ISSN 1607-0763 (Print); ISSN 2408-9516 (Online)
https://doi.org/10.24835/1607-0763-1508

Применение ROC-анализа для оценки результатов работы систем поддержки принятия врачебных решений на примере цифровых маммографических изображений

© Хрустачева М.Ю.[1, 2*], Васильев Ю.А.[1], Памова А.П.[1], Арзамасов К.М.[1, 3]

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

[2] ГБУЗ "Московский многопрофильный клинический центр «Коммунарка»" ДЗ города Москвы; 108814 Москва, п. Коммунарка, ул. Сосенский стан, д. 8, Российская Федерация

[3] ФГБОУ ВО "МИРЭА – Российский технологический университет" (РТУ МИРЭА); 119454 Москва, просп. Вернадского, д. 78, Российская Федерация

В настоящее время идет активное внедрение искусственного интеллекта (ИИ), а также систем поддержки принятия врачебных решений (СППВР) в здравоохранение. Лучевая диагностика занимает лидирующие позиции по использованию подобных технологий. В представленной работе описан метод оценки эффективности работы СППВР, в том числе программного обеспечения (ПО) на основе технологий искусственного интеллекта (ПО на основе ТИИ), подходящий любой медицинской организации, перед которой стоит задача оценки применимости подобного ПО.

Цель исследования: наглядно продемонстрировать применение веб-инструмента для ROC-анализа для оценки результатов работы СППВР на примере цифровых маммографических изображений.

Материал и методы. Был использован ретроспективный набор данных маммографических исследований с результатами отчета калибровочного тестирования при смене версионности одного из сервисов ИИ, участника Московского эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения этих технологий в системе здравоохранения. Версии ИИ-сервиса от 15.02.2023 и 30.05.2023. Объем выборки – 100 исследований. В данной публикации для оценки результатов работы ИИ-сервиса использован ROC-анализ, который был реализован с помощью веб-инструмента для ROC-анализа.

Результаты. Продемонстрирована работа веб-инструмента для ROC-анализа на примере оценки результатов работы ПО на основе ТИИ для обработки цифровых маммографических изображений.

Заключение. Благодаря использованию представленного веб-инструмента для ROC-анализа при необходимости может быть реализована проверка любой СППВР, в том числе ПО на основе ТИИ, а также оценка его производительности без применения дополнительных инструментов.

Ключевые слова: лучевая диагностика; искусственный интеллект; маммография; ROC-кривая; чувствительность; специфичность; системы поддержки принятия врачебных решений; AUC

Авторы подтверждают отсутствие конфликтов интересов.

Данная статья подготовлена авторским коллективом в рамках НИОКР "Разработка платформы повышения качества ИИ-сервисов для медицинской диагностики" (№ ЕГИСУ: 123031400006-0)

Для цитирования: Хрустачева М.Ю., Васильев Ю.А., Памова А.П., Арзамасов К.М. Применение ROC-анализа для оценки результатов работы систем поддержки принятия врачебных решений на примере цифровых маммографических изображений. *Медицинская визуализация.* https://doi.org/10.24835/1607-0763-1508

Поступила в редакцию: 23.10.2024. **Принята к печати:** 11.03.2025. **Опубликована online:** 18.04.2025.

ROC analysis for clinical decision support systems (CDSS) results in digital mammography images

© **Margarita Yu. Khrustacheva[1, 2*], Yuri A. Vasiliev[1], Anastasia P. Pamova[1], Kirill M. Arzamasov[1, 3]**

[1] Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department; 24/1, Petrovka str, Moscow 127051, Russian Federation

[2] Moscow Clinical Center "Kommunarka" of Moscow Healthcare Department; 8, Sosensky stan str., Kommunarka settl., Moscow 108814, Russian Federation

[3] MIREA – Russian Technological University (RTU MIREA); 78, Vernadsky avenue, Moscow 119454, Russian Federation

Artificial intelligence and clinical decision support systems (CDSS) are being actively implemented in healthcare. Radiology is at the forefront of the use of such technologies. In this article, we describe a method for evaluating the performance of CDSS, including software based on artificial intelligence technologies (AI-based software), which is suitable for any medical organization that needs to assess the applicability of such software.

Purpose. The purpose of this study is to demonstrate the use of a web-based ROC analysis tool for evaluating the performance of clinical decision support systems (CDSS) using digital mammography images as an example.

