



ОРИГИНАЛЬНАЯ СТАТЬЯ

DOI: 10.21045/1811-0185-2023-8-54-67

УДК: 618.19–006

ПЕРВЫЕ 10000 МАММОГРАФИЧЕСКИХ ИССЛЕДОВАНИЙ, ВЫПОЛНЕННЫХ В РАМКАХ УСЛУГИ «ОПИСАНИЕ И ИНТЕРПРЕТАЦИЯ ДАННЫХ МАММОГРАФИЧЕСКОГО ИССЛЕДОВАНИЯ С ИСПОЛЬЗОВАНИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА»

**Ю.А. Васильев^а, А.В. Владзимирский^б, К.М. Арзамасов^с,
И.М. Шулькин^д, Л.Е. Аксенова^е ✉, Л.Д. Пестренин^ф,
С.С. Семенов^г, Д.В. Бондарчук^h, И.В. Смирнов^и**

^{а, б, с, д, е, ф, г, и} Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы, г. Москва, Россия.

^а <https://orcid.org/0000-0002-0208-5218>; ^б <https://orcid.org/0000-0002-2990-7736>;

^с <https://orcid.org/0000-0001-7786-0349>; ^д <https://orcid.org/0000-0002-7613-5273>;

^е <https://orcid.org/0000-0003-0885-1355>; ^ф <https://orcid.org/0000-0002-1786-4329>;

^г <https://orcid.org/0000-0003-2585-0864>; ^h <https://orcid.org/0000-0001-8752-0591>.

✉ Автор для корреспонденции: Аксенова Л.Е.

АННОТАЦИЯ

Актуальность. Технологии искусственного интеллекта имеют огромный потенциал для повышения эффективности скрининговых программ по выявлению злокачественных новообразований молочной железы. Учитывая высокую социальную, демографическую и экономическую значимость массовых профилактических исследований, не подлежит сомнению тот факт, что диагностическая точность искусственного интеллекта должна соответствовать точности врачей-рентгенологов или даже превышать ее. В связи с этим необходимы исследования, посвященные сравнению точности программного обеспечения на основе технологии искусственного интеллекта и врачей-рентгенологов на потоке при проведении маммографических исследований.

Цель: оценить качество оказания медицинской услуги «Описание и интерпретация данных маммографического исследования с использованием искусственного интеллекта» в рамках скрининга.

Материал и методы. Выборка для анализа состояла из 9684 цифровых маммографических исследований. Для каждого исследования категория BI-RADS определялась врачом-рентгенологом и с помощью программного обеспечения на основе технологий искусственного интеллекта (ПО на основе ТИИ), зарегистрированного в РФ в качестве медицинского изделия. 45 исследований из этой выборки со значимыми расхождениями в оценке врача и ПО подверглись экспертному аудиту, в результате которого была выставлена категория BI-RADS по мнению врача-эксперта.

Результаты. При оценке средневзвешенных значений статистически значимых различий между результатами врача и ПО на основе ТИИ для 9684 цифровых маммографических исследований не наблюдалось. Оценка согласованности врача и ПО показала, что совпадения наблюдаются в 43,89% случаев для шкалы BI-RADS и в 80,69% – 84,10% для бинарных шкал. Наличие случая, при котором патология, определенная с помощью ПО и подтвержденная при пересмотре результатов экспертом, была пропущена врачом, говорит о перспективности использования ПО на основе ТИИ для оценки маммографических исследований и требует проведения дальнейших исследований.

Заключение. При оценке маммографических исследований согласованность между решением на основе ТИИ и врачом-рентгенологом, достигает 84,10%, при этом ПО чаще присваивает более высокую категорию BI-RADS. Экспертный пересмотр части этих расхождений показал потенциальное снижение количества пропусков злокачественных новообразований молочных желез с помощью ПО.

Ключевые слова: профилактические исследования, маммография, искусственный интеллект, диагностическая точность.

Для цитирования: Васильев Ю.А., Владзимирский А.В., Арзамасов К.М., Шулькин И.М., Аксенова Л.Е., Пестренин Л.Д., Семенов С.С., Бондарчук Д.В., Смирнов И.В. Первые 10000 маммографических исследований, выполненных в рамках услуги «Описание и интерпретация данных маммографического исследования с использованием искусственного интеллекта». Менеджер здравоохранения. 2023; 8:54–67. DOI: 10.21045/1811-0185-2023-8-54-67

© Васильев Ю.А., Владзимирский А.В., Арзамасов К.М., Шулькин И.М., Аксенова Л.Е., Пестренин Л.Д., Семенов С.С., Бондарчук Д.В., Смирнов И.В., 2023 г.



Введение

Злокачественные новообразования (ЗНО) молочной железы представляют собой значимую медицинскую и социально-экономическую проблему, поскольку являются одной из самых распространенных онкологических патологий и занимают первое место в структуре смертности от онкологических заболеваний у женского населения [1].

На сегодняшний день с целью раннего выявления ЗНО молочной железы проводится скрининговое обследование – рентгеновская маммография (ММГ). При этом важным ограничением является тот факт, что точность интерпретации маммографии опытными рентгенологами сильно различается [2]. С целью повышения выявляемости ЗНО было предложено проведение скрининговых исследований с участием двух врачей, что легло в основу действующих нормативно-правовых актов: маммография обеих молочных желез в двух проекциях с двойным прочтением рентгенограмм проводится для женщин в возрасте от 40 до 75 лет 1 раз в 2 года [3]. Скрининг на ЗНО молочной железы является эффективной мерой для выявления ранней стадии заболевания и повышения выживаемости онкологических больных, а также способствует снижению смертности и распространенности ЗНО на поздних стадиях [4].

Прогресс в области технологий искусственного интеллекта (ТИИ) и, в частности, компьютерного зрения привел к появлению программного обеспечения (ПО) на основе ТИИ для оценки цифровых ММГ, с помощью которого осуществляется анализ ММГ, разметка подозрительных областей и, в ряде случаев, дифференциальная диагностика выявленных изменений без дополнительных исследований и финансовых затрат [5, 6].

Существуют исследования, которые демонстрируют сопоставимую с врачом-рентгенологом чувствительность отдельного ИИ-алгоритма [7], а отдельные авторы указывают на превосходство показателей диагностической точности ИИ, по сравнению с врачом [8]. Наши предыдущие исследования указывают на высокие показатели диагностической точности и перспективность применения ПО на основе ТИИ для профилактических маммографических исследований [9]. В исследовании McKinney SM et al. была показана возможность использования алгоритмов искусственного интеллекта в модели двойного чтения при сохранении производительности и снижении нагрузки на второго читателя на 88% [8]. Другим подходом к скринингу

на основе ТИИ являлась модель, в которой через второе чтение проходили только те случаи, в которых с помощью ПО была выявлена патология [7].

Для оценки точности моделей скрининга, включающих использование решений на основе искусственного интеллекта, принято использовать такие показатели, как чувствительность, специфичность, а также количество ложноположительных и ложноотрицательных заключений. Кроме того, возможна оценка согласованности между врачом и ПО с помощью количественной оценки вероятности совпадений и расхождений.

Таким образом, на сегодняшний день представляются актуальными исследования, направленные на оценку эффективности применения ПО на основе ТИИ на потоке при проведении маммографического скрининга. Результаты эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы [10–12] позволили осуществить качественный переход от научных исследований к системному внедрению процесса интерпретации маммографии (ММГ) с помощью ИИ-сервисов в практическое здравоохранение столицы. Организационно это стало возможно благодаря включению новой медицинской услуги «Описание и интерпретация данных маммографического исследования с использованием искусственного интеллекта» в территориальную программу государственных гарантий бесплатного оказания гражданам медицинской помощи в городе Москве.

Целью настоящей работы являлась качественная оценка работы ПО на основе ТИИ при интерпретации цифровых ММГ в рамках новой медицинской услуги «Описание и интерпретация данных маммографического исследования с использованием искусственного интеллекта».

