (58%) метаанализах из 38 по стандартной методике теста Эггера (Egger's test) с помощью оценки асимметрии воронкообразного графика (funnel plot asymmetry). Из них риск предвзятости публикаций оценен как высокий в 6 (27%) метаанализах и как низкий – в 16 (73%) метаанализах.

Оценку неоднородности (heterogeneity) диагностических статистик проводили в 35 (92%) метаанализах из 38 с использованием I^2 индекса. Из них в 28 (80%) метаанализах уровень неоднородности высокий, в 4 (11%) метаанализах – низкий. В обзоре [23] обнаружили высокий уровень неоднородности для радиомических исследований и низкий – для исследований, использовавших глубокое машинное обучение. В обзоре [40] обнаружили высокий уровень неоднородности для исследований прямой кишки и низкий уровень – для исследований мочевого пузыря и CTV (clinical target volume). В обзоре [24] обнаружили высокий уровень неоднородности для радиомических данных и низкий уровень – для оценок рентгенологов.

Оценка качества методологии систематических обзоров по AMSTAR-2

Из 16 вопросов инструмента AMSTAR-2 к систематическим обзорам в области диагностической точности применимо 9 вопросов. Результаты представлены в Приложении 3: https://medvis.vidar.ru/jour/article/view/1425 – дополнительные файлы.

В 20 (53%) из 38 включенных систематических обзоров отмечено, что протокол разработан до проведения исследования и исследования проспективно зарегистрированы в системе PROSPERO (International Prospective Register of Systematic Reviews, https://www.crd.york.ac.uk/PROSPERO/) [44].

Требования к поиску литературы выполнены частично во всех включенных систематических обзорах. Во всех систематических обзорах проводили поиск работ более чем в двух базах данных, авторы подробно описывают поисковую стратегию и использованные ключевые слова и приводят критерии исключения, в том числе и языковые. Помимо публикаций в рецензируемых журналах, серую литературу использовали в 1 (3%) включенном систематическом обзоре [10]. Также, помимо поиска статей в базах данных, поиск ссылок в списках литературы отобранных публикаций, т.н. snowballing, использовали в 13 (34%) из 38 систематических обзоров [10, 14, 15, 19, 22, 24, 27, 31, 34, 38, 45, 46]. Авторы систематических обзоров контактировали с авторами исследований для уточнения информации в 2 (5%) из 38 случаев [18, 31].

Поиск и отбор литературы для систематического обзора осуществляли не менее двух авторов в 31 (82%) из 38 случаев, еще один автор выступал в качестве арбитра. Степень согласованности между авторами в отборе исследований не оценивали ни в одном систематическом обзоре.

Извлечение информации из отобранных статей осуществляли не менее двух авторов в 31 (82%) из 38 систематических обзоров, третий автор выступал качестве арбитра. Степень согласованности между авторами в извлечении информации не оценивали ни в одном обзоре.

Ни в одном из включенных систематических обзоров не было приведено списка исключенных исследований. Также ни в одном из включенных систематических обзоров не было приведено источников финансирований исследований, включенных в систематический обзор.

Исследовали причины неоднородности диагностических оценок (например, с использование метарегрессии) и обсуждали возможные причины этой неоднородности в 24 (63%) из 38 систематических обзоров. Еще в 4 (11%) из 38 систематических обзоров уровень неоднородности оценен как низкий.

Возможность риска предвзятости отобранных публикаций обсуждается в 25 (66%) из 38 систематических обзоров, из них в 22 метаанализах количественную оценку предвзятости публикаций проводили по стандартной методике теста Эггера (Egger's test) с помощью оценки асимметрии воронкообразного графика (funnel plot asymmetry).

Авторы 30 (79%) систематических обзоров из 38 сообщают о конфликте интересов или его отсутствии.

Диагностическая точность ИИ

В качестве основных диагностических параметров чувствительность, специфичность и площадь под ROC-кривой (ROC AUC) используются в 35 (92%) из 38 метаанализов (табл. 2). Совместно все три критерия приведены в 23 метаанализах.

Таблица 2. Диагностические параметры ИИ из метаанализов: чувствительность, специфичность, ROC AUC
Table 2. Diagnostic parameters of AI from meta-analyses: sensitivity, specificity, ROC AUC

Первый автор, год First author, year	Чувствительность ИИ, % Sensitivity of AI, %	Специфичность ИИ, % Specificity of AI, %	ROC AUC ИИ, % ROC AUC of AI, %
J.H. Yoon et al., 2023 [9]	80.6 (95% CI 74.3–85.7)	85.7 (95% CI 74.1–92.6)	87.0–90.0 (min–max)
S.E. Hickman et al., 2023 [10]	75.4 (95% CI 65.6–83.2)	90.6 (95% CI 82.9–95.0)	89.0 (95% CI 84.0–98.0)
L.T. Thong et al., 2023 [18]	94.6 (95 % CI 91.4–96.7)	93.6 (95 % CI 88.5–96.6)	
X. Zheng et al., 2022 [21]	66.1–83.8 (min–max)	59.8–77.4 (min–max)	79.0 (95% CI 77.0–82.0)– 83.0 (95% CI 78.0–88.0)
M. Liu et al., 2023 [55]	87.0 (95% CI 82.0–90.0)	87.0 (95% CI 82.0–91.0)	93.0 (95% CI 91.0–95.0)
T. Sugibayashi et al., 2023 [17]	84.0 (95% CI 79.0–89.0)	96.0 (95% CI 94.0–98.0)	97.0 (95% CI 96.0–98.0)
R.Y.L. Kuo et al., 2022 [11]	91.0 (95% CI 84.0–95.0)	91.0 (95% CI 81.0–95.0)	96.0 (95% CI 94.0– 98.0)
X. Zhang et al., 2022 [16]	90.0 (95% CI 87.0–92.0)	92.0 (95% CI 90.0–94.0)	97.0 (95% CI 95.0–98.0)
J.R. Lex et al., 2023 [13]	89.3 (std 8.5)	87.5% (std 9.9)	
P. Potipimpanon et al., 2022 [32]	86.0 (95% CI 81.0–91.0)	78.0 (95% CI 73.0–83.0)	89.0 (95% CI 86.0–92.0)
C.A. Campello et al., 2023 [30]	81.7 (95% CI 77.2–85.4)	84.8% (95% CI 76.0–90.8)	
F. Nabizadeh et al., 2023 [46]	92.0 (95%CI 90.0–95.0)	93.0 (95%CI 90.0–96.0)	93.0 (95%CI 89.0–96.0)
M. Din et al., 2023 [25]	91.2% (95% CI 82.2%–95.8)		93.6
M.D. Jørgensen et al., 2023 [56]	96.0 (95% CI 93.0–97.0)	97.0 (95% CI 90.0–99.0)	98.0 (95% CI 97.0–99.0)
H.Y. Kim et al., 2021 [43]	77.0 (95% CI 70.0–83.0)	74.0 (95% CI 64.0–82.0)	
A. Adamou et al., 2023 [45]			
S. Agarwal et al., 2023 [22]	90.0 (95% CI 85.0–94.0)	90.0 (95% CI 83.0–95.0)	80.0–99.1
S. Bedrikovetski et al., 2021a [23]			80.8 (95% CI 73.9–87.6)– 91.7 (95% CI 88.2–95.2)
S. Soffer et al., 2023 [36]	88.0 (95% CI 80.3–92.7)	86.0 (95% CI 75.6–92.4)	85.0–95.0
C. Yang et al., 2022 [40]			
L. Gao et al., 2021 [14]	96.0 (95% CI 93.0–100.0)	95.0 (95% CI 91.0–99.0)	86.0–100.0
R. Rescinito et al., 2023 [33]	84.0 (95% CI 80.0–88.00)	87.0 (95% CI 83.0–90.0)	
Q. Wang et al., 2022 [29]	87.0 (95% CI 85.0–89.0)	85.0 (95% CI 82.0–87.0)	94.0 (95% CI 91.0–96.0)
T.N. Poly et al., 2021 [15]	95.0 (95% CI 94.0–95.0)	96.0 (95% CI 96.0–97.0)	98.0
T.E. Komolafe et al., 2023 [28]	91.0 (95% CI 88.0–93.0)	92.0 (95% CI 88.0–94.0)	95.0 (95% CI 88.0–92.0)
F. Abesi et al., 2023 [34]			
F.F. Badr, F.M. Jadu, 2022 [35]	93.0 (95% CI 84.0–97.0)		
W. Xing et al., 2023 [20]			75.0 (95 % CI 67.0–84.0)
N. Menon et al., 2023 [19]	86.5 (95% CI 81.1–90.6) – 86.7 (95% CI 81.4–90.7)	76.1 (95% CI 69.9–81.4) – 87.1 (95% CI 78.0–92.8)	
B.G. Taib et al., 2023 [39]	90.8–97.9	84.4–97.6	
S. Sadr. et al., 2023 [27]	92.5 (95% CI 86.2–96.0)	85.2 (95% CI 81.0–88.5)	
M. Islam et al., 2022 [37]	93.8	91.7	96.0
K. Evangelista et al., 2022 [57]	84.0 (95% CI 58.0–100.0)	89.0 (95% CI 74.0–98.0)	92.0 (95% CI 72.0–100.0)
Y. Xue et al., 2022 [12]	88.0 (95% CI 85.0–90.0)	81.0 (95% CI 74.0–86.0)	92.0 (95% CI 89.0–94.0)
H.-L. Xu et al., 2022 [26]	88.0 (95%CI 85.0–90.0)	85.0 (95%CI 82.0–88.0)	93.0 (95% CI 91.0–95.0)
E.A. Dumitrescu et al., 2022 [31]	92.0 (95% CI 89.0–95.0)	90.0 (95% CI 83.0–94.0)	95.0
S. Bedrikovetski et al., 2021b [24]	72.0–100.0	56.6–76.2	79.8–89.5
Y. Xu et al., 2021 [38]	84.8 (95% CI 69.2–93.2)– 94.3 (95% CI 92.7–95.5)	89.4 (95% CI 63.1–97.7)– 96.5 (95% CI 94.6–97.7)	95.0 (95% CI 93.0–97.0)– 98.0 (95% CI 96.0–99.0)