Materials and methods. A retrospective dataset of mammography studies was used, based on the results of the calibration test report during the version change of one of the AI service participating in the Experiment on the use of innovative computer vision technologies for medical image analysis and subsequent applicability in the healthcare system of Moscow, with versions of the AI service dated 15.02.2023 and 30.05.2023. The sample size consisted of 100 trials. In this publication, ROC analysis implemented using a web-based tool will be used to evaluate the results of the AI service.

Results. The functionality of a web-based tool for ROC analysis was demonstrated using the example of evaluating the performance of AI-based software for processing digital mammography images.

Conclusion. By using the presented web-based ROC analysis tool, the verification of CDSS, including AI-based software, as well as the assessment of its performance, can be performed without the need for additional tools if necessary.

Keywords: X-rays; mammography; ROC curve; sensitivity and specificity; clinical decision support systems; AUC

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

This paper was prepared by a team of authors as a part of the research and development project "Development of a platform for improving quality of AI services for medical diagnostics" (EGISU No.: 123031400006-0)

For citation: Khrustacheva M.Yu., Vasilev Yu.A., Pamova A.P., Arzamasov K.M. ROC analysis for clinical decision support systems (CDSS) results in digital mammography images. *Medical Visualization.* https://doi.org/10.24835/1607-0763-1508

Received: 23.10.2024. **Accepted for publication:** 11.03.2025. **Published online:** 18.04.2025.

Введение

Активная цифровизация здравоохранения, наблюдаемая в последние годы, явилась основой для внедрения в медицинскую практику систем поддержки принятия врачебных решений (СППВР), в том числе на основе технологий искусственного интеллекта (ПО на основе ТИИ) [1]. Ключевые направления развития ТИИ отражены в Национальной стратегии развития искусственного интеллекта на период до 2030 г. (указ Президента Российской Федерации от 15.02.2024 № 124). Определены основные направления внедрения ИИ в здравоохранении. В первую очередь, это компьютерное зрение, которое позволяет распознавать образы, анализировать текстуры и формы, а также выявлять патологии на изображениях (компьютерная томография (КТ), магнитно-резонансная томография (МРТ), ультразвуковая диагностика (УЗИ)) [2]. Компьютерное зрение в медицинской диагностике является СППВР, позволяющей выявить, сегментировать (визуально), классифицировать и описать изменения на медицинском изображении. В зависимости от решаемой задачи функционал СППВР может быть разный: анализ текстовых данных (медицинские карты пациента), анализ временных рядов (электрокардиограмма, электроэнцефалограмма и др.), комплексные системы (использующие несколько типов данных). Цель СППВР – помочь врачам в оказании медицинской помощи, предоставляемой пациентам [3].

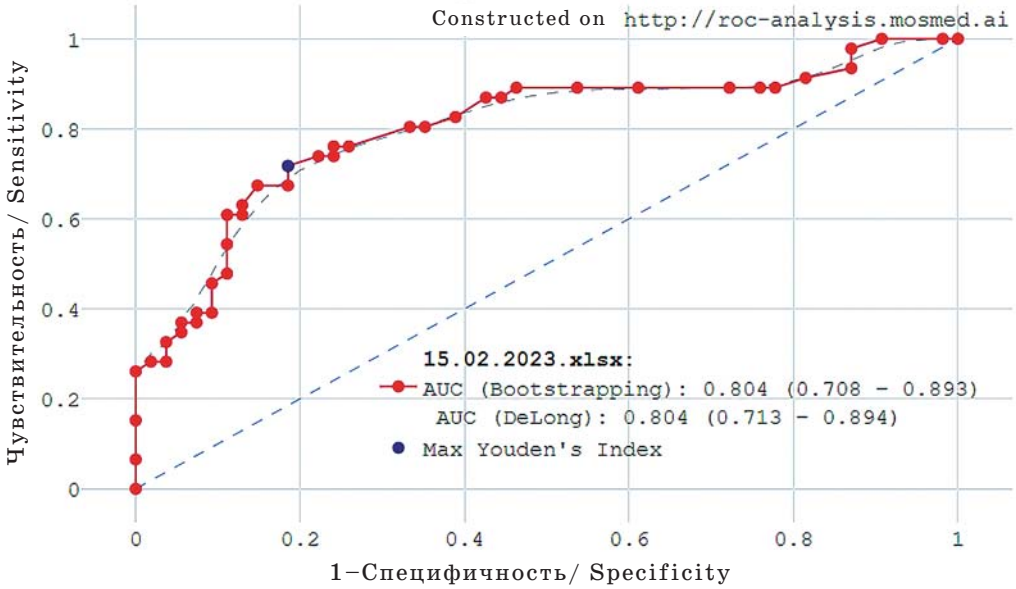

Рис. 1. ROC-кривая, полученная с помощью веб-инструмента для ROC-анализа.
Fig. 1. ROC-curve obtained via web-tool for ROC analysis.