Материал и методы

Дизайн исследования

В настоящем исследовании в качестве материала выступали результаты обработки исследований в рамках новой медицинской услуги с кодом 1601 «Описание и интерпретация данных маммографического исследования с использованием искусственного интеллекта», включающие 10417 цифровых маммографических исследований, которые были выполнены пациенткам в медицинских организациях Департамента Здравоохранения города Москвы в рамках территориальной программы





государственных гарантий бесплатного оказания гражданам медицинской помощи в городе Москве в течение первых двух недель февраля 2023 года.



Процедура проведения исследования

Метод исследования – цифровая маммография обеих МЖ. Каждое маммографическое исследование описывал врач-рентгенолог, работающий по субспециализации «маммография». Часть исследований, которые были отправлены на пересмотр, оценивал врач-маммолог со стажем более 5 лет в оценке маммографических исследований. Описание маммографического исследования проводилось с помощью «Системы отчетности и данных по визуализации молочной железы – Breast Imaging Reporting and Data System» (BI-RADS) путем присваивания исследованию категорий от 0 до 5 в зависимости от степени риска наличия злокачественных образований МЖ [13, 14]. Также каждому маммографическому исследованию была присвоена категория BI-RADS с помощью программного обеспечения на основе ТИИ, которое было зарегистрировано в качестве медицинского изделия в 2021 году [15, 16]. Согласно производителю, такие параметры диагностической точности как AUC, чувствительность, специфичность и точность для данного ПО, равны 0,91, 0,8, 0,92 и 0,86, соответственно.

В качестве общей категории BI-RADS для каждой пациентки было выбрано максимальное значение категории относительно двух молочных желез.

Статистический анализ результатов

Из общей выборки 10417 исследований были исключены исследования, по которым не было результатов работы ИИ-сервиса и / или врачом была выставлена категория BI-RADS0. Таким образом, дальнейшее сравнение результатов осуществлялось для 9684 исследований, из которых была сформирована выборка – 45 случаев с крайними значениями категорий BI-RADS1 и 4–5, которые выставлял врач. Для данной выборки был выполнен дополнительный пересмотр врачом-экспертом с целью формирования итогового заключения. При этом в 37 исследованиях мнения врача и ПО расходились, а в 8 исследованиях в обоих случаях была выставлена категория 0 (норма), относительно бинарной шкалы 1 (рис. 1).

Сравнение результатов ПО на основе ТИИ проводили:

- относительно врача (для 9684 исследований);
- врач относительно эксперта (45 исследований);
- ПО на основе ТИИ относительно эксперта (45 исследований).

В первом случае проводили анализ всех данных (9684 исследований), во втором и третьем анализировали только те исследования, которые были направлены на пересмотр (45 исследований для бинарной шкалы BI-RADS1, 29 исследований для бинарной шкалы BI-RADS2) (рис. 2).

Оценку результатов проводили для следующих шкал: диагностическая шкала BI-RADS1–6, бинарная шкала 1, полученная из диагностической

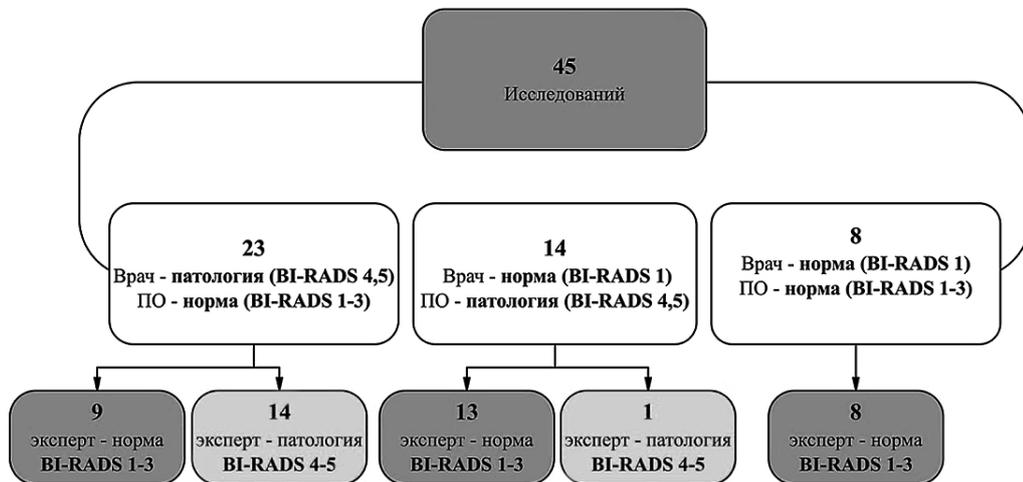


Рис. 1. Визуализация формирования выборки, а также распределения результатов оценки маммографии врачом, ИИ-сервисом и экспертом

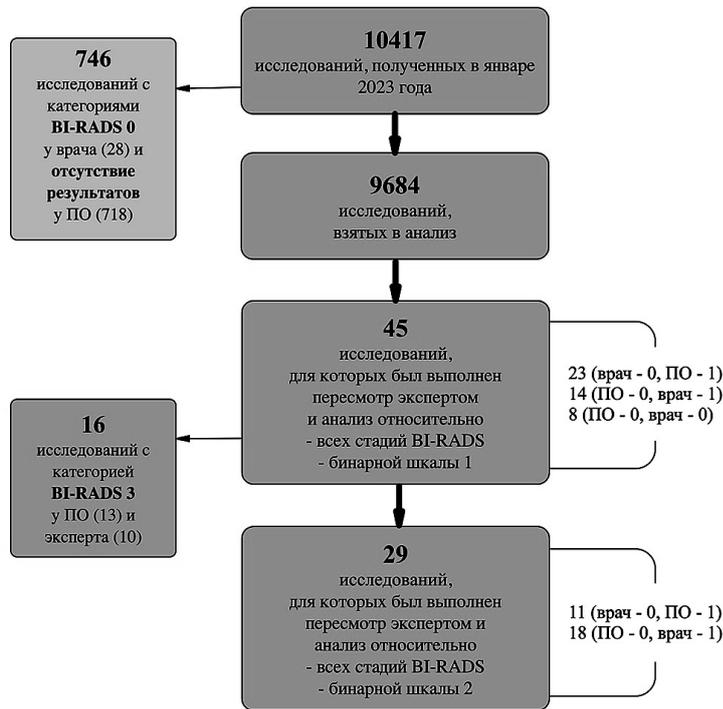


Рис. 2. Блок-схема количественного описания исследований, включенных в анализ

(BI-RADS1–3 = 0, отсутствие признаков ЗНО; BI-RADS4–6 = 1, наличие признаков ЗНО) [17], и бинарная шкала 2 (BI-RADS1,2 = 0, отсутствие признаков ЗНО; BI-RADS4–6 = 1, наличие признаков ЗНО) [18].

Результаты оценивали относительно шкалы BI-RADS за исключением 6-ой категории, которая оценивалась как 5-ая, так как ПО не имело данных

о гистологическом подтверждении. Также оценку проводили относительно бинарных шкал (рис. 3). При исключении категории BI-RADS3 из бинарной шкалы 2 были удалены все строки с результатом BI-RADS3 у врача, ПО на основе ТИИ и эксперта.