Таблица 3. Диагностическая точность ИИ и врачей
Table 3. Diagnostic accuracy of AI and physicians

Первый автор, год First author, year	Область исследования Field of study	Модальность Modality	Диагностик Diagnostician	Чувствительность, ИИ/Врач Sensitivity, AI/Physician	Достоверность различий для чувствительности ИИ/врач Significance of differences for sensitivity AI/Physician p-level	Диагностик Diagnostician	Специфичность, ИИ/Врач Specificity, AI/Physician	Достоверность различий специфичности ИИ/врач Significance of differences for specificity AI/Physician p-level	ROC AUC ИИ/Врач ROC AUC AI/Physician	Достоверность различий ROC AUC для ИИ/Врач Significance of differences for ROC AUC AI/Physician p-level	Число врачей и их опыт Number of physicians and their experience
J.H. Yoon et al., 2023 [9]	Диагностика злокачественных новообразований молочной железы Diagnostics of malignant neoplasms of the breast	Цифровая маммография и томосинтез Digital mammography and tomosynthesis	ИИ / AI	80.6 (95% CI 74.3–85.7)	0.031	ИИ / AI	85.7 (95% CI 74.1–92.6)	0.221	87.0–90.0	0.152	Число врачей 2–24 (min-max), общее число врачей–143, опыт не указан Number of physicians 2–24 (min–max), total number–143, experience is not specified
			Врач / Physician	73.6 (95% CI 68.7–78.0)		Врач / Physician	89.6% (95% CI 82.7–93.9)		81.0–96.0		
S.E. Hickman et al., 2023 [10]	Диагностика злокачественных новообразований молочной железы Diagnostics of malignant neoplasms of the breast	Цифровая маммография Digital mammography	ИИ / AI	75.4 (95% CI 65.6–83.2)	0.70	ИИ / AI	90.6 (95% CI 82.9–95.0)	0.73	89.0 (95% CI: 84.0–98.0)		Число врачей 4–101 (min–max), общее число врачей – 211, опыт от 1 года до 44 лет Number of physicians 4–101 (min–max), total number – 211, experience from 1 year to 44 yeas
			Врач / Physician	73.0 (95% CI 60.7–82.6)		Врач / Physician	88.6 (95% CI 72.4–95.8)		85.0 (95% CI: 78.0–97.0)		
T. Sugibayashi et al., 2023 [17]	Диагностика пневмоторакса Diagnostics of pneumothorax	Рентгенография грудной клетки (59), КТ (4) Chest X-ray (59), CT (4)	ИИ / AI	84.0 (95% CI 79.0–89.0)		ИИ / AI	96.0 (95% CI 94.0–98.0)		97.0 (95% CI: 96.0–98.0)		
				85.0 (95% CI 73.0–92.0)		Врач / Physician	98.0 (95% CI 95.0–99.0)				
R.Y.L. Kuo et al., 2022 [11]	Диагностика переломов Diagnostics of fractures	КТ (2), рентгенография (30) CT (2), radiography (30)	ИИ / AI	91.0 (95% CI 84.0–95.0)		ИИ / AI	91.0 (95% CI 81.0–95.0)		96.0 (95% CI: 94.0–98.0)		Число врачей-клиницистов 3–58 (min–max), опыт не указан Number of clinicians 3–58 (min-max), experience is not specified
			Врач / Physician	94.0 (95% CI 90.0–96).0		Врач / Physician	94.0 (95% CI 91.0–95.0)		98.0 (95% CI: 96.0–99.0)		
X. Zhang et al., 2022 [16]	Диагностика переломов Diagnostics of fractures	КТ (5), рентгенография (34) CT (5), radiography (34)	ИИ / AI	90.0 (95% CI 87.0–92.0)		ИИ / AI	92.0 (95% CI 90.0–94.0)		97.0 (95% CI: 95.0–98.0)		
			Врач / Physician	90.0 (95% CI 85.0–93.0)		Врач / Physician	95.0 (95% CI 93.0–96.0)		97.0 (95% CI: 96.0–99.0)		
P. Potipimpanon et al., 2022 [32]	Диагностика узлов щитовидной железы Diagnostics of thyroid nodules	УЗИ Ultrasonography (US)	ИИ / AI	86.0 (95% CI 81.0–91.0)		ИИ / AI	78.0 (95% CI 73.0–83.0)		89 (95% CI: 86.0–92.0)		Число рентгенологов 1–10, общее число радиологов – 54, опытных радиологов 0–3 в разных исследованиях, общее число опытных радиологов–30, в трех статьях опыт радиологов не указан Number of radiologists 1–10, total number–54, experienced radiologists 0–3, total number of experienced radiologists–30, experience is not specified in three articles
			Врач / Physician	85.0 (95% CI 80.0–89.0)		Врач / Physician	82.0 (95% CI 77.0–86.0)		91.0 (95% CI: 88.0–93.0)		
C.A. Campello et al., 2023 [30]	Диагностика злокачественных новообразований печени Diagnostics of malignant liver tumors	УЗИ, УЗИ с контрастным усилением US, Contrast-enhanced US	ИИ (натив) / AI (native)	81.7 (95% CI 77.2–85.4)		ИИ (натив) AI (native)	84.8 (95% CI 76.0–90.8)				15 рентгенологов, из них старших рентгенологов (опыт 15 – 25 лет) более 4, младших рентгенологов (опыт 5 – 15 лет) более 3. Данные в метаанализе приведены по трем исследованиям 15 radiologists, senior radiologists from them (experience 15 – 25 years) more than 4, junior radiologists (experience 5 – 15 years) more than 3. Data in the meta-analysis are based on three studies
			Врач (контраст) / Physician (contrast)	87.1% (95% CI 81.8–91.0)		Врач (контраст) Physician (contrast)	87.0 (95% CI 83.1–90.1)				
T.N. Poly et al., 2021 [15]	Диагностика COVID-19 Diagnostics of COVID-19	КТ (8), рентгенография (8) CT (8), RG (8)	ИИ / AI	95.0 (95% CI 94.0–95.0)		ИИ / AI	96.0 (95% CI 96.0–97.0)		98.0		
			Врач / Physician	95% CI 75.0–89.0		Врач / Physician	95% CI 83.0–90.0				
S. Bedrikovetski et al., 2021b [24]	Диагностика метастазов в лимфатические узлы у больных со злокачественными новообразованиями брюшной полости Diagnostics of metastases to the lymph nodes in patients with malignant neoplasms of the abdominal cavity	КТ (10), МРТ (10) CT (10), MRI (10)	ИИ / AI	72.0–100.0		ИИ / AI	56.6–76.2		79.8–89.5		
				43.2–77.9		Врач / Physician	55.9–100		63.3–77.4		
Y. Xu et al., 2021 [38]	Детекция колоректальных полипов Detection of colorectal polyps	Эндоскопия Endoscopy	ИИ, детекция / AI, detection	84.8 (95% CI 69.2–93.2)		ИИ, детекция AI, detection	96.5 (95% CI 94.6–97.7)		98.0 (95% CI: 96.0–99.0)		
			ИИ, классификация / AI, classification	94.3 (95% CI 92.7–95.5)		ИИ, классификация AI, classification	89.4 (95% CI 63.1–97.7)		95.0 (95% CI: 93.0–97.0)		
			Врач–эксперт, классификация / Expert, classification	94.4 (95% CI 89.2–97.2)		Врач–эксперт, классификация Expert, classification	84.8 (95% CI 73.2–91.9)		96.0 (95% CI: 94.0–98.0)		
			Врач–неэксперт, объединенное Non expert, united	85.9 (95% CI 76.9–91.8)		Врач–неэксперт, объединенное Non-expert, united	81.1 (95% CI 71.8–87.8)		90.0 (95% CI: 87.0–0.93)		