Взаимодействие врача как с СППВР, в том числе с ПО на основе ТИИ, позволяет ему сконцентрироваться на решении более сложных диагностических задач [4]. Вопрос, который беспокоит медицинское сообщество: насколько же точными являются прогнозы алгоритмов? Последние исследования свидетельствуют о том, что ПО на основе ТИИ показывает результаты, сопоставимые с работой врача-рентгенолога. Очевидным фактом является то, что ТИИ имеют огромный потенциал для дальнейшего развития. Например, врач-рентгенолог может оценить предыдущие маммографические исследования и сопоставить результаты с текущими данными, на момент проведения исследования, а сервисы ИИ таких возможностей пока что не имеют, однако это является временным затруднением в связи с техническими сложностями реализации [5].

Одновременно с этим в настоящее время идут дискуссии об этических аспектах и правовом регулировании СППВР, в том числе ПО на основе ТИИ, в здравоохранении. Необходимо отметить, что на данный момент ни зарубежные, ни российские регулирующие органы не создали всеобъемлющую нормативно-правовую базу и систему контроля подобных технологий. Регистрируется все больше медицинских изделий с ТИИ и в качестве СППВР, а соответственно требуется минимизировать количество ошибок и обеспечить высокие стандарты качества таких технологий [6, 7].

На данном этапе приемочное тестирование СППВР, в том числе ПО на основе ТИИ, может осуществляться медицинской организацией. Приемочное тестирование – один из методов тестирования ПО, при котором проверяется его готовность к работе с конкретными задачами, перед передачей конечному пользователю. Эту задачу можно решить различными способами [8]. В представленной публикации будет рассмотрен ROC-анализ (от английского receiver operating curve). ROC-анализ – это метод, позволяющий наглядно сравнить и оценить качество ПО на основе ТИИ (и другие СППВР) с бинарной классификацией в основе путем построения характеристических ROC-кривых и вычисления метрик диагностической точности с определением порога принятия решений. ROC-кривая представляет собой график, отражающий зависимость верно классифицированных положительных случаев (чувствительность – ось ординат) от неверно классифицированных отрицательных случаев ([1 – специфичность] – ось абсцисс) (рис. 1) [9].

ROC-анализ впервые был использован во время Второй мировой войны для анализа и интерпретации радарных сигналов. Оценивалась способность оператора радара отличать “шум” от реальных сигналов – самолета или корабля. В 1970-х и 1980-х годах ROC-анализ стал широко использоваться в клинической онкологии и других областях медицины для оценки диагностических тестов и моделей прогнозирования [10].

В настоящее время данный вид анализа широко используется при оценке диагностической точности ПО на основе ТИИ как в рентгенологии,

так и в медицине в целом [11]. В связи с простотой интерпретации и распространенностью этого метода предлагаем ознакомиться с возможностью оценки результатов работы СППВР, в том числе ПО на основе ТИИ, при помощи ROC-анализа, который обеспечивает понятную оценку диагностической эффективности такого ПО. Он позволяет исследователям анализировать непрерывные данные (это количественные данные, которые могут принимать любые значения в пределах определенного интервала, они измеряются, а не подсчитываются, и могут быть представлены как целыми, так и дробными числами), создавать графические представления и сравнивать различные бинарные тесты, отражающие работу ПО на основе ТИИ (СППВР) [7]. В представленной публикации объясняются принципы построения ROC-кривой и основные метрики диагностической точности для оценки ТИИ.

Цель исследования: наглядная демонстрация применения веб-инструмента для ROC-анализа, который применяется для оценки результатов работы СППВР и ПО на основе ТИИ – на примере работы ПО с цифровыми маммографическими изображениями.