Согласованность результатов, выраженная в количественной оценке вероятности совпадений и расхождений, а также коэффициент межэкспертного

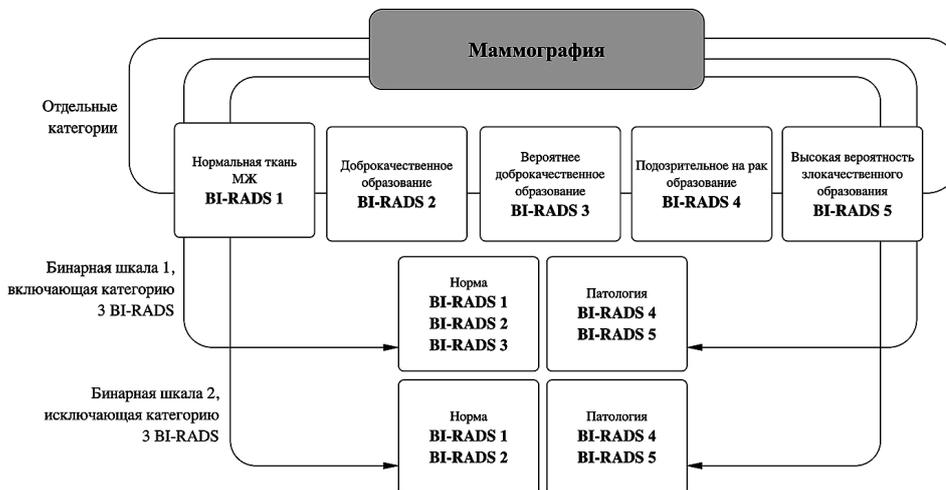


Рис. 3. Значения категорий шкалы BI-RADS, а также бинарные шкалы, на основании которых проводился анализ цифровой маммографии





согласия (коэффициент Каппа Коэна) рассчитывали в парах: ПО / врач, врач / эксперт и ПО / эксперт, относительно отдельных категорий BI-RADS и бинарных шкал 1 и 2.

Оценку точности ПО на основе ТИИ проводили относительно врача (9684 исследования) и эксперта (45 исследований для шкалы BI-RADS1 и 29 исследований для шкалы BI-RADS2) путем расчета следующих метрик: AUC ROC (micro), точности (1), чувствительности (2) и специфичности (3).

(1) Точность = (ИП + ИО) / (ИП + ЛО + ИО + ЛП);

(2) Чувствительность = ИП / (ИП + ЛО);

(3) Специфичность = ИО / (ЛП + ИО).

Также, с целью определения надежности ПО, в работе проводили оценку количества ложноположительных (ЛП), ложноотрицательных (ЛО) результатов, ИП) истинноположительных и ИО (истинноотрицательных) результатов. Ложноположительными считались исследования из группы «Отсутствие целевой патологии» (BI-RADS0 по бинарной шкале), которые ошибочно относили к группе «Присутствие целевой патологии» (BI-RADS1 по бинарной шкале), а ложноотрицательным считался результат отнесения ММГ из группы «Присутствие целевой патологии» к группе «Отсутствие целевой патологии». Истинноположительными и истинноотрицательными результатами являлись исследования, правильно отнесенные к группам «Присутствие целевой патологии» и «Отсутствие целевой патологии», соответственно.

ROC-анализ проводился с использованием веб-инструмента (<https://roc-analysis.mosmed.ai/>) для

выполнения ROC анализа результатов диагностических тестов [19].

Все количественные расчеты совпадений, расхождений, коэффициента Каппа Коэна, а также расчеты метрик ЛП, ЛО, ИП, ИО, AUC ROC, точности, чувствительности и специфичности проводили с использованием языка программирования Python и библиотек pandas, numpy, sklearn.

Результаты

1. Сравнение оценок маммографии, выставленных ПО на основе ТИИ, врачом-рентгенологом и экспертом, по шкале с категориями 1–5 BI-RADS и бинарным шкалам

Количественные характеристики результатов определения категорий BI-RADS врачом и ПО на основе ТИИ, относительно всех 9684 исследований, а также экспертом, врачом и ПО, относительно выборки из 45 исследований, приведены в *таблице 1*. Значения средневзвешенных значений и стандартных отклонений для результатов ПО на основе ТИИ и врача, относительно всех 9684 исследований, показали, что статистически значимых различий между этими результатами нет.

Количественное описание результатов маммографических исследований относительно бинарных шкал приведено в *таблице 2*. Можно отметить, что ПО на основе ТИИ чаще определяет исследование, как содержащее патологию (17,08–21,38%), в отличие от врача (2,46–3,1%), относительно всех исследований. При этом, распределение

Таблица 1

Распределение категорий BI-RADS, выставленных ПО на основе ТИИ, врачом и экспертом в общей выборке, включающей 9684 исследования

	BI-RADS1 (% от общего количества)	BI-RADS2 (% от общего количества)	BI-RADS3 (% от общего количества)	BI-RADS4 (% от общего количества)	BI-RADS5 (% от общего количества)	Средневзвешенное значение категории BI-RADS	Стандартное отклонение категории BI-RADS	Общее количество исследований
Врач	1225 (12,65)	7686 (79,37)	535 (5,53)	203 (2,1)	35 (0,36)	1,982	0,546	9684
ПО на основе ТИИ	1595 (16,47)	3968 (40,98)	2467 (25,48)	1629 (16,82)	25 (0,26)	2,434	0,963	
Эксперт	2 (4,44)	19 (42,22)	9 (20)	15 (33,33)	-	N/A	N/A	45
Врач	22 (48,89)	-	-	22 (48,89)	1 (2,2)	N/A	N/A	
ПО на основе ТИИ	8 (17,78)	15 (33,33)	8 (17,78)	14 (31,11)	-	N/A	N/A	

Примечание: N/A – выборка не имеет нормального распределения ввиду особенностей ее формирования



Таблица 2

Распределение по бинарным шкалам категорий BI-RADS, выставленных ПО на основе ТИИ, врачом и экспертом в выборках из 9684 и 45 исследований

Шкала	Оценщик результата	Количество исследований в группе 0 (% от общего количества)	Количество исследований в группе 1 (% от общего количества)	Всего исследований
Бинарная шкала 1	Врач	9446 (97,54)	238 (2,46)	9684
	ПО на основе ТИИ	8030 (82,92)	1654 (17,08)	
Бинарная шкала 2	Врач	6604 (96,90)	211 (3,10)	6815
	ПО на основе ТИИ	5358 (78,62)	1457 (21,38)	
Бинарная шкала 1	Эксперт	30 (66,67)	15 (33,33)	45
	Врач	22 (48,89)	23 (51,11)	
	ПО на основе ТИИ	31 (68,89)	14 (31,11)	
Бинарная шкала 2	Эксперт	14 (48,28)	15 (51,72)	29
	Врач	11 (37,93)	18 (62,07)	
	ПО на основе ТИИ	18 (62,07)	11 (37,93)	

в выборке из 45 исследований показывает обратную ситуацию для решения на основе искусственного интеллекта (31,11–37,93) относительно врача (51,11–62,07), обусловленную способом формирования этой выборки.

Оценка согласованности врача и ПО на основе ТИИ, относительно 9684 исследований, показала, что совпадения с врачом наблюдаются в 43,89%

случаев для шкалы BI-RADS; в 84,10% случаев для бинарной шкалы 1; в 80,69% случаев для бинарной шкалы 2 (таблица 3).

При оценке результатов, которые были получены с помощью решения на основе искусственного интеллекта, было показано, что превалирует количество случаев завышения категорий BI-RADS, определенных врачом. Так, в случае отдельных категорий

Таблица 3

Оценка согласованности (совпадений и расхождений, а также показателей завышения, занижения и коэффициента Каппа Коэна) между врачом и ПО на основе ТИИ, ПО на основе ТИИ и экспертом, врачом и экспертом, относительно шкалы BI-RADS и бинарных шкал

Шкала	Группы сравнения	Количество совпадений (% от всех ММГ)	Расхождения			Коэффициент Каппа Коэна	Общее количество ММГ
			Количество случаев (% от всех ММГ)	из них			
				Количество завышений* (% от всех расхождений)	Количество занижений* (% от всех расхождений)		
Бинарная шкала 1	ПО и врач	8144 (84,10)	1540 (15,9)	1478 (95,97)	62 (4,03)	0,15	9684
	ПО и эксперт	18 (40)	27 (60)	13 (48,15)	14 (51,85)	-0,372	
	Врач и эксперт	35 (77,78)	10 (22,22)	9 (90)	1 (10)	0,558	
Бинарная шкала 2	ПО и врач	5499 (80,69)	1316 (19,31)	1281 (97,34)	35 (2,66)	0,166	6815
	ПО и эксперт	5 (17,24)	24 (82,76)	10 (41,67)	14 (58,33)	-0,641	
	Врач и эксперт	24 (82,76)	5 (17,24)	4 (80)	1 (20)	0,652	
Шкала BI-RADS	ПО и врач	4250 (43,89)	5434 (56,11)	4154 (76,44)	1280 (23,56)	0,12	9684
	ПО и эксперт	5 (11,11)	40 (88,89)	20 (50)	20 (50)	-0,248	
	Врач и эксперт	15 (33,33)	30 (66,67)	10 (33,33)	20 (66,67)	0,142	

Примечание: * – повышение/понижение категории, выставленной врачом, ПО на основе ТИИ; повышение/понижение категории, выставленной экспертом, врачом или ПО на основе ТИИ.