Был дополнительно рассчитан интервал обобщенных значений чувствительности, специфичности и ROC AUC между метаанализами. Использовали только данные, где были приведены средние значения по метаанализу. В случае, если диагностическая точность была дана по подгруппам, брали ее среднее значение.

Среднее значение (mean) чувствительности ИИ составляет 89%, медиана (median) – 88,7%, полученные на 33 (n) метаанализах с минимальным и максимальным значением 75,4 и 97,9% соответственно, стандартное отклонение (std) 5,3. Для специфичности ИИ обобщенные параметры следующие: mean = 88,8%, median = 89,5%, n = 31, min–max 74–97,6%, std = 5,6%. Для ROC AUC ИИ обобщенные параметры следующие: mean = 92%, median =93,8%, n = 26, min–max 75–98%, std = 5,6%.

Диагностическая точность ИИ по сравнению с врачами-рентгенологами

В табл. 3 приведены значения диагностических показателей для ИИ и врачей.

Значения диагностической точности совместно для ИИ и для врачей найдены в 10 (26%) метаанализах из 38. В одном метаанализе [30] диагностическую значимость ИИ оценивали по нативным изображениям, а для врачей – по изображениям с контрастом, поэтому этот метаанализ рассмотрен отдельно.

Из 10 метаанализов, где приведены совместно значения диагностической точности для ИИ и врачей, 5 (50%) – онкологической направленности: диагностика злокачественных новообразований молочной железы [10], диагностика узлов щитовидной железы [32], диагностика злокачественных новообразований печени [30], диагностика метастазов в лимфатические узлы у больных со злокачественными новообразованиями брюшной полости [24]. 2 (20%) метаанализа направлены на диагностику переломов [11, 16], 1 (10%) – диагностику пневмоторакса [17], 1 (10%) – диагностику COVID-19 [15], 1 (10%) – детекцию колоректальных полипов [38].

Для диагностики злокачественных новообразований молочной железы использовали цифровую маммографию [9, 10] и томосинтез [9]. УЗИ проводилось при диагностике узлов щитовидной железы [32] и злокачественных новообразований печени [30]. В исследованиях, где изучали диагностику метастазов в лимфатические узлы у больных со злокачественными новообразованиями брюшной полости [24], использовали КТ либо МРТ. Для диагностики переломов [11, 16] применяли КТ и рентгенографию, для диагностики пневмоторакса [17] – рентгенографию грудной клетки и КТ, для диагностики COVID-19 [15] – КТ и рентгенографию, для детекции колоректальных полипов [38] – эндоскопию.

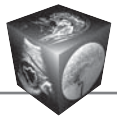
Интервалы для диагностических параметров ИИ и врачей посчитаны нами с включением метаанализов, в которых были приведены обобщенные значения интересующих диагностических параметров. Средние значения, медиана, среднеквадратичное отклонение, минимальное и максимальное значение чувствительности для ИИ и для врачей (число включенных метаанализов 7 [9, 10, 11, 16, 17, 32, 38]) следующие: 85,2; 86,0; 5,7; 75,4–91,0 и 84,4; 85,0; 8,2; 73,0–94,0% соответственно. Те же показатели для ИИ и для врачей (число включенных метаанализов 7 [9, 10, 11, 16, 17, 32, 38]) следующие: 89,5; 91,0; 5,9; 78,0–96,0 и 90,0; 94,0; 6,1; 82,0–98,0% соответственно. Те же показатели для ИИ и для врачей (число включенных метаанализов 5 [10, 11, 16, 32, 38]) следующие: 93,5; 92,5; 4,1; 89,0–97,0 и 92,8%; 94,0 5,2; 85,0–98,0% соответственно.

Из 10 метаанализов, где исследовали диагностическую точность и ИИ, и врачей, в 5 данные о числе врачей и их опыте не приводятся. В 3 метаанализах [10, 15, 32] содержатся данные по числу врачей и их опыту, в 2 метаанализах [9, 11] – только данные по числу врачей. Подробная информация приведена в табл. 3.

В 3 метаанализах приведены значения уровня значимости при сравнении диагностических параметров ИИ и врачей. В обзоре [9], рассматривающем диагностику злокачественных новообразований молочной железы с использованием цифровой маммографии и томосинтеза, показано, что чувствительность ИИ (80,6%) достоверно выше ($p = 0,031$), чем врачей (73,6%); специфичность ИИ (85,7%) и врачей (89,6%) не различается достоверно ($p = 0,221$); ROC AUC также не различается достоверно ($p = 0,152$) между ИИ (87–90%) и врачами (81–96%). Общее число врачей в метаанализе [9] – 143, варьирует между разными исследованиями от 2 до 24, опыт врачей не указан.

