



¹Научно-практический
клинический центр
диагностики
и телемедицинских
технологий, Москва

²ООО «ГУТА-
КЛИНИК», Москва

³ Российский
технологический
университет МИРЭА,
Москва

⁴ Самарский
государственный
медицинский
университет

Диагностическая точность ИИ-сервисов при оценке маммографических исследований по шкале плотности ACR.

Согласованность заключений ИИ-сервисов между собой и с мнением врача-эксперта

М.Ю. Хрустачева¹, А.С. Азарян^{1, 2}, Л.Д. Пестренин^{1, 3},
А.С. Доможирова¹, К.М. Арзамасов^{1, 3, 4}

Адрес для переписки: Маргарита Юрьевна Хрустачева, myukhrustacheva@mail.ru

Для цитирования: Хрустачева М.Ю., Азарян А.С., Пестренин Л.Д. и др. Диагностическая точность ИИ-сервисов при оценке маммографических исследований по шкале плотности ACR. Согласованность заключений ИИ-сервисов между собой и с мнением врача-эксперта. Эффективная фармакотерапия. 2025; 21 (47): 14–20..

DOI 10.33978/2307-3586-2025-21-47-14-20

Маммографическая плотность молочной железы – важный диагностический признак. При повышении ее плотности снижается диагностическая точность исследования, кроме того, с ростом плотности повышается риск рака молочной железы. Стоит сказать, что не только высокая плотность может говорить о возможных патологических изменениях. Так, низкая плотность может быть маркером повышенного риска сердечно-сосудистых заболеваний. На основании последних проведенных исследований по всему миру можно говорить о том, что пониженная плотность молочной железы указывает на вероятность повышенного риска артериальной гипертензии, ишемической болезни сердца, сердечной недостаточности, цереброваскулярной болезни, гиперхолестеринемии, а также сахарного диабета.

Цель – оценить согласованность заключений трех ИИ-сервисов: ИИ-сервис № 1, ИИ-сервис № 2, ИИ-сервис № 3 по шкале плотности ACR между собой и с мнением врача-эксперта, а также оценить метрики их диагностической точности.

Материал и методы. Проведено смешанное исследование, сочетающее ретроспективный диагностический анализ и аналитическое исследование. На основании данных ЕРИС ЕМИАС отобраны 99 анонимизированных маммографических исследований женщин старше 18 лет (средний возраст 56 лет) за период с 13.11.2020 по 04.10.2021, исключая исследования с артефактами на снимках. Каждое исследование независимо оценено врачом-экспертом (рентгенолог с опытом ≥ 5 лет и ученой степенью) и тремя ИИ-сервисами (ИИ-сервис № 1, ИИ-сервис № 2, ИИ-сервис № 3) с определением плотности молочной железы по шкале ACR BI-RADS. Основной целью исследования являлась оценка согласованности между мнением эксперта и заключениями ИИ-сервисов, которая анализировалась методом внутриклассовой корреляции Пирсона.

Результаты. В нашем исследовании параметры диагностической точности ИИ-сервисов, определяющих плотность молочной железы по бинарной шкале ACR, были следующими: ROC AUC – от 0,866 (0,787–0,947) до 0,904 (0,838–0,972), чувствительность – от 0,833 (0,700–0,967) до 0,867 (0,745–0,988), специфичность – от 0,899 (0,827–0,970) до 0,957 (0,908–1,000), точность – от 0,879 (0,814–0,943) до 0,919 (0,866–0,973).

При оценке диагностической точности для отдельных категорий плотности молочных желез по полной шкале ACR были получены следующие значения: ROC AUC – от 0,817 (0,722–0,917) до 0,995 (0,985–1,005), чувствительность – от 0,714 (0,521–0,908) до 1,000 (1,000–1,000), специфичность – от 0,784 (0,671–0,897) до 1,000 (1,000–1,000), точность – от 0,828 (0,754–0,903) до 0,990 (0,970–1,000).

Заключение. Все ИИ-сервисы показали высокую чувствительность (0,833–0,867) и специфичность (0,899–0,957), что является критически важным для скрининговых инструментов, целью которых является минимизация как ложноотрицательных, так и ложноположительных результатов.