это значение составило 76,44%, для бинарной шкалы 1–95,97%, а для бинарной шкалы 2–97,34%.

Сравнение оценок ММГ, сделанных ПО на основе ТИИ и экспертом, показывает их совпадение лишь в трети (для бинарной шкалы 1) и четверти случаев (для бинарной шкалы 2). В то же время, оценки врача и эксперта в большинстве случаев совпадали для бинарной шкалы 1 (77,78%) и для бинарной шкалы 2 (82,76%), тогда как при сравнении по шкале BI-RADS процент совпадений составил 33,33%.

При оценке согласованности результатов ПО и врача с помощью коэффициента Каппа Коэна можно отметить, что его значение практически одинаково для всех трех типов шкал и определяется как «незначительное согласие». Данный показатель для результатов эксперта и врача составил от 0,12 до 0,166, что говорит о низкой степени согласованности, а для результатов эксперта и ПО имеет отрицательные значения, которые указывают на то, что согласованность меньше, чем ожидалось бы при случайном совпадении.

2. Характеристики точности ПО на основе ТИИ, относительно врача-рентгенолога и эксперта

Точность ПО на основе ТИИ была оценена с помощью метрик AUC ROC, чувствительности, специфичности и точности, относительно врача и эксперта (таблица 4).

Метрика AUC ROC, измеренная для ПО относительно врача, на выборке из 9684 исследований, имела незначительные различия, при оценке относительно различных бинарных шкал (0,792 и 0,82),

и была ниже при оценке относительно шкалы BI-RADS (0,65). В то же время AUC ROC для решения на основе искусственного интеллекта, измеренное относительно эксперта, имела значительно меньшие значения от 0,176 до 0,316 для бинарных шкал и 0,41 для шкалы BI-RADS, что обусловлено формированием выборки из ложных случаев срабатывания. При оценке точности врача относительно эксперта AUC ROC составила 0,817 и 0,824 для бинарных шкал 1 и 2 соответственно и 0,54 для шкалы BI-RADS.

Значения AUC ROC для отдельных категорий шкалы BI-RADS, измеренные для ПО на основе ТИИ относительно врача и эксперта, а также для врача относительно эксперта, представлены на рис. 4. Тем не менее, в последних двух случаях расчет данных значений проводился на выборке, содержащей 45 исследований, что может представлять ограниченную достоверность результатов.

В таблице 5 представлено количество ложноположительных, ложноотрицательных, истинноположительных и истинноотрицательных оценок ММГ, ПО относительно врача-рентгенолога и эксперта.

Для ПО на основе ТИИ было показано значительно большее количество ложноположительных результатов (15,26% и 18,8% по бинарным шкалам) при оценке относительно врача, в отличие от ложноотрицательных результатов (0,64% и 0,51% по бинарным шкалам). Относительно эксперта решение на основе искусственного интеллекта, наоборот, в большем количестве случаев показывал ЛО результаты (31,11% и 48,28% по бинарным шкалам), в отличие от ЛП результатов (28,89% и 34,48%). При оценке врача относительно эксперта по бинарным

Таблица 4

Характеристики точности ПО на основе ТИИ относительно врача-рентгенолога и эксперта; врача-рентгенолога относительно эксперта

Шкала	Объект оценки относительно истинного значения (врача или эксперта)	AUC (micro)	Чувствительность	Специфичность	Точность	Всего исследований
Бинарная шкала 1	ПО относительно врача	0,792	0,74	0,844	0,84	9684
	ПО относительно эксперта	0,316	0,067	0,567	0,4	
	Врач относительно эксперта	0,817	0,933	0,7	0,777	
Бинарная шкала 2	ПО относительно врача	0,82	0,834	0,806	0,807	6815
	ПО относительно эксперта	0,176	0,067	0,286	0,172	
	Врач относительно эксперта	0,824	0,933	0,714	0,828	
Шкала BI-RADS	ПО относительно врача	0,65	0,85	0,846	0,83	9684
	ПО относительно эксперта	0,41	0,4	0,802	0,702	
	Врач относительно эксперта	0,54	0,927	0,694	0,878	

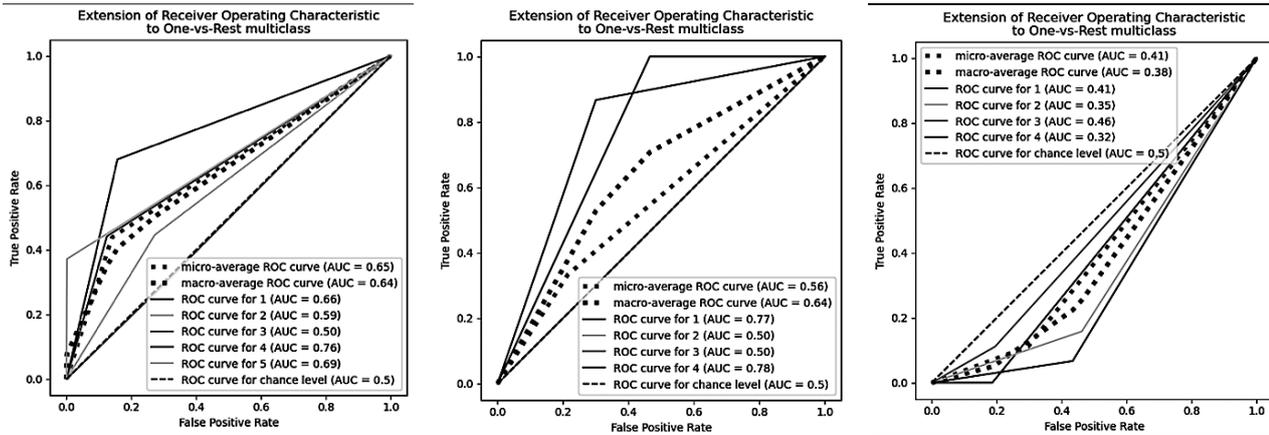


Рис. 4. Графики AUC ROC для оценки точности ПО на основе ТИИ относительно врача (1), врача относительно эксперта (2) и ПО на основе ТИИ относительно эксперта (3)

Таблица 5

Четырехпольная таблица результатов оценки маммографии (ММГ)

Шкала	Группы сравнения	Количество ЛП оценок (% от всех ММГ)	Количество ЛО оценок (% от всех ММГ)	Количество ИП оценок (% от всех ММГ)	Количество ИО оценок (% от всех ММГ)	Всего ММГ
Бинарная шкала 1	ПО относительно врача	1478 (15,26)	62 (0,64)	176 (1,82)	7968 (82,28)	9684
	ПО относительно эксперта	13 (28,89)	14 (31,11)	1 (2,22)	17 (37,78)	45
	Врач относительно эксперта	9 (20)	1 (2,22)	14 (31,11)	21 (46,67)	45
Бинарная шкала 2	ПО относительно врача	1281 (18,8)	35 (0,51)	176 (2,58)	5323 (78,11)	6815
	ПО относительно эксперта	10 (34,48)	14 (48,28)	1 (3,45)	4 (13,79)	29
	Врач относительно эксперта	4 (13,79)	1 (3,45)	14 (48,28)	10 (34,48)	29

шкалам доля ЛП результатов (20% и 13,79%) превышала долю ЛО результатов (2,22% и 3,45%).