Материал и методы

Был использован ретроспективный набор данных маммографических исследований с результатами калибровочного тестирования при смене версионности одного из сервисов ИИ, участника Московского эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения этих технологий в системе здравоохранения (далее – Эксперимент). Калибровочное тестирование – один из этапов тестирования ИИ-сервиса, целью которого является оценка показателей диагностической точности и выбора оптимального порога принятия решений для разделения двух классов (нормы и патологии). Объем выборки – 100 исследований, проанализированных двумя версиями одного ИИ-сервиса, версия 1 – от 15.02.2023 и обновленная версия 2 – от 30.05.2023. Для оценки результатов работы ПО на основе ТИИ используется ROC-анализ.

Чтобы объяснить принципы построения и интерпретации кривой принятия решений (ROC-кривая), необходимо ввести понятия: чувствительность, специфичность и AUC, которые являются метриками диагностической точности (по отношению к ПО на основе ТИИ в области компьютерного зрения) и используются для оценки эффективности работы ТИИ.

Чувствительность – показатель, который отражает вероятность, что ПО на основе ТИИ предсказывает “патологию”, когда на снимке (цифровом изображении) действительно есть патологические изменения.

Специфичность – показатель, который отражает вероятность того, что ПО на основе ТИИ предсказывает “норму”, когда на снимке действительно нет патологических изменений.

AUC-ROC (area under curve, площадь под кривой) – показатель, который отображает производительность ТИИ в числовом выражении от 0 до 1. При AUC = 0 ПО на основе ТИИ определяет класс, к которому принадлежит объект, противоположный реальному, например, определяет “заболевания нет”, хотя на самом деле имеется патология. При приближении AUC к 1 ПО на основе ТИИ лучше определяет класс, к которому принадлежит объект, например, “заболевание есть”, когда оно действительно есть, или “заболевания нет”, когда его действительно нет. Если AUC равен 1, это означает, что предсказания ПО на основе ТИИ совершенно точные. При AUC = 0,5 – ПО на основе ТИИ работает как случайный классификатор.

Также в литературе можно встретить следующую экспертную шкалу: AUC ROC от 0,9 вплоть до 1,0 качество модели (ПО на основе ТИИ или СППВР) – превосходное, от 0,8 до 0,9 – очень хорошее, от 0,7 до 0,8 – хорошее, от 0,6 до 0,7 – среднее, а от 0,5 до 0,6 – неудовлетворительное [12].

В данной публикации рассматривается ПО на основе ТИИ, использующееся в маммографии (ММГ), которое производит бинарную классификацию изображения: наличие целевой патологии – (1) или на ММГ-изображении отсутствует целевая патология – (0). Применение ПО на основе ТИИ в ММГ повышает эффективность скрининговых программ, что, в конечном итоге, приводит к увеличению выявляемости и снижению смертности от злокачественных новообразований (ЗНО) молочной железы [13]. Чтобы объяснить принципы вычисления метрик диагностической точности, введем еще несколько необходимых определений.

Истинно положительный результат (True Positive (TP)) – это значение, показывающее количество пациентов, у которых действительно есть наличие целевой патологии и при этом ПО на основе ТИИ также показало наличие целевой патологии.

Истинно отрицательный результат (True Negative (TN)) – это значение, показывающее количество пациентов, у которых отсутствует целевая патология, и при этом ПО на основе ТИИ также показало отсутствие целевой патологии.

Ложноположительный результат (False Positive (FP)) – это значение, показывающее пациентов, у которых отсутствует целевая патология, а ПО на основе ТИИ показало наличие целевой патологии.

Таблица 1. Матрица путанности

Table 1. Confusion matrix

Тип предсказания **Type of prediction**	**Факт** **Fact**	
Предсказание модели Model prediction	Случай описан врачом как отрицательный (патологии нет) The case was described by the doctor as negative (no pathology)	Случай описан врачом как положительный (патология есть) The case was described by the doctor as positive (pathology)
Предсказание отрицательного случая (патологии нет) Prediction of a negative case (no pathology)	Истинно отрицательный результат (TN) True-negative result (TN)	Ложноотрицательный результат (FN) False-negative result (FN)
Предсказание положительного случая (патология есть) Prediction of a positive case (pathology)	Ложноположительный результат (FP) False Positive result (FP)	Истинно положительный результат (TP) True-positive result (TP)

Ложноотрицательный результат (False Negative (FN)) – это значение, показывающее пациентов, у которых есть наличие целевой патологии, а ПО на основе ТИИ показало отсутствие целевой патологии [14].