Ключевые слова: маммография, искусственный интеллект, плотность молочной железы, маммографический скрининг, рак молочной железы



Введение

Маммография остается самым эффективным и широко используемым во всем мире методом скрининга на злокачественные новообразования (ЗНО) молочных желез, а также одним из наиболее эффективных методов снижения смертности от ЗНО молочной железы за счет ранней диагностики [1]. Несмотря на то что высокая плотность тканей молочной железы является менее значимым фактором риска, чем другие факторы, например наследственная предрасположенность, она встречается чаще [2]. При повышении плотности молочной железы снижается диагностическая точность маммографии [3–5], что обуславливает необходимость применения более дорогостоящих методов обследования, таких как МРТ молочных желез. Согласно исследованиям, МРТ признана наиболее эффективным методом дообследования при высокой плотности тканей [6]. Для улучшения ранней выявляемости ЗНО при высокой плотности эффективно дополнительное УЗИ, однако его внедрение ограничено дефицитом квалифицированных специалистов. Цифровой томосинтез, ставший стандартом диагностики, повышает выявляемость ЗНО, но его диагностическая точность также снижается при высокой плотности ткани железы [7]. В свою очередь, низкая плотность может быть маркером повышенного риска сердечно-сосудистых заболеваний (ССЗ). На основании последних исследований, проведенных по всему миру, можно говорить о том, что пониженная плотность молочной железы указывает на вероятность повышенного риска артериальной гипертензии, ишемической болезни сердца, сердечной недостаточности, цереброваскулярной болезни, гиперхолестеринемии, а также сахарного диабета [8].

При проведении диагностики лучевыми методами, как одними из самых массовых, здравоохранение сталкивается со множеством проблем: выгорание врачей, низкая доступность исследований для населения, недостаточный охват, а также дефицит кадров и финансовых ресурсов. Для решения этих вопросов продолжается активное внедрение ИИ-сервисов в лучевую диагностику [9], что обуславливает необходимость независимого определения этих параметров при внедрении ИИ-сервисов в практическое здравоохранение.

Цель

Оценить диагностическую точность и согласованность заключений трех ИИ-сервисов между собой и с мнением врача-эксперта при оценке маммографических исследований по шкале плотности ACR.

Материал и методы

Выполнено смешанное исследование (количественный компонент – ретроспективное диагностическое исследование, качественный компонент – аналитическое исследование).

Данное исследование основано на результатах Эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения

в системе здравоохранения г. Москвы, утвержденного этическим комитетом (выписка из протокола № 2 НЭК МРО POPP от 20 февраля 2020 г.), также зарегистрированного на ClinicalTrials (NCT04489992).

Набор данных

Из Единого радиологического информационного сервиса Единой медицинской информационно-аналитической системы Москвы (ЕРИС ЕМИАС) случайным образом были отобраны и выгружены 99 маммографических исследований, выполненных за период с 13.11.2020 по 04.10.2021.

Критерии включения: исследования выполнены пациентам женского пола в возрасте старше 18 лет.

Критерии исключения: маммограммы с артефактами на изображениях.

Средний возраст пациенток составил 56,0 года.

Перед выгрузкой все исследования были анонимизированы. Далее для каждого исследования врач-эксперт (рентгенолог с опытом работы не менее 5 лет и ученым степенью/званием по специальности «рентгенология») определил значение плотности по шкале ACR (табл. 1). Полная шкала ACR включает в себя следующие типы плотности:

- Молочная железа представлена практически полностью жировой тканью.
- Отдельные участки фиброгlandулярной (фиброзно-железистой) плотности.
- Гетерогенная плотность молочной железы, плотная фиброгlandулярная ткань может маскировать небольшие образования.
- Железа почти полностью состоит из фиброгlandулярной ткани, что значительно снижает чувствительность маммографии.

Для обеспечения сопоставимости наших результатов с результатами, представленными в научной литературе, мы также перевели полную шкалу ACR в бинарную (табл. 2).

ИИ-сервисы

Подготовленный нами набор данных, состоящий из 396 изображений (99 исследований), был обработан тремя ИИ-сервисами российских коммерческих компаний, которые далее по тексту будут представлены в анонимизированном виде: ИИ-сервис № 1, ИИ-сервис № 2,

Таблица 1. Распределение исследований по категориям плотности ACR на основе оценок врача-эксперта

Локализация	Тип А	Тип В	Тип С	Тип D
Правая	18 (18,2%)	51 (51,5%)	24 (24,2%)	6 (6,1%)
Левая	18 (18,2%)	50 (50,5%)	25 (25,2%)	6 (6,1%)

Таблица 2. Распределение исследований по категориям плотности по бинарной шкале плотности ACR на основе оценок врача-эксперта

Молочная железа	0 – категории плотности А, В	1 – категории плотности С, D
Правая	69 (69,7%)	30 (30,3%)
Левая	68 (68,7%)	31 (31,3%)



Таблица 3. Диагностическая точность ИИ-сервисов при оценке маммографических исследований по бинарной шкале ACR