В обзоре [10], рассматривающем диагностику злокачественных новообразований молочной железы с использованием цифровой маммографии, показано, что чувствительность ИИ (75,4%) и врачей (73%) не различается достоверно ($p = 0,7$); специфичность ИИ (90,6%) и врачей (88,6%) также не различается достоверно ($p = 0,73$). Общее число врачей в метаанализе [10] – 211, варьирует между разными исследованиями от 4 до 101, опыт врачей варьирует от 1 года до 44 лет.

В обзоре [38], рассматривающем детекцию колоректальных полипов с использованием эндоскопии, тест на значимость различий между ИИ



и врачами приводится для параметра DOR (Diagnostic odds ratio). Результаты по DOR для ИИ сравнивали с врачами-экспертами и с врачами-не экспертами. DOR не различается достоверно (p = 0,9654) между ИИ и врачами-экспертами, DOR достоверно (p = 0,0342) лучше у ИИ и по сравнению с врачами-не экспертами. Число врачей и их опыт не указаны.

Еще в 4 метаанализах, рассматривающих диагностику пневмоторакса [17] с использованием рентгенографии грудной клетки и КТ, диагностику переломов [11, 16] с использованием рентгенографии и КТ, диагностику узлов щитовидной железы с использованием УЗИ [32], авторы отмечают, что не обнаружили разницы в диагностических параметрах между ИИ и врачами, при этом приводятся значения для чувствительности, специфичности и ROC AUC, но не приводятся значения p-value. В метаанализах [16, 17] данные по числу врачей и их опыту не приводятся. В метаанализе [11] число врачей-клиницистов варьировало от 3 до 58 между разными исследованиями, данные по опыту не приводятся. В метаанализе [32] общее число рентгенологов составило 54, из них число опытных рентгенологов – 30, число рентгенологов в разных исследованиях варьирует от 1 до 10, при этом число опытных рентгенологов – от 0 до 3, определения опытности рентгенологов в годах не приводится, в 3 (10%) исследованиях из 29, использованных в метаанализе, не указан опыт рентгенологов. Метаанализ [30], рассматривающий диагностику злокачественных новообразований печени с использованием УЗИ, отличается тем, что врачи анализировали изображения, полученные с помощью контрастного усиления, а ИИ тестировал нативные изображения. И хотя уровни значимости при сравнении диагностических интервалов ИИ и врачей не приводятся, авторы метаанализа отмечают высокие диагностические показатели ИИ. Так, средние значения чувствительности ИИ и врачей составляют 81,7 и 87,1% соответственно, средние значения специфичности ИИ и врачей – 84,8 и 87,0% соответственно. Данные по числу врачей, с которыми проводилось сравнение, и их опыту в метаанализе не приведены.

Результаты внедрения

Доказанные результаты внедрения ИИ приведены в одном (3%) метаанализе [22] из 38. Авторы обнаружили 3 (19%) статьи [47–49] из 16 проанализированных ими, где приводятся доказанные результаты внедрения ИИ в клиническую практику. Все 3 исследования выполнены в области нейровизуализации с использованием КТ. 2 исследования [47, 48] поместили модель ИИ в начало клинического пути перед интерпретацией рентгенологом (предварительная сортировка). В исследовании [47] ИИ показал сокращение среднего времени на отчет для несрочных обследований, которые ИИ отметил, как подлежащие пересмотру, с 512 до 19 мин. В исследовании [48] ИИ продемонстрировал достоверное снижение среднего времени для отчета по отмеченным случаям для амбулаторных больных (с 674 до 70 мин, p < 0,001), стационарных больных (с 390 до 352 мин, p = 0,002), но не для экстренных случаев (p = 0,37). Вред и отсрочка по времени из-за ложноотрицательных результатов при этом не оценивались, хотя число ложноотрицательных результатов составляло 26 (7,5%) из 347 в исследовании [47] и 205 (11,6%) из 1760 в исследовании [48].

В 2 исследованиях [47, 49] ИИ применялся в качестве второго читателя после интерпретации рентгенолога, рассматривали расхождения между ИИ и рентгенологами. ИИ выявил 1,2 и 0,03% соответственно ложноотрицательных результата рентгенологов.

Обсуждение

Число исследований, оценивающих применимость ИИ в диагностической практике, нарастает экспоненциально. Нами проведен систематический анализ области применимости ИИ в лучевой диагностике, дана оценка основных проблемных точек в этой области, которые могут препятствовать внедрению в практику этого современного инструмента.

Область применения ИИ в лучевой диагностике крайне широка. Половина проанализированных метаанализов выполнена в области диагностики злокачественных новообразований. Это одна из самых распространенных задач для ИИ в лучевой диагностике.

Часто (более 50%) в метаанализ включают исследования разных модальностей, что повышает неоднородность данных и затрудняет их обобщение. Лидирующим методом в лучевой диагностике с применением ИИ является КТ, за ней следует МРТ, далее – рентгенография и УЗИ.

Оценка качества исходных публикаций

Ключевым для внедрения ИИ в диагностическую практику являются данные и методика, с помощью которых диагностическая точность ИИ тестируется.

Стоит отметить большой разброс в объеме выборок (числе пациентов) между исходными публикациями, составляющий в некоторых случаях 3–4 порядка. В части исследований выборка мала

и может составлять менее двух десятков пациентов. Наблюдается недостаточное число проспективных исследований и исследований с внешней валидацией данных, что отмечается в большинстве метаанализов.

Для оценки риска систематической ошибки в подавляющем числе метаанализов использовали QUADAS-2, в настоящее время это один из основных инструментов оценки качества работ. В половине случаев исследования оценены как методологически приемлемые. В остальных случаях риск систематической ошибки оценен как высокий или средний. Основных причин высоких рисков систематической ошибки и снижения качества исследований несколько. Значительная часть рисков обусловлена использованием несбалансированных по объему и составу выборок. Например, включаются маленькие по объему выборки, а также выборки, где доля положительных случаев значительно выше доли отрицательных случаев. Необоснованность критериев исключения также является одним из важных факторов снижения качества исследований. Следующими наиболее частыми причинами повышения рисков являются низкое число либо отсутствие проспективных исследований, преобладание случай–контроль дизайна, отсутствие внешней валидации результатов, использование в качестве обучающей, тестовой и валидирующей перекрывающихся выборок. Зачастую недостаточно детальное описание методики исследования приводит к снижению ценности его результатов. Например, в некоторых исследованиях не ясны характер и способ заслепления, число и опыт врачей, осуществляющих контрольное чтение, время между индексным и референсным тестами. В целом наблюдается низкий уровень стандартизаций используемых в разных исследованиях методик. В частности, неоднородность подходов и методик дает высокую неоднородность диагностических оценок даже в одной области исследований.

В большинстве метаанализов риск предвзятости публикаций оценен как низкий. Тем не менее отсутствие публикаций с отрицательными результатами дает основание предполагать наличие этого риска.

Оценка качества включенных систематических обзоров

В первую очередь стоит отметить, что в настоящее время отсутствуют специализированные инструменты для оценки качества систематических обзоров в области диагностической точности ИИ. Так, из 16 вопросов AMSTAR-2 в этой области применимо только 9. Назрела необходимость разработки такого специализированного инструмента.