Для вычисления количества TP, TN, FP, FN строится матрица путанности (confusion matrix). Матрица путанности – это табличное представление описания производительности ПО на основе ТИИ на наборе данных, для которых известны фактические значения “ground-truth” (т.е. врач до использования ТИИ определил наличие патологии или норму, и эта информация считается достоверной). Это позволяет легко идентифицировать путаницу между классами, другими словами, когда один класс ошибочно помечен как другой со стороны ПО на основе ТИИ или СППВР.

В табл. 1 наглядно показано, как формируются вышеописанные показатели.

Допущение различными СППВР, в том числе ПО на основе ТИИ, ошибок определяется его чувствительностью (Se) и специфичностью (Sp).

Под чувствительностью (Sensitivity) подразумевается доля истинно положительных случаев среди субъектов с наличием патологии:

$$\text{Чувствительность} = \frac{\text{Истинно положительные (TP)}}{\text{Истинно положительные (TP)} + \text{Ложноотрицательные (FN)}}$$

Под специфичностью (Specificity) подразумевается доля истинно отрицательных случаев среди субъектов с отсутствием патологии

$$\text{Специфичность} = \frac{\text{Истинно отрицательные (TN)}}{\text{Истинно отрицательные (TN)} + \text{Ложноположительные (FP)}}$$

Дополнительно к этому необходимо указать, что есть и другие метрики диагностической точности ПО на основе ТИИ.

Precision-модели представляет собой долю верных предсказаний относительно всего количества наблюдений, иногда эту метрику называют “точность”, однако достаточно редко [15]:

$$\text{Precision} = \frac{\text{Истинно положительные (TP)}}{\text{Истинно положительные (TP)} + \text{Ложноположительные (FP)}}$$

Существует метрика accuracy (точность), под которой в большинстве случаев и подразумевают истинную “точность” в русскоязычных статьях:

$$\text{Точность (accuracy)} = \frac{\text{Истинно положительные (TP)} + \text{Истинно отрицательные (TN)}}{\text{Истинно положительные (TP)} + \text{Ложноположительные (FP)} + \text{Истинно отрицательные (TN)} + \text{Ложноотрицательные (FN)}}$$

Также важным значением в области ROC-анализа является порог принятия решений (cut-off) – определенное выбранное значение в ряду всех результатов предсказания ПО на основе ТИИ, которое определяет границу классов на ROC-кривой [16].

Индекс Юдена (J) (Youden’s index) – это разница между долей истинно положительных результатов (чувствительностью теста) и долей ложноположительных результатов [17].

Минимальное расстояние (d) – это значение, которое отражает минимальное расстояние от ROC-кривой до точки на графике с координатами (0;1).

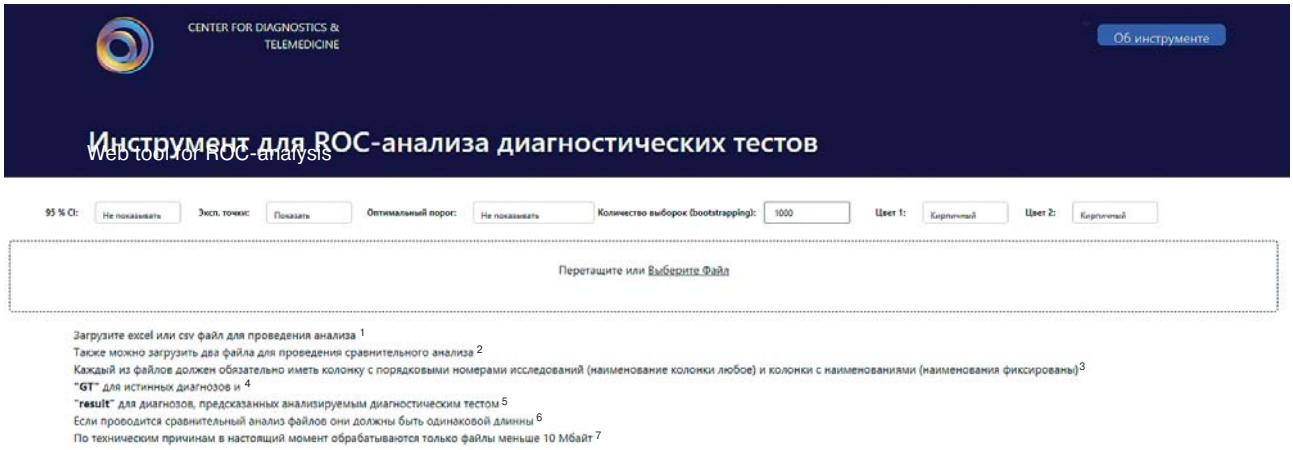

Рис. 2. Внешний интерфейс веб-инструмента для ROC-анализа.