	Молочная железа	ИИ-сервис № 1	ИИ-сервис № 2	ИИ-сервис № 3
ROC AUC	П	0,907 [0,837–0,979]	0,861 [0,777–0,947]	0,936 [0,881–0,992]
	Л	0,852 [0,767–0,936]	0,845 [0,762–0,928]	0,904 [0,838–0,972]
	П + Л	0,895 [0,821–0,969]	0,866 [0,787–0,947]	0,904 [0,838–0,972]
Чувствительность	П	0,857 [0,728–0,987]	0,821 [0,680–0,963]	0,929 [0,833–1,000]
	Л	0,733 [0,575–0,892]	0,733 [0,575–0,892]	0,867 [0,745–0,988]
	П + Л	0,833 [0,700–0,967]	0,833 [0,700–0,967]	0,867 [0,745–0,988]
Специфичность	П	0,958 [0,911–1,000]	0,901 [0,832–0,971]	0,944 [0,890–0,997]
	Л	0,971 [0,931–1,000]	0,957 [0,908–1,000]	0,942 [0,887–0,997]
	П + Л	0,957 [0,908–1,000]	0,899 [0,827–0,970]	0,942 [0,887–0,997]
Точность	П	0,929 [0,879–0,980]	0,879 [0,814–0,943]	0,939 [0,892–0,986]
	Л	0,899 [0,840–0,958]	0,889 [0,827–0,951]	0,919 [0,866–0,973]
	П + Л	0,919 [0,866–0,973]	0,879 [0,814–0,943]	0,919 [0,866–0,973]

Примечание. В квадратных скобках представлены значения 95%-ных доверительных интервалов (ДИ); П – правая молочная железа; Л – левая молочная железа; П + Л – максимальное значение по шкале ACR для правой и левой молочных желез у одной пациентки.

ИИ-сервис № 3. ИИ-сервисы определили значение плотности по шкале ACR.

Статистический анализ

Для оценки диагностической точности ИИ-сервисов были рассчитаны следующие параметры с 95%-ными доверительными интервалами (ДИ): ROC AUC, чувствительность, специфичность, точность. В качестве истинных значений (Ground Truth) использовались оценки врача-эксперта. Расчет параметров был выполнен с помощью веб-инструмента, разработанного ГБУЗ «Научно-практический клинический центр диагностики и телемедицинских технологий» (НПКЦ ДиТ) ДЗМ (<https://roc-analysis.mosmed.ai>). Параметры диагностической точности рассчитывались для бинарной шкалы ACR.

Подготовленный набор данных состоял из 99 маммографических исследований. С обоснованием объема выборки можно ознакомиться в наших предыдущих работах [10].

Для количественной оценки степени согласованности между заключениями различных ИИ-сервисов и мнением врача-эксперта при определении категорий плотности молочной железы по шкале ACR был применен метод внутриклассовой корреляции (Intraclass Correlation Coefficient, ICC) Пирсона. Данный метод представляет собой мощный статистический инструмент, позволяющий оценить не только степень соответствия между измерениями, но и их согласованность в условиях, когда оценивается один и тот же параметр разными методами или наблюдателями. В нашем исследовании использовалась двухсторонняя модель ICC с доверительными интервалами 95%, что соответствует рекомендациям по оценке согласованности между различными методами измерений. Значения ICC интерпретировались согласно классификации Landis и Koch: 0,81–1,00 – практически полная согласованность; 0,61–0,80 – существенная согласованность; 0,41–0,60 – умеренная согласованность; 0,21–0,4 – слабая согласованность; 0–0,20 – очень слабая согласованность.

Результаты

В таблице 3 представлены значения ROC AUC, чувствительности, специфичности и точности с 95%-ными доверительными интервалами для трех ИИ-сервисов при бинарной оценке маммографических исследований.

Анализ данных, приведенных в табл. 3, демонстрирует высокую и сопоставимую общую диагностическую точность всех трех сервисов. Так, значения ROC AUC находятся в диапазоне от 0,866 (ИИ-сервис № 2) до 0,904 (ИИ-сервис № 3). Все ИИ-сервисы показали высокую чувствительность (0,833–0,867) и специфичность (0,899–0,957), что является критически важным для скрининговых инструментов, целью которых является минимизация как ложноотрицательных, так и ложноположительных результатов. Наибольшую специфичность (0,957), а следовательно, и наименьшую частоту ложноположительных результатов продемонстрировал ИИ-сервис № 1 – общая точность алгоритмов находится в диапазоне от 0,879 до 0,919. В таблице 4 представлены значения параметров диагностической точности ИИ-сервисов для отдельных типов рентгенологической плотности ткани молочной железы (ACR A–D).