Количество совпадений оценок ИИ-сервиса, врача-рентгенолога и эксперта в 45 исследованиях, отправленных эксперту на пересмотр, представлено на рис. 5.

При переводе отдельных категорий BI-RADS в бинарную шкалу 1, появились случаи согласия врача, ПО на основе ТИИ и эксперта в связи с тем, что категории 1 и 3 попали в группу «Норма». При оценке по бинарной шкале 1 эксперт согласился

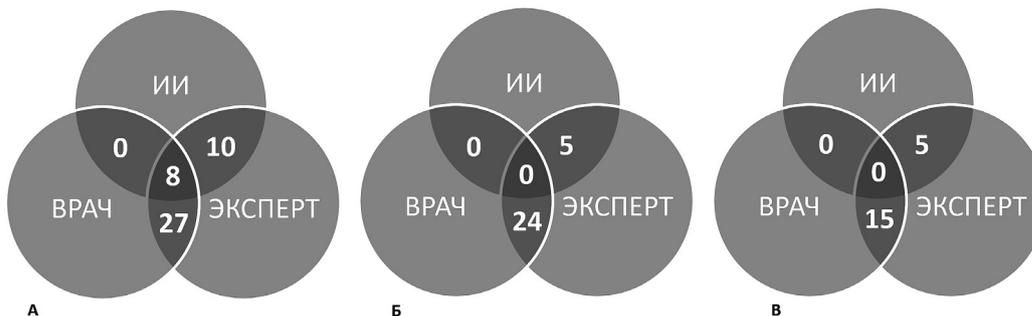


Рис. 5. Совпадения оценок ПО на основе ТИИ, врача-рентгенолога и эксперта: - в 45 исследованиях, оцененных относительно бинарной шкалы 1 (А); - в 29 исследованиях, оцененных относительно бинарной шкалы 2 (Б); - в 45 исследованиях, оцененных относительно полной шкалы BI-RADS (В)





с категорией, выставленной ПО, и не согласился с категорией, выставленной врачом, в 10 случаях из 45 (22,2%); согласился с врачом, но не согласился с ПО в 27 случаях из 45 (60,0%).

При оценке по бинарной шкале 2 эксперт согласился с категорией, выставленной ПО, и не согласился с категорией, выставленной врачом, в 5 случаях из 29 (17,2%); согласился с врачом, но не согласился с ПО в 24 случаях из 29 (82,8%).

При оценке по шкале BI-RADS эксперт согласился только с ПО на основе ТИИ в 5 случаях из 45 (11,1%), только с врачом – в 15 случаях из 45 (33,3%), не согласился ни с ПО на основе ТИИ, ни с врачом в 25 случаях из 45 (55,6%).

Обсуждение

В ходе настоящей работы нами было проанализировано 9684 маммографических исследований, проведенных в рамках территориальной программы государственных гарантий бесплатного оказания гражданам медицинской помощи в городе Москве. Количество случаев, которые были дополнительно оценены с помощью эксперта, составило 45.

С целью обеспечить сравнимость полученных данных с результатами, полученными в исследованиях других авторов, оценки по шкале BI-RADS в настоящем исследовании были переведены в бинарную шкалу 1 [17] и 2 [18].

Категория BI-RADS3 соответствует «вероятно, доброкачественным изменениям». Вероятность развития ЗНО молочной железы у пациентов с рентгенологическими признаками, соответствующими данной категории, составляет от 0 до 2% [20]. Именно поэтому исследования с категорией BI-RADS3 в полной мере не могут быть отнесены ни к группе «Отсутствие признаков ЗНО», ни к группе «Присутствие признаков ЗНО». В связи с этим использование в исследовании, в том числе, бинарной шкалы 2, исключающей 3 категорию, представляется оправданным с методологической точки зрения. Полагаем, что необходимы дальнейшие исследования, направленные на изучение точности категорий BI-RADS, выставляемых врачом-рентгенологом, путем экспертного пересмотра достаточных выборок маммографических исследований.

В результате сравнения категорий BI-RADS по бинарной шкале 2 между решением на основе искусственного интеллекта и врачом было показано, что в 19,31% случаев их мнения расходятся. При этом, согласно Coolen AMP et al., расхождения, измеренные с помощью такой же бинарной шкалы

BI-RADS, между врачами рентгенологами наблюдаются в 28,16% случаях, что значительно отличается от результатов, полученных в настоящем исследовании [18].

При оценке результатов по бинарной шкале 1 было показано, что мнения ПО на основе ТИИ и врача расходятся в 15,9% случаев. В статье Chang Sen LQ et al. сравнение, проведенное с использованием аналогичной шкалы, в которой категория BI-RADS3 была отнесена к одной группе с категориями BI-RADS1 и 2, показало, что расхождения между врачами-рентгенологами наблюдаются в 30,44% случаях. Данные различия могут быть обусловлены разным размером выборок. Так, в настоящем исследовании было проанализировано 9684 исследований, в то время как в исследовании Chang Sen LQ et al. размер выборки составил 1669 [17].

В настоящей работе значение коэффициента Каппа Коэна между врачом и ПО на основе ТИИ, измеренное относительно всех 9684 исследований, составило 0,15, что приравнивается к значению «незначительное согласие». При этом в работе Moradi M et al. показано, что согласованность между двумя врачами-рентгенологами при оценке 642 исследований, измеренная с помощью коэффициента Каппа Коэна, составляет 0,74 [21]. Согласно авторам другого исследования, коэффициент Каппа Коэна составляет 0,85, при этом измерение проводилось на 10763 исследованиях [22]. Таким образом, можно отметить, что согласованность между ПО на основе ТИИ и врачом, измеренная с помощью коэффициента Каппа Коэна, в нашем исследовании оказалась значительно меньше, чем в других исследованиях.

В дополнение к числу совпадений и расхождений при оценке качества скрининга важными показателями является количество ЛП и ЛО результатов. Согласно полученным данным, число ЛП и ЛО результатов, рассчитанное для ПО относительно врача, составляет 15,26% и 18,8% и 0,64% и 0,51%, а относительно эксперта – 28,89% и 34,48% и 31,11% и 48,28% соответственно. В то же время Lehman CD и соавторами было показано, что количество ЛП и ЛО результатов, рассчитанное для ПО на основе ТИИ относительно врача, составляет 4,76% и 2,32% [11]. Согласно Flemban AF, частота случаев гипердиагностики (ЛП результатов) при скрининговой маммографии среди женщин в возрасте 40 лет и старше, составляет 12,6% [23]. Подобные случаи увеличивают стоимость скрининга, усложняют оценку программ



и не приносят пользу пациенту, так как выявленное заболевание не угрожает его здоровью.

Тем не менее, полученные результаты, измеренные относительно двух выборок – 9684 и 45 исследований, различаются, в связи с тем, что вторая выборка содержала только случаи выставления крайних категорий 1, 4, 5 BI-RADS, определенных врачом.

Показатель AUC ROC, измеренный на ограниченной выборке для оценки точности ПО на основе ТИИ относительно эксперта, составил 0,176–0,41. Такие значения ($<0,5$) указывают на некорректную работу классификатора, что было обусловлено формированием этой выборки из исследований с некорректным срабатыванием алгоритма, поэтому данные значения некорректно использовать для оценки качества работы искусственного интеллекта. В то же время AUC ROC для врача-рентгенолога, измеренный относительно результата эксперта, был равен 0,54–0,824. Согласно Rodriguez-Ruiz A et al., средний AUC ROC для ПО на основе ТИИ составил 0,89 относительно доказанных гистологически результатов, а для врачей-радиологов – 0,87, что говорит о повышении точности диагностики с помощью ПО, но не согласуется с результатами настоящего исследования [24]. В настоящем исследовании не было возможности проведения гистологического подтверждения наличия патологии, однако ранее проведенное исследование демонстрирует более высокие показатели диагностической точности на гистологически валидированном наборе данных [25]. Таким образом, учитывая различия в размерах выборок, а также используемые шкалы BI-RADS, необходимо провести дополнительные исследования, чтобы с большей уверенностью сравнивать подобные результаты.