Fig. 2. External interface of the web tool for ROC-analysis.

[1] Upload excel or csv file for analyzation
[2] Also you can upload two files for comparison
[3] File should contain a column with study id (with any column name) and columns named as
[4] "**GT**" with ground truth (true) values
[5] "**result**" with predicated values
[6] Files should have same length
[7] Due to technical reasons files larger than 10 Mb cannot be analyzed now

В представленной работе демонстрируется использование веб-инструмента для ROC-анализа для обработки результатов диагностических тестов, представлен на сайте: https://roc-analysis.mosmed.ai [18].

Чтобы воспользоваться этим инструментом, результаты работы ПО на основе ТИИ необходимо представить в виде таблицы формата CSV или XLS со следующими значениями трех столбцов (иногда в других источниках их называют "поля") (табл. 2):

1. Столбец (поле) с номером пациента или каким-то маркером по каждому пациенту (уникальным идентификатором, на англ. "uid").

2. "Ground truth" – GT или "эталон", столбец, содержащий мнение врача в виде бинарного набора чисел есть – (1) или нет – (0) целевой патологии на цифровой маммограмме. GT является в этом случае эталонным решением (т.е. единственно достоверным), с которым будут сравниваться результаты работы ПО на основе ТИИ.

3. Столбец "Result", который должен содержать вероятность наличия или отсутствия целевой патологии по результатам работы ПО на основе ТИИ (вероятность наличия целевой патологии должна быть представлена в диапазоне от 0 – точно нет, до 1 – точно есть).

Таблица 2. Пример таблицы для работы с веб-инструментом

Table 2. The example of the table for using web tool

Номер исследования Number of investigation	GT	Result
1	1	19
2	1	2
3	0	98

Рассмотрим детально возможности персональной настройки будущей ROC-кривой на рис. 2.

1. Поле "95% CI" (от англ. confidence interval), т.е. 95% доверительный интервал (ДИ) – позволяет настроить отображение доверительного интервала.

2. Поле "Эксп. Точки" – позволяет настроить отображение экспоненциальных точек на графике.

3. Поле "Оптимальный порог" (cut-off) – позволяет настроить отображение оптимального порога принятия решений на ROC-кривой двумя способами: через индекс Юдена (J) или минимальное расстояние (d).

4. Поле "Количество выборок (bootstrapping)" – позволяет настроить число выборок от 1000 до 100 000. Это вычислительный статистический метод. Он позволяет оценить распределение выборочной статистики путем многократной генерации выборок методом Монте-Карло на основе уже имеющейся выборки. Бутстрап помогает коррек-

тировать смещение, тестировать гипотезы и строить доверительные интервалы [19]. То есть случайным образом из переданного инструменту набора данных будут отбираться случайные пациенты, формируя случайные выборки сколько угодно раз (в зависимости от указаний пользователя в этом поле).

После настройки интерфейса и параметров инструмента, а также сформированной таблицы, соответствующей критериям, описанным выше, можно приступить к непосредственной работе с инструментом ROC-анализа.

Результаты исследования

После загрузки файла в формате CSV или XLS был получен график с ROC-кривой, которая отражает результаты работы ПО на основе ТИИ первой версии от 15.02.2023 г. (рис. 3).