Результаты демонстрируют существенную вариабельность диагностической эффективности ИИ-алгоритмов в зависимости от типа плотности ткани. ROC-AUC сохраняет стабильно высокие значения (более 0,8) для всех сервисов во всех категориях плотности, что подтверждает их общую надежность в дифференциальной диагностике. Особого внимания заслуживает прогрессирование метрик в категориях с высокой плотностью (C и D), где ИИ-сервис № 3 достигает исключительных показателей с ROC-AUC 0,995 [0,985–1,005] для категории D. Однако анализ чувствительности выявляет разнонаправленные тенденции между сервисами. ИИ-сервис № 1 показывает наименьшую чувствительность в категориях ACR A (0,714) и ACR C (0,731), но демонстрирует один из лучших результатов для категории ACR D (0,750). ИИ-сервис № 2 сохраняет относитель-



Таблица 4. Диагностическая точность ИИ-сервисов для отдельных типов рентгенологической плотности ткани молочной железы по полной шкале BI-RADS

	Молочная железа	ACR A			ACR B			ACR C			ACR D		
		ИИ-сервис № 1	ИИ-сервис № 2	ИИ-сервис № 3	ИИ-сервис № 1	ИИ-сервис № 2	ИИ-сервис № 3	ИИ-сервис № 1	ИИ-сервис № 2	ИИ-сервис № 3	ИИ-сервис № 1	ИИ-сервис № 2	ИИ-сервис № 3
ROC AUC	П	0,849 [0,755– 0,942]	0,898 [0,813– 0,986]	0,851 [0,751– 0,952]	0,828 [0,750– 0,905]	0,828 [0,751– 0,908]	0,889 [0,827– 0,952]	0,848 [0,757– 0,940]	0,828 [0,730– 0,928]	0,938 [0,877– 1,001]	0,864 [0,622– 1,113]	0,995 [0,985– 1,005]	0,995 [0,985– 1,005]
	Л	0,927 [0,863– 0,994]	0,903 [0,892– 0,982]	0,851 [0,751– 0,952]	0,839 [0,762– 0,916]	0,819 [0,746– 0,895]	0,870 [0,805– 0,937]	0,826 [0,731– 0,920]	0,787 [0,690– 0,888]	0,903 [0,829– 0,978]	0,875 [0,633– 1,123]	0,989 [0,975– 1,004]	0,995 [0,985– 1,005]
	П + Л	0,851 [0,751– 0,952]	0,881 [0,789– 0,976]	0,851 [0,751– 0,952]	0,850 [0,778– 0,924]	0,829 [0,755– 0,908]	0,870 [0,805– 0,937]	0,838 [0,748– 0,929]	0,817 [0,722– 0,917]	0,903 [0,829– 0,978]	0,864 [0,622– 1,113]	0,989 [0,975– 1,004]	0,995 [0,985– 1,005]
Чувствительность	П	0,762 [0,580– 0,944]	0,810 [0,642– 0,977]	0,714 [0,521– 0,908]	0,840 [0,738– 0,942]	0,840 [0,738– 0,942]	0,920 [0,845– 0,995]	0,750 [0,577– 0,923]	0,750 [0,577– 0,923]	0,917 [0,806– 1,000]	0,750 [0,326– 1,000]	1,000 [1,000– 1,000]	1,000 [1,000– 1,000]
	Л	0,905 [0,779– 1,000]	0,857 [0,707– 1,000]	0,714 [0,521– 0,908]	0,875 [0,781– 0,969]	0,854 [0,954– 0,954]	0,917 [0,838– 0,995]	0,692 [0,515– 0,870]	0,615 [0,428– 0,802]	0,846 [0,707– 0,985]	0,750 [0,326– 1,000]	1,000 [1,000– 1,000]	1,000 [1,000– 1,000]
	П + Л	0,714 [0,521– 0,908]	0,762 [0,580– 0,944]	0,714 [0,521– 0,908]	0,917 [0,838– 0,995]	0,854 [0,754– 0,954]	0,917 [0,838– 0,995]	0,731 [0,560– 0,901]	0,731 [0,560– 0,901]	0,846 [0,707– 0,985]	0,750 [0,326– 1,000]	1,000 [1,000– 1,000]	1,000 [1,000– 1,000]
Специфичность	П	0,936 [0,882– 0,990]	0,987 [0,962– 1,000]	0,987 [0,962– 1,000]	0,816 [0,708– 0,925]	0,816 [0,708– 0,925]	0,857 [0,759– 0,955]	0,947 [0,896– 0,998]	0,907 [0,841– 0,973]	0,917 [0,806– 1,000]	0,979 [0,950– 1,000]	0,989 [0,969– 1,000]	0,989 [0,969– 1,000]
	Л	0,949 [0,900– 0,998]	0,949 [0,900– 0,998]	0,987 [0,962– 1,000]	0,804708 [0,695– 0,913]	0,784 [0,671– 0,897]	0,824 [0,719– 0,928]	0,959 [0,913– 1,000]	0,959 [0,913– 1,000]	0,959 [0,913– 1,000]	1,000 [1,000– 1,000]	0,979 [0,950– 1,000]	0,989 [0,969– 1,000]
	П + Л	0,987 [0,962– 1,000]	1,000 [1,000– 1,000]	0,987 [0,962– 1,000]	0,784 [0,671– 0,897]	0,804 [0,695– 0,913]	0,824 [0,719– 0,928]	0,945 [0,893– 0,997]	0,904 [0,837– 0,972]	0,959 [0,913– 1,000]	0,979 [0,950– 1,000]	0,979 [0,950– 1,000]	0,989 [0,969– 1,000]
Точность	П	0,899 [0,840– 0,958]	0,949 [0,906– 0,993]	0,929 [0,879– 0,980]	0,828 [0,754– 0,903]	0,828 [0,754– 0,903]	0,889 [0,827– 0,951]	0,899 [0,840– 0,958]	0,869 [0,802– 0,935]	0,949 [0,906– 0,993]	0,970 [0,936– 1,000]	0,990 [0,970– 1,000]	0,990 [0,970– 1,000]
	Л	0,939 [0,892– 0,968]	0,929 [0,897– 0,980]	0,929 [0,879– 0,980]	0,838 [0,766– 0,911]	0,818 [0,742– 0,894]	0,869 [0,802– 0,935]	0,889 [0,827– 0,951]	0,869 [0,802– 0,935]	0,929 [0,879– 0,980]	0,990 [0,970– 1,000]	0,980 [0,952– 1,000]	0,990 [0,970– 1,000]
	П + Л	0,929 [0,879– 0,980]	0,949 [0,906– 0,993]	0,929 [0,879– 0,980]	0,848 [0,778– 0,919]	0,828 [0,754– 0,903]	0,869 [0,802– 0,935]	0,889 [0,827– 0,951]	0,859 [0,790– 0,927]	0,929 [0,879– 0,980]	0,970 [0,936– 1,000]	0,980 [0,952– 1,000]	0,990 [0,970– 1,000]