Результаты оценки качества методологии систематических обзоров показали, что она нуждается в улучшении. Так, только в половине рассмотренных систематических обзоров зарегистрирован протокол проведения своего исследования проспективно. Большинство обзоров не используют так называемую серую литературу, например опубликованные отчеты исследований, что может приводить к предвзятым оценкам диагностической точности и их завышению. В систематических обзорах часто не хватает количественных оценок степени согласованности результатов авторов, параллельно проводивших отбор литературы и извлечение информации, данных об исключенных статьях. В систематических обзорах не приводят информацию об источниках финансирования анализируемых в них исследований, что важно в медицинских работах, так как может приводить к предвзятости полученных результатов.

Диагностическая точность ИИ и внедрение в практику

В подавляющем большинстве (36 из 38) метаанализов отмечается высокая диагностическая точность ИИ. Высока доля истинно положительных результатов (чувствительность), доля истинно отрицательных результатов (специфичность), а также существенны значения ROC AUC, отражающей соотношение между истинно позитивными исходами и ложнопозитивными исходами, а также некоторые другие обобщенные диагностические критерии (например, DOR). Высокие диагностические показатели ИИ обнаружены в области диагностики злокачественных новообразований, остеопороза, переломов, прогнозов по беременности, диагностике COVID-19 в сравнении с другими видами пневмонии, автоматической сегментации изображений, прогноза исходов терапии при онкологических заболеваниях, а также в области нейровизуализации.

Диагностическая точность ИИ не уступает диагностической точности врачей, а в некоторых случаях может превосходить ее. В частности, обнаружено [38], что диагностическая точность ИИ сравнима с показателями врачей экспертного уровня, превосходя показатели обычных врачей. Тем не менее число работ, проводящих прямые количественные сравнения между ИИ и врачами, попрежнему мало. Кроме того, далеко не во всех этих работах подробно указаны число и опыт врачей, что существенно снижает возможность интерпретации и использования полученных результатов.

Во всех метаанализах отмечается высокий потенциал для внедрения ИИ в практику. Однако существует целый ряд методологических сложностей в этой области. В 2 метаанализах отмечена невысокая диагностическая точность ИИ, требующая повышения показателей, это работы, посвященные стадированию злокачественных новообразований легкого [21] и классификации ответов на терапию при метастазах головного мозга [43]. В обоих метаанализах включены радиомические работы, при этом качество радиомики не было оценено.

В метаанализах, где показаны высокие диагностические показатели и перспективность ИИ, отмечается ряд задач, требующих решения. Во-первых, необходимость увеличения объема данных [50, 51] и создания открытых пространств для тестирования результатов ИИ. Далее, необходимость стандартизации методик [52, 53], так как высокая неоднородность данных в метаанализах делает выводы менее убедительными. Также отмечается необходимость внешней валидации для генерализации результатов. Валидация на внутренних данных склонна переоценивать значения ROC AUC, что ограничивает интеграцию моделей ИИ в клинических условиях. Необходимо приводить подробные данные о числе и опыте врачей, с которыми сравнивали диагностическую точность ИИ. Кроме того, отмечается отсутствие комплексных исследований последствий клинического внедрения алгоритмов ИИ.

Из всех проанализированных нами метаанализов только в одном приводится анализ результатов реального внедрения ИИ в диагностическую практику. Этот анализ основан на трех исследованиях. Результаты использования ИИ на практике неоднозначны. С одной стороны, ИИ дал выигрыш по времени при несрочных обследованиях, но не дал выигрыша при срочных обследованиях. В качестве инструмента верификации первого прочтения ИИ эффективно выявлял ложноотрицательные результаты рентгенологов, но при этом эффективность выявления ложноположительных результатов оказалась неудовлетворительной. На данном этапе требуется больше исследований по внедрению ИИ в диагностическую практику для выявления преимуществ и недостатков метода.

Можно заключить, что в настоящий момент результаты исследований показывают высокую диагностическую точность ИИ, но важно, что эти результаты зачастую получены в исследованиях с некорректным дизайном, способом проведения исследований и их отчетности, что может привести к систематической ошибке и переоценке эффективности этих алгоритмов [54].

Заключение

Существуют различные области потенциального применения ИИ в лучевой диагностике. Наибольшее число исследований выполнено в области диагностики злокачественных новообразований – 47% из проанализированных нами метаанализов. Из них наиболее часто встречаются исследования, посвященные диагностике злокачественных новообразований пищеварительной (28%), дыхательной (22%) и репродуктивной (22%) систем. Менее часто встречаются исследования в области диагностики поражений мозга (11%) и в стоматологии (11%).

Лидирующей модальностью в исследованиях является КТ (45% метаанализов), далее следует МРТ (29%), а за ней – рентгенография (16%) и УЗИ (16%).

Существует целый ряд методологических проблем, приводящих к снижению ценности результатов, полученных в области оценки потенциала применения ИИ в лучевой диагностике. Риск систематической ошибки оценен как низкий только в половине (51%) метаанализов, как средний – в 12%, как высокий – в 37%. Снижение качества исследований связано с несбалансированностью выборок по объему и составу, низкой долей проспективных исследований и исследований с внешней валидацией данных. Недостаточно подробное описание использованных методик и отсутствие данных в открытом доступе затрудняют валидацию результатов и стандартизацию методик.

Качество методологии самих систематических обзоров в области диагностической точности ИИ также нуждается в повышении, а именно методы планирования систематического обзора, методы поиска и отбора литературы, методы отчетности. Назрела необходимость разработки специализированного инструмента оценки качества систематических обзоров в области диагностической точности ИИ.

По данным современных исследований алгоритмы ИИ не уступают врачам по показателям диагностической точности, а в некоторых случаях могут превышать ее. Средние значения чувствительности, специфичности и ROC AUC: 85,2% (median = 86,0; std = 5,7%; min–max = 75,4–91,0%), 89,5% (91,0%; 5,9%; 78,0–96,0%), 93,5% (92,5%, 4,1%; 89,0–97,0%). Для врачей-рентгенологов указанные показатели составляют: 84,4% (85,0%; 8,2%; 73,0–94,0%), 90,0% (94,0%; 6,1%; 82,0–98,0%), 92,8% (94,0% 5,2%; 85,0–98,0%) соответственно.

ИИ является перспективным методом для внедрения в клиническую практику. Однако для повышения достоверности результатов необходимы: стандартизация методик исследования ИИ в диаг-

ностике и их детальное описание, использование сбалансированных по объему и составу выборок, валидация результатов на внешних данных. Несмотря на высокие диагностические показатели ИИ, число доказанных результатов внедрения метода в клиническую практику невелико – только один (3%) метаанализ рассматривает доказанные результаты внедрения ИИ, но приведенные результаты при этом противоречивы. На практике ИИ давал выигрыш по времени для несрочных больных, но не давал выигрыша по времени для экстренных больных. Доля ложноотрицательных результатов ИИ при этом была достаточно высока (7,5–11,6%). Остается также неясным, какой проигрыш по времени давали на практике ложноотрицательные результаты ИИ.

Финансирование. Работа выполнена при финансовой поддержке НИР “Научное обоснование методов лучевой диагностики опухолевых заболеваний с использованием радиомического анализа” (№ ЕГИСУ: № 123031500005-2) в соответствии с Приказом от 21.12.2022 № 1196 “Об утверждении государственных заданий, финансовое обеспечение которых осуществляется за счет средств бюджета города Москвы государственным бюджетным (автономным) учреждениям, подведомственным Департаменту здравоохранения города Москвы, на 2023 год и плановый период 2024 и 2025 годов” Департамента здравоохранения города Москвы.