Чувствительность ПО на основе ТИИ относительно врача-рентгенолога, в зависимости от шкалы оценки, составила от 0,74 до 0,85, что является более низким значением по сравнению с данными (0,89–1,0) мультицентровой проспективной валидации технологий искусственного интеллекта. Специфичность находилась в диапазоне от 0,806 до 0,846, что соответствует значениям, представленным в мультицентровых исследованиях (0,69–0,92) [26]. В то же время значение чувствительности при

оценке по бинарной шкале 2 в настоящем исследовании (0,834) оказалось выше, чем в предыдущей работе, выполненной в нашем учреждении в 2022 году (0,68 для лучшего ИИ-сервиса), что может свидетельствовать об улучшении возможностей ПО на основе ТИИ по распознаванию ЗНО молочной железы [9].

Также обращает на себя внимание тот факт, что даже в небольшой выборке исследований, отправленных на пересмотр эксперту, в 1 случае из 45 ПО поставило категорию BI-RADS4, которую подтвердил эксперт, тогда как врачом была поставлена категория BI-RADS1. Это указывает на перспективность использования ПО на основе ТИИ для оценки ММГ и требует проведения дальнейших исследований.

Ограничением данного исследования является отсутствие подтверждения результатов ММГ с помощью дополнительных методов визуализации, проведенной биопсии или информации о наличии интервального рака. Другим ограничением исследования можно считать небольшой размер выборки ММГ, которая была пересмотрена экспертом, а также превалированием в данной выборке исследований с расхождениями между врачом и ПО на основе ТИИ. Кроме того, в настоящем исследовании не была установлена первичность результата (оценивал ли врач маммографии раньше ПО или наоборот), в связи с чем фактор влияния решения на основе искусственного интеллекта на результат врача не учитывался.

Выводы:

1. Согласованность между ИИ-сервисом и врачом-рентгенологом при оценке маммографий, выполняемых в рамках скрининга, достигает 84,10%.
2. В большинстве случаев расхождений (до 97,34%) ИИ-алгоритм ставил более высокую категорию BI-RADS, по сравнению с врачом.
3. Наличие случая, при котором патология, определенная с помощью ПО и подтвержденная при пересмотре результатов экспертом, была пропущена врачом, говорит о перспективности использования ПО на основе ТИИ для оценки ММГ и требует проведения дальнейших исследований.

Финансирование: Исследование выполнено в рамках государственного задания «Научные методологии устойчивого развития технологий искусственного интеллекта в медицинской диагностике», регистрационный номер ЕГИСУ: 123031500004–5.





СПИСОК ИСТОЧНИКОВ



1. Каприна А.Д., Старинской В.В., Шахзадовой А.О. Злокачественные новообразования в России в 2021 году. Москва: МНИОИ им. П.А. Герцена – филиал ФГБУ «НМИЦ радиологии» Минздрава России; 2022. – 252 с.
2. Elmore J.G., Jackson S.L., Abraham L., Miglioretti D.L., Carney P.A., Geller B.M. et al. Variability in Interpretive Performance at Screening Mammography and Radiologists' Characteristics Associated with Accuracy. // Radiology. 2009 Dec;253(3):641–51.
3. Приказ Минздрава России (Министерство здравоохранения РФ) от 27 апреля 2021 г. № 404н «Об утверждении Порядка проведения профилактического медицинского осмотра и диспансеризации определенных групп взрослого населения».
4. Ren W., Chen M., Qiao Y., Zhao F. Global guidelines for breast cancer screening: A systematic review. // The Breast. 2022 Aug;64:85–99.
5. Pasyukov D.V., Egoshin I.A., Kolchev A.A., Kliouchkin I.V., Pasyukova O.O. The value of computer aided detection system in breast cancer difficult to detect at screening mammography. Rejr. 2019;9(2):107–18.
6. Васильев Ю.А., Тыров И.А., Владимирский А.В., Арзамасов К.М., Шулькин И.М., Кожихина Д.Д. и др. Двойной просмотр результатов маммографии с применением технологий искусственного интеллекта: новая модель организации массовых профилактических исследований // Digital Diagnostics. – 2023. – Т. 4. – № 2. DOI: 10.17816/DD321423
7. Leibig C., Brehmer M., Bunk S., Byng D., Pinker K., Umutlu L. Combining the strengths of radiologists and AI for breast cancer screening: a retrospective analysis. // Lancet Digit Health. 2022;4(7): e507-e519. DOI: 10.1016/S2589-7500(22)00070-X
8. McKinney S.M., Sieniek M., Godbole V. et al. International evaluation of an AI system for breast cancer screening [published correction appears in Nature. 2020 Oct;586(7829): E19]. Nature. 2020;577(7788):89–94. DOI: 10.1038/s41586-019-1799-6
9. Арзамасов К.М., Семенов С.С., Кирпичев Ю.С. и др. Оценка диагностической точности ИИ-алгоритмов для выявления злокачественных новообразований по данным маммографии // Медицинская физика. – 2022. – № 1(93). – С. 13–14.
10. Морозов С.П., Владимирский А.В., Шулькин И.М. и др. Целесообразность применения технологий искусственного интеллекта в лучевой диагностике (результаты первого года Московского Эксперимента по компьютерному зрению) // Врач и информационные технологии. – 2022. – № 1. – С. 12–29. – DOI: 10.25881/18110193_2022_1_12
11. Владимирский А.В., Васильев Ю.А., Арзамасов К.М. и др. Компьютерное зрение в лучевой диагностике: первый этап Московского эксперимента. – Москва: Общество с ограниченной ответственностью «Издательские решения», 2022. – 388 с. – ISBN 978-5-0059-3043-9.
12. Gusev A., Morozov S., Vladzimirskyy A. et al. Development of Artificial Intelligence in Healthcare in Russia // Intelligent Systems Reference Library. – 2022. – Vol. 212. – P. 259–279. – DOI: 10.1007/978-3-030-83620-7_11.
13. American College of Radiology. BI-RADS Breast imaging reporting and data system. Breast imaging atlas: mammography, breast ultrasound, breast MR-imaging. Virginia. Reston, 2003. – 268 p.
14. Морозов С.П., Ветшева Н.Н., Диденко В.В. и др. Организация программы популяционного скрининга злокачественных новообразований молочной железы среди женского населения: методические рекомендации // Серия «Лучшие практики лучевой и инструментальной диагностики». – Вып. 55 – М.: ГБУЗ «НПКЦ ДиТ ДЗМ», 2020. – 44 с.
15. Патент № 2697733 С1 Российская Федерация, МПК G06T 1/40, G06T 7/11, G06K 9/50. Система обработки рентгенографических изображений и вывода результата пользователю: № 2019118035: заявл. 10.06.2019: опубл. 19.08.2019 / И.И. Чернин, В.И. Чернин.
16. Павлович П.И., Броннов О.Ю., Капнинский А.А., Абович Ю.А., Рычагова Н.И. Сравнительное исследование результатов анализа данных цифровой маммографии системы на основе искусственного интеллекта «Цель» и врачей-рентгенологов // Digital Diagnostics. – 2021. – Т. 2. – № 2S. – С. 22–23. DOI: 10.17816/DD83184
17. Chang Sen L.Q., Mayo R.C., Lesslie M.D., Yang W.T., Leung J.W.T. Impact of Second-Opinion Interpretation of Breast Imaging Studies in Patients Not Currently Diagnosed With Breast Cancer. // J Am Coll Radiol. 2018;15(7):980–987.e1. DOI: 10.1016/j.jacr.2018.03.055
18. Coolen A.M.P., Lameijer J.R.C., Voogd A.C., et al. Characteristics of screen-detected cancers following concordant or discordant recalls at blinded double reading in biennial digital screening mammography. // Eur Radiol. 2019;29(1):337–344. DOI: 10.1007/s00330-018-5586-9
19. Морозов С.П., Андрейченко А.Е., Четвериков С.Ф. и др. Свидетельство о государственной регистрации программы для ЭВМ № 2022617324 Российская Федерация. Веб-инструмент для выполнения ROC анализа результатов диагностических тестов: № 2022616046: заявл. 05.04.2022: опубл. 19.04.2022; заявитель Государственное бюджетное учреждение здравоохранения города Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы».