Примечание. В квадратных скобках представлены значения 95%-ных доверительных интервалов (ДИ); П – правая молочная железа; Л – левая молочная железа; П + Л – максимальное значение по шкале ACR для правой и левой молочных желез у одной пациентки.

ную стабильность показателей в категориях ACR A–C (0,854–0,731) с пиковым значением в категории ACR D (0,762). Наиболее впечатляющие результаты демонстрирует ИИ-сервис № 3, поддерживающий стабильно высокую чувствительность во всех категориях и достигающий идеального значения 1,000 в категориях ACR C и ACR D. Показатель специфичности последовательно улучшается от категории ACR A к ACR D для всех сервисов: ИИ-сервис № 1 прогрессирует от 0,979 (D) до 0,987 (A), ИИ-сервис № 2 – от 0,804 (B) до 1,000 (D), а ИИ-сервис № 3 – от 0,824 (B) до 0,989 (C) и 0,979 (D). Данная динамика свидетельствует о повышенной точности алгоритмов в работе с плотными тканями, где традиционные методы диагностики сталкиваются

с наибольшими сложностями. В категориях с низкой плотностью (ACR A и ACR B) сервисы показывают относительно стабильные показатели ROC AUC в диапазоне 0,829–0,903, при этом специфичность остается высокой для категории ACR A (0,987 у ИИ-сервиса № 1), но заметно снижается для ACR B (0,784–0,804). Для высокоплотных категорий (ACR C и ACR D) наблюдаются различные показания: в категории С некоторые сервисы демонстрируют улучшение показателей (ROC AUC 0,903 – у ИИ-сервиса № 3), а в категории ACR D достигают исключительно высоких значений (AUC 0,995 – у ИИ-сервис № 3). Чувствительность достигает 100% у ИИ-сервис № 3 для категории ACR D, однако эти показатели сопровождаются широкими



Таблица 5. Согласованность между ИИ-сервисами по правой и по левой молочным железам

Молочная железа	ИИ-сервисы	Внутрикласовая корреляция
Правая	ИИ-сервис № 1 и ИИ-сервис № 3	0,751 [0,651–0,826]
	ИИ-сервис № 1 и ИИ-сервис № 3	0,820 [0,724–0,881]
	ИИ-сервис № 1 и ИИ-сервис № 2	0,870 [0,813–0,911]
Левая	ИИ-сервис № 1 и ИИ-сервис № 2	0,843 [0,776–0,892]
	ИИ-сервис № 2 и ИИ-сервис № 3	0,797 [0,712–0,859]
	ИИ-сервис № 2 и ИИ-сервис № 3	0,806 [0,722–0,866]

Таблица 6. Согласованность между заключениями ИИ-сервисов и мнением врача-эксперта по правой и по левой молочным железам