Синей точкой на графике отмечен оптимальный порог принятия решений для этой ROC-кривой, рассчитанный методом максимизации индекса Юдена. Также для каждой точки на графике мы можем узнать значения: чувствительности, специфичности и индекса Юдена, рассчитанных для этой точки. Можно настраивать порог принятия решений в каждой точке на графике, в том числе таким образом, чтобы получить 100% чувствительность или 100% специфичность, в зависимости от стоящих перед ПО на основе ТИИ задач. Настройка порога со 100% специфичностью на скрининговых исследованиях, таких как ММГ, не является оправданной. При скрининговых исследованиях важен высокий показатель чувствительности, который отражает истинную способность ПО на основе ТИИ выявлять патологию, в то время как специфичность отражает способность алгоритма определять норму. Для того чтобы продемонстрировать еще одну функцию веб-инструмента, выгрузим две таблицы с результатами работы ПО на основе ТИИ как первой версии от 15.02.23, так и второй версии от 30.05.2023. Получен следующий график (рис. 4).

Обсуждение

При создании ИИ-сервисов разработчик закладывает определенное пороговое значение, так называемый порог принятия решений (cut-off). При превышении этого порога ИИ-сервис будет классифицировать исследование как содержащее целевую патологию. В зависимости от клинической задачи, стоящей перед медицинской организацией, может потребоваться подбор оптимального значения порога принятия решений, отличного от того, который задал разработчик. Данный процесс называется *тонкой настройкой*. Для скрининговых исследований может быть применима настройка на 100% чувствительность, что, в свою очередь, позволит не пропустить целевую патологию. В иных случаях может потребоваться 100% специфичность (чтобы наверняка отсечь тех пациентов, которые со 100% вероятностью не имеют целевой патологии) для того, чтобы сконцентрироваться на тех пациентах, которым может требоваться незамедлительная помощь. Осуществлять такую *тонкую настройку* возможно с применением рассмотренного инструмента (определяя оптимальный порог принятия решений в зависимости от клинической задачи).

На сегодняшний день приемочное тестирование СППВР, в том числе ПО на основе ТИИ, может проводиться медицинской организацией. Для того чтобы принять решение, подходит ли новый сервис СППВР или ПО на основе ТИИ для целей медицинской организации, нужно оценить его соответствие заявленным характеристикам со стороны разработчика. В представленной работе показано, как использовать веб-инструмент для ROC-анализа для решения вышеописанных задач.

В результате работы с инструментом получен график, на котором наглядно представлены кривая ROC, порог принятия решений по индексу Юдена, значения чувствительности и специфичности для каждой точки на графике и в точке с оптимальным порогом принятия решений, вычисленным по индексу Юдена.

Применяя ROC-анализ в режиме сравнения двух ROC-кривых (см. рис. 4), можно оценить, как изменяются метрики диагностической точности у новой версии ИИ-сервиса или СППВР по сравнению с предыдущей [7].

На графике с двумя ROC-кривыми видно разницу в производительности двух версий одного ПО, инструмент вычислил площади под кривой для каждой версии, может показаться, что последняя версия от 30.05.2023 имеет большую производительность на основе значения AUC (AUC = 0,842) (примем за гипотезу H1). Поэтому воспользуемся еще одной функцией веб-инструмента и проверим нулевую гипотезу (H0) о том, что вторая версия ИИ-сервиса на самом деле не отличается от первой. Необходимо уточнить, за H0 обычно принимают утверждение, что ничего не изменилось. Также в режиме сравнения двух ROC-кривых инструмент высчитывает величину *p-value* (величина, используемая в статистике при тестировании гипотез, показывает вероятность получения наблюдаемых результатов при условии, что нулевая гипотеза верна, или вероятность ошибки в случае отклонения нулевой гипотезы). В данном случае *p-value* = 0,258. Сравниваем с заранее заданным уровнем