20. Sickles E.A., D'Orsi C.J., Bassett L.W. et al. ACR BI-RADS® Mammography In: ACR BI-RADS® Atlas, Breast imaging reporting and data system. Reston, VA: American College of Radiology, 2013; P. 121–140
21. Moradi M., Ganji K., Teyfour N., Kolahdoozan F. Performance of double reading mammography in an Iranian population and its effect on patient outcome. // Iran J Radiol. 2013;10(2):51–55. DOI: 10.5812/iranjradiol.11729
22. Lehman C.D., Yala A., Schuster T. et al. Mammographic Breast Density Assessment Using Deep Learning: Clinical Implementation. // Radiology. 2019;290(1):52–58. DOI: 10.1148/radiol.2018180694
23. Flemban A.F. Overdiagnosis Due to Screening Mammography for Breast Cancer among Women Aged 40 Years and Over: A Systematic Review and Meta-Analysis. // Journal of Personalized Medicine. 2023; 13(3):523. <https://doi.org/10.3390/jpm13030523>
24. Rodríguez-Ruiz A., Krupinski E., Mordang J.J. et al. Detection of Breast Cancer with Mammography: Effect of an Artificial Intelligence Support System. // Radiology. 2019;290(2):305–314. DOI: 10.1148/radiol.2018181371
25. Arzamasov K., Vasilev Y., Vladzimirskyy A. et al. An International Non-Inferiority Study for the Benchmarking of AI for Routine Radiology Cases: Chest X-ray, Fluorography and Mammography. // Healthcare. 2023; 11(12):1684. <https://doi.org/10.3390/healthcare11121684>
26. Морозов С.П., Владзимирский А.В., Ледихова Н.В. и др. Московский эксперимент по применению компьютерного зрения в лучевой диагностике: вовлеченность врачей-рентгенологов // Врач и информационные технологии. – 2020. – № 4. – С. 14–23. DOI: 10.37690/1811-0193-2020-4-14-23.

ORIGINAL PAPER

THE FIRST 10,000 MAMMOGRAPHY EXAMS PERFORMED AS PART OF THE “DESCRIPTION AND INTERPRETATION OF MAMMOGRAPHY DATA USING ARTIFICIAL INTELLIGENCE” SERVICE

Y.A. Vasiliev^a, A.V. Vladzimirsky^b, K.M. Arzamasov^c, I.M. Shulkin^d, L.E. Aksenova^e✉, L.D. Pestrenin^f, S.S. Semenov^g, D.V. Bondarchuk^h, I.V. Smirnovⁱ

^{a, b, c, d, e, f, g, i} Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department, Moscow, Russia.

^a <https://orcid.org/0000-0002-0208-5218>;

^b <https://orcid.org/0000-0002-2990-7736>;

^c <https://orcid.org/0000-0001-7786-0349>;

^d <https://orcid.org/0000-0002-7613-5273>;

^e <https://orcid.org/0000-0003-0885-1355>;

^f <https://orcid.org/0000-0002-1786-4329>;

^g <https://orcid.org/0000-0003-2585-0864>;

^h <https://orcid.org/0000-0001-8752-0591>

✉ Corresponding author: Aksenova L.E.

ABSTRACT

Artificial intelligence technologies have great potential in improving the effectiveness of screening programs in the detection of malignant neoplasms of the breast. Given the high social, demographic and economic importance of mass preventive research, there is no doubt that the diagnostic accuracy of artificial intelligence must match or even exceed the accuracy of radiologists. In this regard, studies are needed to compare the accuracy of software based on artificial intelligence technology and radiologists during the mammography examinations in a clinical environment.

Purpose: to assess the quality of the medical service “Description and interpretation of mammography data using artificial intelligence” as part of screening.

Materials and methods. The sample for analysis consisted of 9684 digital mammograms. For each study, the BI-RADS category was determined by a radiologist and using software based on artificial intelligence technologies (AI based software) registered in the Russian Federation as a medical device. Forty-five studies from this sample with significant discrepancies in physician and software assessments were subject to peer review, which resulted in a BI-RADS category according to the physician expert.

Results. When evaluating weighted averages, there were no statistically significant differences between physician results and AI based software for 9684 digital mammography exams. Evaluation of physician and software consistency showed that matches are observed in 43,89% of cases for the BI-RADS scale and in 80,69% – 84,10% for binary scales. The presence of a case in which the pathology identified with the help of software and confirmed during the review of the results by the expert was missed by the doctor indicates the promise of using AI based software for evaluating mammography studies and requires further research.

Findings. When evaluating mammography studies, the agreement between the AI based decision and the radiologist reaches 84,10%, with the software assigning a higher BI-RADS category more often. Expert review of part of these discrepancies showed a potential reduction in the number of missed breast malignancies with the help of software.

Keywords: preventive research, mammography, artificial intelligence, diagnostic accuracy.

For citation: Vasiliev Y.A., Vladzimirsky A.V., Arzamasov K.M., Shulkin I.M., Aksenova L.E., Pestrenin L.D., Semenov S.S., Bondarchuk D.V., Smirnov I.V. The first 10,000 mammography exams performed as part of the “Description and interpretation of mammography data using artificial intelligence” service. *Manager Zdravookhraneniya*. 2023; 8:54–67. DOI: 10.21045/1811-0185-2023-8-54-67