Молочная железа	Сервисы	Внутрикласовая корреляция
Правая	ИИ-сервис № 1 и врач	0,837 [0,766–0,887]
	ИИ-сервис № 2 и врач	0,852 [0,787–0,898]
	ИИ-сервис № 3 и врач	0,848 [0,782–0,896]
Левая	ИИ-сервис № 1 и врач	0,857 [0,792–0,902]
	ИИ-сервис № 2 и врач	0,842 [0,774–0,891]
	ИИ-сервис № 3 и врач	0,833 [0,762–0,885]

доверительными интервалами, что указывает на необходимость осторожной интерпретации. Специфичность последовательно возрастает к категории ACR D, достигая значений 0,979–1,000. Полученные данные подчеркивают критическую важность учета плотности ткани молочной железы при валидации и клиническом внедрении ИИ-решений, поскольку их диагностическая эффективность существенно зависит от данного фактора.

В таблице 5 отражены показатели внутрикласовой корреляции, демонстрирующие уровень согласованности оценок между различными парами ИИ-сервисов для правой и левой молочных желез. Статистический анализ выявил высокую степень согласованности между заключениями ИИ-сервисов в оценке маммографической плотности. Наибольшая согласованность наблюдалась между ИИ-сервисом № 1 и ИИ-сервисом № 2, с ICC = 0,87 для правой молочной железы и 0,843 – для левой, что свидетельствует о практически полном соответствии их оценок. Сравнение согласованности ИИ-сервиса № 1 с ИИ-сервисом № 3 показало ICC в диапазоне 0,751–0,82, а между ИИ-сервисом № 2 и ИИ-сервисом № 3 – 0,797–0,806, что соответствует уровню существенной согласованности. В таблице 6 представлены значения внутрикласовой корреляции между заключениями трех ИИ-сервисов и мнением врача-эксперта при оценке плотности ткани правой и левой молочных желез. При сопоставлении ИИ-сервисов с экспертными оценками все системы продемонстрировали высокие параметры диагностической точности. Наилучшие результаты показали ИИ-сервис № 1 (ICC = 0,857) и ИИ-сервис № 2 (ICC = 0,852) при анализе левой молочной железы, в то время как ИИ-сервис № 3 достиг значений 0,833–0,848. Особого внимания заслуживает анализ доверительных интервалов, которые во всех случаях находились в зоне существенной или практически полной согласованности. Например, для пары

ИИ-сервис № 1 и ИИ-сервис № 2 при оценке правой молочной железы 95%-ный ДИ составил 0,813–0,911, что исключает возможность умеренной или слабой согласованности и подтверждает высокую воспроизводимость результатов.

Обсуждение

В нашем исследовании параметры диагностической точности ИИ-сервисов, определяющих плотность молочной железы по бинарной шкале ACR, были следующими: ROC AUC – 0,866–0,904, чувствительность – 0,833–0,867, специфичность – 0,899–0,957, точность – 0,879–0,919.

При оценке диагностической точности для отдельных категорий плотности молочных желез по полной шкале ACR были получены следующие значения: ROC AUC – 0,817–0,995, чувствительность – 0,714–1,000, специфичность – 0,784–1,000, точность – 0,828–0,990. В другой работе Н. Ji и соавт. (2024) согласованность заключений трех коммерческих ИИ-сервисов и мнения врача-рентгенолога при оценке плотности молочной железы находилась в диапазоне от 0,500 до 0,673, что соответствует уровню от слабой до умеренной. При этом согласованность между самими ИИ-сервисами варьировалась в еще более широком диапазоне – от 0,486 до 0,684. Наибольшее расхождение наблюдалось у одного из сервисов (ИИ-сервис № 3), который демонстрировал низкую согласованность оценки как с врачом-экспертом (0,500), так и с другими алгоритмами (0,486–0,494) [11].

В работе B. Rigaud и соавт. (2022) продемонстрировали согласованность оценок ИИ-сервиса и врача-эксперта на уровне 0,33 по полной шкале плотности ACR и 0,54 для бинарной классификации, что соответствует удовлетворительной и умеренной степени согласованности соответственно. Авторы предложили использовать другие, предварительно обученные модели глубокого обучения, которые продемонстрировали более высокую



согласованность с экспертами: значения в диапазонах 0,61–0,66 – для полной шкалы плотности и 0,70–0,75 – для бинарной классификации указывают на умеренную и существенную степень согласованности [12].

Отдельного внимания заслуживает диагностическая точность рентгенологов при оценке плотности молочных желез. Согласно исследованию M. Biros и соавт. (2024), согласованность оценок ИИ-сервиса и врача-рентгенолога составила 0,708 (95% ДИ 0,562–0,841), что свидетельствует о существенной согласованности, точность заключений ИИ-сервиса составила 0,819 (95% ДИ от 0,736 до 0,903) [13].