Вклад авторов

Авторы подтверждают соответствие своего авторства международным критериям ICMJE. Все авторы внесли существенный вклад в проведение исследования и подготовку статьи, прочли и одобрили финальную версию перед публикацией. Наибольший вклад распределен следующим образом:

Васильев Ю.А., Владзимирский А.В., Омелянская О.В. – разработка концепции исследования, согласование финальной версии рукописи.

Решетников Р.В., Блохин И.А., Коденко М.Р., Нанова О.Г. – обзор литературы, анализ данных, написание текста статьи.

Authors’ participation

Authors confirm that their authorship meets the international criteria of the ICMJE. All authors made significant contributions to the study and preparation of the article, read and approved the final version before publication. The largest contribution is distributed as follows: Vasiliev Yu.A., Vladzymyrskyy A.V., Omelyanskaya O.V. – development of the research concept, approvement of the final version of the manuscript. Reshetnikov R.V., Blokhin I.A., Kodenko M.R., Nanova O.G. – literature review, data analysis, writing the text of the manuscript.

Список литературы [References]

1. Hosny A., Parmar C., Quackenbush J. et al. Artificial intelligence in radiology. *Nat. Rev. Cancer.* 2018; 18 (8): 500–510. http://doi.org/10.1038/s41568-018-0016-5
2. Sogani J., Allen B. Jr., Dreyer K., McGinty G. Artificial intelligence in radiology: the ecosystem essential to improving patient care. *Clin. Imaging*. 2019; 59 (1): A3–A6. https://doi.org/10.1016/j.clinimag.2019.08.001
3. Keane P.A., Topol E.J. With an eye to AI and autonomous diagnosis. *NPJ Digital Medicine.* 2018; 1: 40. https://doi.org/10.1038/s41746-018-0048-y
4. Владзимирский А.В., Гусев А.В., Шарова Д.Е., Шулькин И.М., Попов А.А., Балашов М.К., Омелянская О.В., Васильев Ю.А. Методика оценки уровня зрелости информационной системы для здравоохранения. *Врач и информационные технологии*. 2022; 3: 68–86. https://doi.org/10.25881/18110193_2022_3_68
Vladzimirskyy A.V., Gusev A.V., Sharova D.E. et al. Health information system maturity assessment methodology. *Vrach i informatsionnye tekhnologii.* 2022; 3: 68–86. https://doi.org/10.25881/18110193_2022_3_68 (In Russian)
5. Шарова Д.Е., Гарбук С.В., Васильев Ю.А. Системы искусственного интеллекта в клинической медицине. Первая в мире серия национальных стандартов. *Стандарты и качество.* 2023; 1: 46–51. https://doi.org/10.35400/0038-9692-2023-1-304-22
Sharova D.E., Garbuk S.V., Vasilev Y.A. Artificial intelligence systems in clinical medicine: the world’s first series of national standards. *Standards and Quality.* 2023; 1: 46–51. https://doi.org/10.35400/0038-9692-2023-1-304-22 (In Russian)
6. Kelly S., Kaye S.-A., Oviedo-Trespalacios O. What factors contribute to the acceptance of artificial intelligence? A systematic review. *Telematics and Informatics*. 2023; 77: 101925. https://doi.org/10.1016/j.tele.2022.101925
7. Page M.J., McKenzie J.E., Bossuyt P.M. et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021; 372: n71. https://doi.org/10.1136/bmj.n71
8. Shea B.J., Reeves B.C., Wells G. et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*. 2017; 358: j4008. https://doi.org/10.1136/bmj.j4008
9. Yoon J.H., Strand F., Baltzer P.A.T. et al. Standalone AI for Breast Cancer Detection at Screening Digital Mammography and Digital Breast Tomosynthesis: A Systematic Review and Meta-Analysis. *Radiology*. 2023; 307 (5): e222639. https://doi.org/10.1148/radiol.222639
10. Hickman S.E., Woitek R., Vi Le E.P. et al. Machine Learning for Workflow Applications in Screening Mammography: Systematic Review and Meta-Analysis. *Radiology*. 2022; 302: 88–104. https://doi.org/10.1148/radiol.2021210391
11. Kuo R.Y.L., Harrison C., Curran T.-A. et al. Artificial Intelligence in Fracture Detection: A Systematic Review and Meta-Analysis. *Radiology.* 2022. 304 (1): 50–62. https://doi.org/10.1148/radiol.211785
12. Xue Y., Zhou Y., Wang T. et al. Accuracy of Ultrasound Diagnosis of Thyroid Nodules Based on Artificial Intelligence-Assisted Diagnostic Technology: A Systematic Review and Meta-Analysis. *Int. J. Endocrinol.* 2022: 9492056. https://doi.org/10.1155/2022/9492056
13. Lex J.R., Di Michele J., Koucheki R. et al. Artificial Intelligence for Hip Fracture Detection and Outcome