REFERENCES

1. *Kaprina A.D., Starinskoy V.V., Shakhzadova A.O.* Malignant neoplasms in Russia in 2021. Moscow: MNIOL im. P.A. Herzen – branch of the Federal State Budgetary Institution “NMITS Radiology” of the Ministry of Health of Russia; 2022. – 252 p.
2. *Elmore J.G., Jackson S.L., Abraham L., Miglioretti D.L., Carney P.A., Geller B.M. et al.* Variability in Interpretive Performance at Screening Mammography and Radiologists’ Characteristics Associated with Accuracy. // *Radiology.* 2009 Dec;253(3):641–51.
3. Order of the Ministry of Health of Russia (Ministry of Health of the Russian Federation) dated April 27, 2021 No. 404n “On Approval of the Procedure for Conducting a Preventive Medical Examination and Medical Examination of Certain Groups of the Adult Population”.
4. *Ren W., Chen M., Qiao Y., Zhao F.* Global guidelines for breast cancer screening: A systematic review. // *The Breast.* 2022 Aug;64:85–99.
5. *Pasynkov D.V., Egoshin I.A., Kolchev A.A., Kliouchkin I.V., Pasynkova O.O.* The value of computer aided detection system in breast cancer difficult to detect at screening mammography. // *Rejr.* 2019;9(2):107–18.
6. *Vasilev Y., Tyrov I., Vladzimirskyy A., Arzamasov K., Shulkin I., Kozhikhina D. et al.* Double reading mammograms using artificial intelligence technologies: a new model of mass preventive examination organization // *Digital Diagnostics.* – 2023. – Vol. 4. – N. 2. DOI: 10.17816/DD321423
7. *Leibig C, Brehmer M, Bunk S, Byng D, Pinker K, Umutlu L.* Combining the strengths of radiologists and AI for breast cancer screening: a retrospective analysis. // *Lancet Digit Health.* 2022;4(7): e507-e519. DOI: 10.1016/S2589-7500(22)00070-X
8. *McKinney S.M., Steniek M., Godbole V. et al.* International evaluation of an AI system for breast cancer screening [published correction appears in *Nature.* 2020 Oct;586(7829): E19]. *Nature.* 2020;577(7788):89–94. DOI: 10.1038/s41586-019-1799-6
9. *Arzamasov K.M., Semenov S.S., Kirpichev Yu.S. et al.* Evaluation of the diagnostic accuracy of AI algorithms for the detection of malignant neoplasms according to mammography // *Medical Physics.* – 2022. – No. 1(93). – P. 13–14.
10. *Morozov S.P., Vladzimirskyy A.V., Shulkin I.M., Ledikhova N.V., Arzamasov K.M., Andreychenko A.E., et al.* Feasibility of using artificial intelligence in radiology (first year of Moscow Experiment on computer vision). // *Vrach i informacionnye tehnologii.* 2022;(1):12–29.
11. *Vladimirsky A.V., Vasilev Y.A., Arzamasov K.M. et al.* Computer vision in radiology: the first stage of the Moscow experiment. – Moscow: Limited Liability Company “Publishing Solutions”, 2022. – 388 p. – ISBN 978-5-0059-3043-9.
12. *Gusev A., Morozov S., Vladzimirskyy A. et al.* Development of Artificial Intelligence in Healthcare in Russia // *Intelligent Systems Reference Library.* – 2022. – Vol. 212. – P. 259–279. – DOI: 10.1007/978-3-030-83620-7_11.
13. American College of Radiology. BI-RADS Breast imaging reporting and data system. Breast imaging atlas: mammography, breast ultrasound, breast MR-imaging. Virginia. Reston, 2003. – 268 p.
14. *Morozov S.P., Vetsheva N.N., Didenko V.V. et al.* Organization of a population screening program for malignant neoplasms of the breast among the female population: methodological recommendations // *Series “Best practices of radiological and instrumental diagnostics”.* – Issue. 55 – M.: “Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department”, 2020. – 44 p.
15. Patent No. 2697733 C1 Russian Federation, IPC G06T 1/40, G06T 7/11, G06K 9/50. System for processing radiographic images and displaying the result to the user: No. 2019118035: Appl. 06/10/2019: publ. 19.08.2019 / I.I. Chernin, V.I. Chernin.
16. *Pavlovich P.I., Bronov O.Yu., Kapninsky A.A., Yu.* Comparative study of the results of digital mammography data analysis of a system based on artificial intelligence “Celsus” and radiologists // *Digital Diagnostics.* – 2021. – Vol. 2. – No. 2S. – C. 22–23. DOI: 10.17816/DD83184
17. *Chang Sen L.Q., Mayo R.C., Lesslie M.D., Yang W.T., Leung J.W.T.* Impact of Second-Opinion Interpretation of Breast Imaging Studies in Patients Not Currently Diagnosed With Breast Cancer. // *J Am Coll Radiol.* 2018;15(7):980–987.e1. DOI: 10.1016/j.jacr.2018.03.055
18. *Coolen A.M.P., Lameijer J.R.C., Voogd A.C. et al.* Characteristics of screen-detected cancers following concordant or discordant recalls at blinded double reading in biennial digital screening mammography. // *Eur Radiol.* 2019;29(1):337–344. DOI: 10.1007/s00330-018-5586-9
19. *Morozov S.P., Andreychenko A.E., Chetverikov S.F. et al.* Certificate of state registration of the computer program No. 2022617324 Russian Federation. Web tool for performing ROC analysis of diagnostic test results: No. 2022616046: App. 04/05/2022: publ. April 19, 2022; applicant State Budgetary Institution of Health of the City of Moscow “Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department”.
20. *Sickles E.A., D’Orsi C.J., Bassett L.W. et al.* ACR BI-RADS® Mammography In: ACR BI-RADS® Atlas, Breast imaging reporting and data system. Reston, VA: American College of Radiology, 2013. – P. 121–140.
21. *Moradi M., Ganji K., Teyfouri N., Kolahdoozan F.* Performance of double reading mammography in an Iranian population and its effect on patient outcome. // *Iran J Radiol.* 2013;10(2):51–55. DOI: 10.5812/iranjradiol.11729
22. *Lehman C.D., Yala A., Schuster T. et al.* Mammographic Breast Density Assessment Using Deep Learning: Clinical Implementation. // *Radiology.* 2019;290(1):52–58. DOI: 10.1148/radiol.2018180694



- 23.** *Flemban A.F.* Overdiagnosis Due to Screening Mammography for Breast Cancer among Women Aged 40 Years and Over: A Systematic Review and Meta-Analysis. // Journal of Personalized Medicine. 2023; 13(3):523. <https://doi.org/10.3390/jpm13030523>
- 24.** *Rodríguez-Ruiz A., Krupinski E., Mordang J.J. et al.* Detection of Breast Cancer with Mammography: Effect of an Artificial Intelligence Support System. // Radiology. 2019;290(2):305–314. DOI: 10.1148/radiol.2018181371
- 25.** *Arzamasov K., Vasilev Y., Vladzimirskyy A. et al.* An International Non-Inferiority Study for the Benchmarking of AI for Routine Radiology Cases: Chest X-ray, Fluorography and Mammography. // Healthcare. 2023; 11(12):1684. <https://doi.org/10.3390/healthcare11121684>
- 26.** *Morozov S.P., Vladzimirskii A.V., Ledikhova N.V. et al.* Moscow experiment on the use of computer vision in radiation diagnostics: the involvement of radiologists. // Vrach i informatsionnye tekhnologii. – 2020. – No. 4. – P. 14–23. DOI: 10.37690/1811-0193-2020-4-14-23

ИНФОРМАЦИЯ ОБ АВТОРАХ / ABOUT THE AUTHORS

Васильев Юрий Александрович – к.м.н., директор, Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы, г. Москва, Россия.

Yuri A. Vasilyev – PhD, Director, Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department, Moscow, Russia.

E-mail: VasilevYAI@zdrav.mos.ru

Владзимирский Антон Вячеславович – д.м.н., заместитель директора по научной работе, Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы, г. Москва, Россия.

Anton V. Vladimirsky – Doctor of Medical Sciences, Deputy Director for Research, Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department, Moscow, Russia.

E-mail: VladzimirskijAV@zdrav.mos.ru

Арзамасов Кирилл Михайлович – к.м.н., руководитель отдела медицинской информатики, радиомики и радиогеномики, Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы, г. Москва, Россия.

Kirill M. Arzamasov – PhD, Head of the Department of Medical Informatics, Radiomics and Radiogenomics, Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department, Moscow, Russia.

E-mail: ArzamasovKM@zdrav.mos.ru

Шулькин Игорь Михайлович – заместитель директора по перспективному развитию, Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы, г. Москва, Россия.

Igor M. Shulkin – Deputy Director for Prospective Development, Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department, Moscow, Russia.

E-mail: ShulkinIM@zdrav.mos.ru

Аксенова Любовь Евгеньевна – магистр, младший научный сотрудник, Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы, г. Москва, Россия.

Lyubov E. Aksenova – magister, junior researcher, Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department, Moscow, Russia.

E-mail: AksenovaLE@zdrav.mos.ru

Пестренин Лев Дмитриевич – младший научный сотрудник, Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы, г. Москва, Россия.

Lev D. Pestrenin – junior researcher, Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department, Moscow, Russia.

E-mail: PestreninLD@zdrav.mos.ru

Семенов Серафим Сергеевич – врач-рентгенолог, Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы, г. Москва, Россия.

Serafim S. Semenov – radiologist, Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department, Moscow, Russia.

E-mail: SemenovSS3@zdrav.mos.ru

Бондарчук Дмитрий Владимирович – заведующий отделом – врач-рентгенолог, Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы, г. Москва, Россия.

Dmitri V. Bondarchuk – head of the department – radiologist, Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department, Moscow, Russia.

E-mail: BondarchukDV@zdrav.mos.ru

Смирнов Иван Викторович – врач-рентгенолог, Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы, г. Москва, Россия.

Ivan V. Smirnov – radiologist, Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department, Moscow, Russia.

E-mail: SmirnovIV8@zdrav.mos.ru