Проводя параллель между исследованиями, можно говорить о результатах работы российского ПО на основе технологий искусственного интеллекта не только как о сопоставимых с зарубежными аналогами, но и превосходящих их. Наиболее значимое преимущество наблюдается в категориях с высокой плотностью тканей молочной железы (ACR C и ACR D), где российские системы показывают близкие к идеальным значения чувствительности и специфичности. Преимущество в точности обусловлено комплексной методологией разработки, сочетающей масштабные данные и глубокую интеграцию экспертных знаний. Стоит отметить, что российские разработчики при обучении моделей ИИ активно привлекают врачей-рентгенологов, в то время как зарубежные компании зачастую используют информацию из открытых источников [14,15].

Полученные данные убедительно свидетельствуют о том, что современные ИИ-алгоритмы способны воспроизводить экспертные оценки маммографической плотности с точностью, соответствующей критериям «практически полной согласованности». Важно отметить, что различия между показателями ICC для правой и левой молочных желез (в среднем 0,01–0,02) не достигли статистической значимости ($p > 0,05$), что подтверждает стабильность работы алгоритмов независимо от стороны исследования и отсутствие систематической ошибки, связанной с локализацией. Эти результаты имеют важное клиническое значение, так как демонстрируют потенциал ИИ-технологий для стандартизации оценки маммографической плотности в рутинной практике.

Заключение

Проведенное исследование продемонстрировало высокую диагностическую согласованность заключений российских ИИ-сервисов как между собой, так и с экспертной оценкой врача-рентгенолога при определении маммографической плотности по шкале ACR. Полученные значения внутриклассовой корреляции (ICC 0,751–0,87 между ИИ-сервисами и 0,833–0,857 при сравнении с экспертом) соответствуют уровням существенной и практически полной согласованности согласно критериям Landis и Koch, что не только сопоставимо с зарубежными аналогичными исследованиями, но и превосходит их результаты. Наблюдается высокая согласованность оценок плотности как между ИИ-сервисами, так и между ИИ-сервисами и врачом-экспертом. Параметры диагностической точности по бинарной шкале ACR также находятся на высоком уровне. По полной шкале чувствительность ИИ-сервисов неодинакова к разным категориям: ИИ-сервис № 1 определяет лучше категорию ACR B, ИИ-сервис № 3 – ACR D, ИИ-сервис № 2 – ACR D. Но даже при этом уровне чувствительности – нормальные, а значения специфичности и ROC AUC – высокие, что в совокупности позволяет использовать данные ИИ-сервисы в клинической практике. Их использование может способствовать стандартизации оценки плотности молочной железы. Таким образом, современные российские ИИ-сервисы, использующиеся для оценки маммографической плотности, не только соответствуют эксперциальному уровню, но и демонстрируют сопоставимость результатов с зарубежными аналогичными исследованиями. ☀

Финансирование.

*Статья подготовлена авторским коллективом
в рамках НИОКР «Программные
решения на основе технологий искусственного
интеллекта для практического здравоохранения».*

Благодарности.

*Авторы выражают благодарность компаниям-
разработчикам АО «МЕДИЦИНСКИЕ ТЕХНОЛОГИИ
Лтд», ООО «Медицинские Скрининг Системы»,
ООО «Платформа Третье Мнение»
за предоставление своих решений для тестирования
в рамках настоящей публикации.*

Литература

- Edmonds C.E., O'Brien S.R., Conant E.F. Mammographic breast density: current assessment methods, clinical implications, and future directions. *Semin. Ultrasound CT MR.* 2023; 44 (1): 35–45.
- Lee C.I., Chen L.E., Elmore J.G. Risk-based breast cancer screening: implications of breast density. *Med. Clin. North. Am.* 2017; 101 (4): 725–741.
- Лабазанова П.Г., Будanova М.В., Бурдина И.И. и др. Маммографическая плотность – маркер повышенного риска развития рака молочной железы. *Медицинский алфавит.* 2021; 19: 41–48.
- Голотюк М.А., Бережной А.А., Казанцева Н.В. и др. Клинико-диагностическая характеристика BRCA-ассоциированного рака молочной железы. *Онкология. Журнал им. П.А. Герцена.* 2022; 11 (6): 18–25.
- Арзамасов К.М., Васильев Ю.А., Владзимирский А.В. и др. Применение компьютерного зрения для профилактических исследований на примере маммографии. *Профилактическая медицина.* 2023; 26 (6): 117–123.
- Hussein H. Comparative effectiveness of imaging modalities for dense breast tissue assessment. *J. Med. Imaging.* 2023; 10 (2): 45–52.