Prediction A Systematic Review and Meta-analysis. *JAMA Network Open*. 2023; 6 (3): e233391. https://doi.org/10.1001/jamanetworkopen.2023.3391
14. Gao L., Jiao T., Feng Q., Wang W. Application of artificial intelligence in diagnosis of osteoporosis using medical images: a systematic review and meta-analysis. *Osteoporosis International*. 2021; 32: 1279–1286. https://doi.org/10.1007/s00198-021-05887-6
15. Poly T.N., Islam M., Li Y.-C. J. et al. Application of Artificial Intelligence for Screening COVID-19 Patients Using Digital Images: Meta-analysis. *JMIR Med. Inform*. 2021; 9 (4): e21394. https://medinform.jmir.org/2021/4/e21394
16. Zhang X., Yang Y., Shen Y.-W. et al. Diagnostic accuracy and potential covariates of artificial intelligence for diagnosing orthopedic fractures: a systematic literature review and meta-analysis. *Eur. Radiol*. 2022; 32 (10): 7196–7216. https://doi.org/10.1007/s00330-022-08956-4
17. Sugibayashi T., Walston S.L., Matsumoto T. et al. Deep learning for pneumothorax diagnosis: a systematic review and meta-analysis. *Eur. Respirat. Rev.* 2023; 32: 220259. https://doi.org/10.1183/16000617.0259-2022
18. Thong L.T., Chou H.S., Chew H.S.J., Lau Y. Diagnostic test accuracy of artificial intelligence-based imaging for lung cancer screening: A systematic review and meta-analysis. *Lung Cancer.* 2023; 176: 4–13. https://doi.org/10.1016/j.lungcan.2022.12.002
19. Menon N., Guidozzi N., Chidambaram S., Markar S.R. Performance of radiomics-based artificial intelligence systems in the diagnosis and prediction of treatment response and survival in esophageal cancer: asystematic review and meta-analysis of diagnostic accuracy. *Dis. Esophagus*. 2023; 36 (6): doad034. https://doi.org/10.1093/dote/doad034
20. Xing W., Gao W., Lv X. et al. Artificial intelligence predicts lung cancer radiotherapy response: A meta-analysis. *Artif. Intell. Med.* 2023; 142: 102585. https://doi.org/10.1016/j.artmed.2023.102585
21. Zheng X., He B., Hu Y. et al. Diagnostic Accuracy of Deep Learning and Radiomics in Lung Cancer Staging: A Systematic Review and Meta-Analysis. *Front. Public Health.* 2022; 10: 938113. https://doi.org/10.3389/fpubh.2022.938113
22. Agarwal S., Wood D., Grzeda M. et al. Systematic Review of Artificial Intelligence for Abnormality Detection in High-volume Neuroimaging and Subgroup Meta-analysis for Intracranial Hemorrhage Detection. *Clin. Neuroradiol.* 2023; 33: 943–956. https://doi.org/10.1007/s00062-023-01291-1
23. Bedrikovetski S., Dudi-Venkata N.N., Kroon H.M. et al. Artificial intelligence for pre-operative lymph node staging in colorectal cancer: a systematic review and meta-analysis. *BMC Cancer*. 2021a; 21. https://doi.org/10.1186/s12885-021-08773-w
24. Bedrikovetski S., Dudi-Venkata N.N., Maicas G. et al. Artificial intelligence for the diagnosis of lymph node metastases in patients with abdominopelvic malignancy: A systematic review and meta-analysis. *Artif. Intell. Med.* 2021b; 113: 102022. https://doi.org/10.1016/j.artmed.2021.102022
25. Din M., Agarwal S., Grzeda M. et al. Detection of cerebral aneurysms using artificial intelligence: a systematic review and meta-analysis. *J. NeuroIntervent. Surg*. 2023; 15: 262–271. https://doi.org/10.1136/jnis-2022-019456
26. Xu H.-L., Gong T.-T., Liu F.-H. et al. Artificial intelligence performance in image-based ovarian cancer identification: A systematic review and meta-analysis. *eClinicalMedicine.* 2022; 53: 101662. https://doi.org/10.1016/j.eclinm.2022.101662
27. Sadr S., Mohammad-Rahimi H., Motamedian S.R. et al. Deep Learning for Detection of Periapical Radiolucent Lesions: A Systematic Review and Meta-analysis of Diagnostic Test Accuracy. *J. Endodont*. 2023; 49 (3): 248–261.e3. https://doi.org/10.1016/j.joen.2022.12.007
28. Komolafe T.E., Cao Y., Nguchu B.A. et al. Diagnostic Test Accuracy of Deep Learning Detection of COVID-19: A Systematic Review and Meta-Analysis. *Acad. Radiol.* 2021; 28 (11): 1507–1523. https://doi.org/10.1016/j.acra.2021.08.008
29. Wang Q., Ma J., Zhang L., Xie L. Diagnostic performance of corona virus disease 2019 chest computer tomography image recognition based on deep learning: Systematic review and meta-analysis. *Medicine.* 2022; 101: 42 (e31346). http://dx.doi.org/10.1097/MD.0000000000031346
30. Campello C.A., Castanha E.B., Vilardo M. et al. Machine learning for malignant versus benign focal liver lesions on US and CEUS: a meta-analysis. *Abdom. Radiology*. 2023; 48: 3114–3126. https://doi.org/10.1007/s00261-023-03984-0
31. Dumitrescu E.A., Ungureanu B.S., Cazacu I.M. et al. Diagnostic Value of Artificial Intelligence-Assisted Endoscopic Ultrasound for Pancreatic Cancer: A Systematic Review and Meta-Analysis. *Diagnostics*. 2022; 12: 309. https://doi.org/10.3390/diagnostics12020309
32. Potipimpanon P., Charakorn N., Hirunwiwatkul P. A comparison of artificial intelligence versus radiologists in the diagnosis of thyroid nodules using ultrasonography: a systematic review and meta-analysis. *Eur. Arch. of Oto-Rhino-Laryngol*. 2022; 279: 5363–5373. https://doi.org/10.1007/s00405-022-07436-1
33. Rescinito R., Ratti M., Payedimarri A.B., Panella M. Prediction Models for Intrauterine Growth Restriction Using Artificial Intelligence and Machine Learning: A Systematic Review and Meta-Analysis. *Healthcare.* 2023; 11: 1617. https://doi.org/10.3390/healthcare11111617
34. Abesi F., Jamali A.S., Zamani M. Accuracy of artificial intelligence in the detection and segmentation of oral and maxillofacial structures using cone-beam computed tomography images: a systematic review and meta-analysisn. *Polish J. Radiol*. 2023; 88: e256–e263. https://doi.org/10.5114/pjr.2023.127624
35. Badr F.F., Jadu F.M. Performance of artificial intelligence using oral and maxillofacial CBCT images: A systematic review and meta-analysis. *Nigerian J. Clin. Pract*. 2022; 25: 1918–1927. https://doi.org/10.4103/njcp.njcp_394_22
36. Soffer S., Klang E., Shimon O. et al. Deep learning for pulmonary embolism detection on computed tomography pulmonary angiogram: a systematic review and meta-analysis. *Scientific Reports*. 2021; 11: 15814. https://doi.org/10.1038/s41598-021-95249-3
37. Islam M., Poly T.N., Walther B.A. et al. Deep Learning for the Diagnosis of Esophageal Cancer in Endoscopic Images: A Systematic Review and Meta-Analysis. *Cancers*. 2022; 14 (23): 5996. https://doi.org/10.3390/cancers14235996
38. Xu Y., Ding W., Wang Y. et al. Comparison of diagnostic performance between convolutional neural networks and human endoscopists for diagnosis of colorectal polyp:

A systematic review and meta-analysis. *PLoS ONE.* 2021; 16 (2): e0246892. https://doi.org/10.1371/journal.pone.0246892
39. Taib B.G., Karwath A., Wensley K. et al. Artificial intelligence in the management and treatment of burns: A systematic review and meta-analyses. *J. Plast. Reconstructive & Aesthetic Surg*. 2023; 77: 133–161. https://doi.org/10.1016/j.bjps.2022.11.049
40. Yang C., Qin L., Xie Y., Liao J. Deep learning in CT image segmentation of cervical cancer: a systematic review and meta-analysis. *Radiation Oncology.* 2022; 17: 175. https://doi.org/10.1186/s13014-022-02148-6
41. Electronic resource: QUADAS-2, University of Bristol: https://www.bristol.ac.uk/population-health-sciences/projects/quadas/quadas-2/
42. Electronic resource: Radiomics Quality Score - RQS 2.0.: https://www.radiomics.world/rqs2
43. Kim H.Y., Cho S.J., Sunwoo L. et al. Classification of true progression after radiotherapy of brain metastasis on MRI using artificial intelligence: a systematic review and meta-analysis. *Neuro-Oncology Advances*. 2021; 3 (1): 1–12. https://doi.org/10.1093/noajnl/vdab080
44. Electronic resource: International Prospective Register of Systematic Reviews. https://www.crd.york.ac.uk/PROSPERO/
45. Adamou A., Beltsios E.T., Bania A. et al. Artificial intelligence-driven ASPECTS for the detection of early stroke changes in non-contrast CT: a systematic review and meta-analysis. *J. NeuroIntervent. Surg.* 2023; 15: e298–e304. https://doi.org/10.1136/jnis-2022-019447
46. Nabizadeh F., Ramezannezhad E., Kargar A. et al. Diagnostic performance of artificial intelligence in multiple sclerosis: a systematic review and meta-analysis. *Neurol. Sci*. 2023; 44: 499–517. h ttps://doi.org/10.1007/s10072-022-06460-7
47. Arbabshirani M.R., Fornwalt B.K., Mongelluzzo G.J. et al. Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. *NPJ Dig. Med.* 2018; 1: 9. https://doi.org/10.1038/s41746-017-0015-z
48. Ginat D. Implementation of machine learning software on the radiology worklist decreases scan view delay for the detection of intracranial hemorrhage on CT. *Brain Sci.* 2021.11(7):832.https://doi.org/10.3390/brainsci11070832
49. Salehinejad H., Kitamura J., Ditkofsky N. et al. A real-world demonstration of machine learning generalizability in the detection of intracranial hemorrhage on head computerized tomography. *Sci. Reports*. 2021; 11: 17051. https://doi.org/10.1038/s41598-021-95533-2
50. Владзимирский А.В., Омелянская О.В., Шулькин.И.М., Четвериков С.Ф., Семенов С.С., Новик В.П., Павлов Н.А., Арзамасов К.М., Андрейченко А.Е., Бобровская Т.М., Блохин И.А., Гомболевский В.А. MosMedData: НДКТ с признаков рака легкого. Свидетельство о государственной регистрации базы данных № 2023621253 Российская Федерация. MosMedData: НДКТ с признаками рака легкого: № 2023620623: заявл. 10.03.2023: опубл. 18.04.2023 / заявитель ГБУЗ города Москвы "Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы".
Vladzimirskyy A.V., Omelyanskaya O.V., Shulkin I.M. et al. LDCT of the lung cancer. Certificate of state registration of the database № 2023621253 Russian Federation. MosMedData: LDCT of the lung cancer: № 2023620623: declared 10.03.2023: published 18.04.2023 / applicant "Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of Moscow Healthcare Department". (In Russian)
51. Морозов С.П., Андрейченко А.Е., Блохин И.А., Гележе П.Б., Гончар А.П., Николаев А.Е., Павлов Н.А., Чернина В.Ю., Гомболевский В.А. MosMedData: датасет 1110 компьютерных томографий органов грудной клетки, выполненных во время эпидемии COVID-19. *Digital Diagnostics.* 2020; 1 (1): 49–59. https://doi.org/10.17816/DD46826
Morozov S.P., Andreychenko A.E., Blokhin I.A. et al. MosMedData: data set of 1110 chest CT scans performed during the COVID-19 epidemic. *Dig. Diagn.* 2020; 1 (1): 49–59. https://doi.org/10.17816/DD46826. (In Russian)
52. Гомболевский В.А., Харламов К.А., Пятницкий И.А. и др. Шаблоны протоколов описаний исследований по специальности "Рентгенология". Компьютерная томография: Методические рекомендации № 23 / Москва: ГБУЗ города Москвы "Научно-практический центр медицинской радиологии Департамента здравоохранения города Москвы", 2016а. 31 с.
Gombolevskiy V.A., Kharlamov K.A., Pyatnitskii I.A. et al. Templates for protocol descriptions of studies in the specialty "Radiology". Computed tomography: Guidelines № 23 / Moscow: "Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of Moscow Healthcare Department", 2016a, 31 p. (In Russian)
53. Гомболевский В.А., Харламов К.А., Пятницкий И.А. и др. Шаблоны протоколов описаний исследований по специальности "Рентгенология". Магнитно-резонансная томография: Методические рекомендации № 21 / Москва: Государственное бюджетное учреждение здравоохранения города Москвы "Научно-практический центр медицинской радиологии Департамента здравоохранения города Москвы", 2016б. 41 с.
Gombolevskiy V.A., Kharlamov K.A., Pyatnitskii I.A. et al. Templates for protocol descriptions of studies in the specialty "Radiology". Magnetic resonance imaging: Guidelines № 21 / Moscow: "Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of Moscow Healthcare Department", 2016b, 41 p. (In Russian)
54. Aggarwal R., Sounderajah V., Martin G. et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. *NPJ Dig. Med.* 2021; 4: 65. https://doi.org/10.1038/s41746-021-00438-z
55. Liu M., Wu J., Wang N. et al. The value of artificial intelligence in the diagnosis of lung cancer: A systematic review and meta-analysis. *PLoS ONE.* 2023; 18 (3): e0273445. https://doi.org/10.1371/journal.pone.0273445
56. Jørgensen M.D., Antulov R., Hess S., Lysdahlgaard S. Convolutional neural network performance compared to radiologists in detecting intracranial hemorrhage from brain computed tomography: A systematic review and meta-analysis. *Eur. J. Radiol.* 2022; 146: 110073. https://doi.org/10.1016/j.ejrad.2021.110073
57. Evangelista K., de Freitas Silva B.S., Yamamoto-Silva F.P. et al. Accuracy of artificial intelligence for tooth extraction decision-making in orthodontics: a systematic review and meta-analysis. *Clin. Oral Invest.* 2022; 26: 6893–6905. https://doi.org/10.1007/s00784-022-04742-0

Для корреспонденции*: Нанова Ольга Геннадьевна – e-mail: nanova@mail.ru

Васильев Юрий Александрович – канд. мед. наук, директор ГБУЗ города Москвы “Научно-практический клинический центр диагностики и телемедицинских технологий ДЗ города Москвы”, Москва. https://orcid.org/0000-0002-0208-5218

Владзимирский Антон Вячеславович – доктор мед. наук, профессор, заместитель директора по научной работе ГБУЗ города Москвы “Научно-практический клинический центр диагностики и телемедицинских технологий ДЗ города Москвы”, Москва. http://orcid.org/0000-0002-2990-7736

Омелянская Ольга Васильевна – руководитель по управлению подразделениями Дирекции наука ГБУЗ города Москвы “Научно-практический клинический центр диагностики и телемедицинских технологий ДЗ города Москвы”, Москва. http://orcid.org/0000-0002-0245-4431

Решетников Роман Владимирович – канд. физ.-мат. наук, руководитель отдела научных медицинских исследований ГБУЗ города Москвы “Научно-практический клинический центр диагностики и телемедицинских технологий ДЗ города Москвы”, Москва. http://orcid.org/0000-0002-9661-0254

Блохин Иван Андреевич – начальник сектора исследований в лучевой диагностике ГБУЗ города Москвы “Научно-практический клинический центр диагностики и телемедицинских технологий ДЗ города Москвы”, Москва. http://orcid.org/0000-0002-2681-9378

Коденко Мария Романовна – младший научный сотрудник ГБУЗ города Москвы “Научно-практический клинический центр диагностики и телемедицинских технологий ДЗ города Москвы”, Москва. https://orcid.org/0000-0002-0166-3768

Нанова Ольга Геннадьевна – канд. биол. наук, ведущий научный сотрудник ГБУЗ города Москвы “Научно-практический клинический центр диагностики и телемедицинских технологий ДЗ города Москвы”, Москва. https://orcid.org/0000-0001-8886-3684

Contact*: Olga G. Nanova– e-mail: nanova@mail.ru

Yuri A. Vasiliev – Cand. of Sci. (Med.), Director of the Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Healthcare Department, Moscow. https://orcid.org/0000-0002-0208-5218

Anton V. Vladzymyrskyy – Doct. of Sci. (Med.), Deputy Director of the Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Healthcare Department, Moscow. http://orcid.org/0000-0002-2990-7736

Olga V. Omelyanskaya – Head of Division Management, Science of the Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Healthcare Department, Moscow. http://orcid.org/0000-0002-0245-4431

Roman V. Reshetnikov – Cand. of Sci. (Phys.-Math.), Head of Scientific Medical Research Department of the Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Healthcare Department, Moscow. http://orcid.org/0000-0002-9661-0254

Ivan A. Blokhin – Head of Research Sector in Diagnostic Radiology of the Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Healthcare Department, Moscow. http://orcid.org/0000-0002-2681-9378

Maria R. Kodenko – Junior researcher of the Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Healthcare Department, Moscow. https://orcid.org/0000-0002-0166-3768

Olga G. Nanova – Cand. of Sci. (Biol.), Leading researcher of the Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Healthcare Department, Moscow. https://orcid.org/0000-0001-8886-3684