7. Payne N.R., Hickman S.E., Black R., et al. Breast density effect on the sensitivity of digital screening mammography in a UK cohort. *Eur. Radiol.* 2025; 35 (1): 177–187.
8. Бочкарева Е.В., Рожкова Н.И., Бутина Е.К. и др. Маммографическая плотность молочной железы и сердечно-сосудистые заболевания у женщин. Обзор литературы. Кардиоваскулярная терапия и профилактика. 2024; 23 (8): 4064.
9. Васильев Ю.А., Тыров И.А., Владзимирский А.В. и др. Автономный искусственный интеллект для сортировки результатов профилактических лучевых исследований. Профилактическая медицина. 2024; 27 (7): 23–29.
10. Четвериков С.Ф., Арзамасов К.М., Андрейченко А.Е. Подходы к формированию выборки для контроля качества работы систем искусственного интеллекта в медико-биологических исследованиях. Современные технологии в медицине. 2023; 2: 19–27.
11. Ji H., Jang M.J., Chang J.M. Variability in breast density estimation and its impact on breast cancer risk assessment. *J. Breast Cancer.* 2024; 27 (5): 334–342.
12. Rigaud B., Weaver O.O., Dennison J.B., et al. Deep learning models for automated assessment of breast density using multiple mammographic image types. *Cancers (Basel).* 2022; 14 (20): 5003.
13. Biroš M., Kvák D., Dandár J., et al. Enhancing accuracy in breast density assessment using deep learning: a multicentric, multi-reader study. *Diagnostics (Basel).* 2024; 14 (11): 1117.
14. Рожкова Н.И., Ройтберг П.Г., Варфоломеева А.А. и др. Сегментационная модель скрининга рака молочной железы на основе нейросетевого анализа рентгеновских изображений. Сеченовский вестник. 2020; 11 (3): 4–14.
15. Васильев Ю.А. и Владзимирский А.В. Искусственный интеллект в лучевой диагностике: Per Aspera ad Astra. M.: Издательские решения, 2025.

Diagnostic Accuracy of AI Services in the Evaluation of Mammographic Studies on the ACR Density Scale. Consistency of the Conclusions of AI Services among Themselves and with the Opinion of an Expert Doctor

M.Yu. Khrustacheva¹, A.S. Azaryan^{1, 2}, L.D. Pestrenin^{1, 3}, A.S. Domozhirova¹, K.M. Arzamasov^{1, 3, 4}

¹ Scientific and Practical Clinical Center for Diagnostics and Telemedicine Technologies, Moscow

² "GUTA-CLINIC" LLC, Moscow

³ MIREA – Russian Technological University, Moscow

⁴ Samara State Medical University

Contact person: Margarita Yu. Khrustacheva, myukhrustacheva@mail.ru

Mammographic breast density is an important diagnostic indicator. An increase in breast density reduces diagnostic accuracy; moreover, as density increases, the risk of breast cancer also rises.

It is worth noting that not only high density can indicate possible pathological changes. Low density can be a marker of increased risk of cardiovascular diseases (CVD). Based on recent studies conducted worldwide, there is evidence of an increased risk of arterial hypertension, ischemic heart disease, heart failure, cerebrovascular disease, hypercholesterolemia, as well as diabetes mellitus.

Aim of the research: to assess the agreement among three AI services: AI-service No 1, AI-service No 2, AI-service No 3 according to the ACR density scale.

Material and methods. A mixed study was conducted, combining retrospective diagnostic analysis and analytical research. Based on data from ERIS EMIAS, 99 anonymized mammographic studies of women over 18 years old (mean age 56 years) were selected for the period from 13.11.2020 to 04.10.2021, excluding studies with artifacts on the images. Each study was independently evaluated by an expert physician (a radiologist with ≥ 5 years of experience and an academic degree) and three AI services (AI-service No 1, AI-service No 2, AI-service No 3) with determination of breast density according to the ACR BI-RADS scale. The main objective of the study was to assess the agreement between the expert and AI services, which was analyzed using the intraclass correlation method (Pearson).

Results. In our study, the diagnostic accuracy parameters of AI services determining breast density according to the binary ACR scale were as follows: ROC AUC – 0.866–0.904, sensitivity – 0.833–0.867, specificity – 0.899–0.957, accuracy – 0.879–0.919. When assessing diagnostic accuracy for individual breast density categories according to the full ACR scale, the following values were obtained: ROC AUC – 0.817–0.995, sensitivity – 0.714–1.000, specificity – 0.784–1.000, accuracy – 0.828–0.990.

Conclusion. All AI-services demonstrated high sensitivity (0.833–0.867) and specificity (0.899–0.957), which is critically important for screening tools aimed at minimizing both false-negative and false-positive results.

Keywords: mammography, artificial intelligence, breast density, mass screening, breast cancer