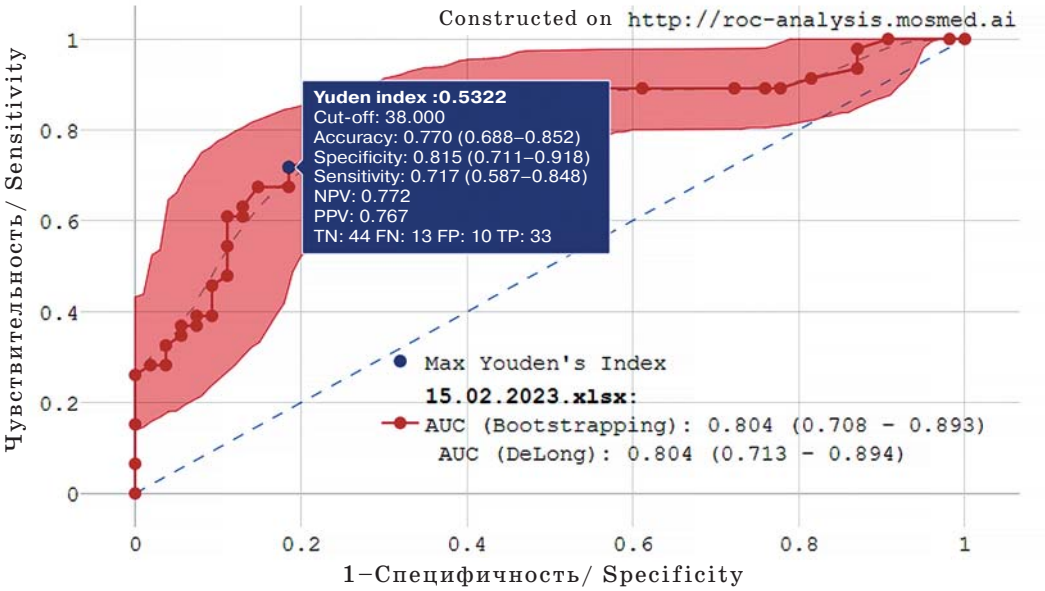

Рис. 3. ROC-кривая, полученная с помощью веб-инструмента для ROC-анализа.
PPV (positive predictive value) – положительная прогностическая значимость. NPV (negative predictive value) – отрицательная прогностическая значимость. Красной областью на графике выделен 95% доверительный интервал.

Fig. 3. ROC-curve obtained via web-based tool for ROC analysis.
PPV – positive predictive value. NPV – negative predictive value. The red area is demonstrated 95% confidence interval. Индекс Юдена – Yuden index. Порог принятия решений – cut-off. Точность – Accuracy. Чувствительность – Sensitivity. Специфичность – Specificity.

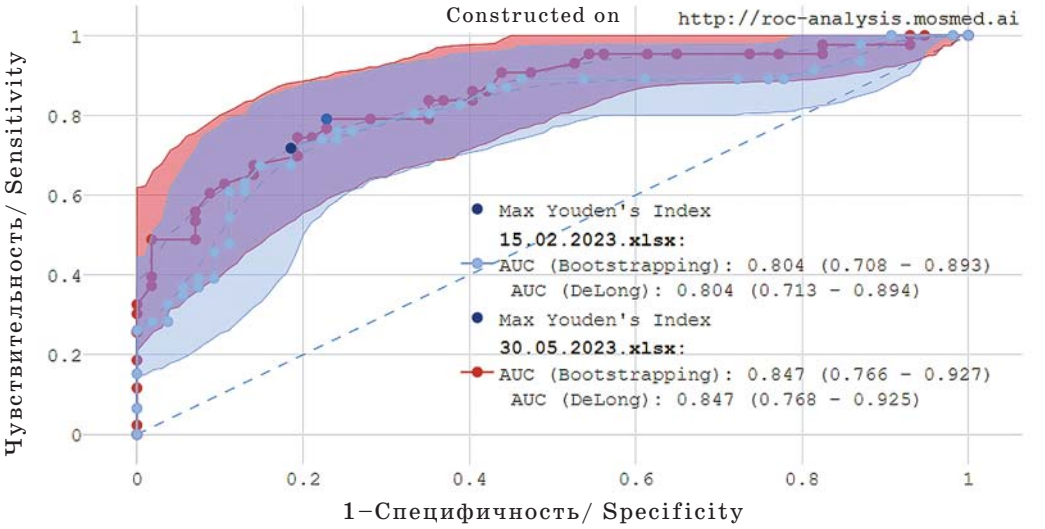

Рис. 4. Сравнение двух ROC-кривых для двух версий ПО на основе ТИИ. Красным цветом на графике обозначена ROC-кривая, отражающая результаты работы версии ПО на основе ТИИ от 15.02.2023, а синим цветом на графике обозначена ROC-кривая, отражающая результаты работы версии от 30.05.2023, на одном наборе данных, состоящем из 100 исследований.

Fig. 4. Comparison of the two ROC curves for the two versions of the AI software. The red ROC curve in the graph represents the ROC curve for the AI service version of the software dated 02/15/2023, and the blue ROC curve in the graph represents the ROC curve for the version dated 05/30/2023, on a single dataset of 100 studies.

Таблица 3. Чувствительность и специфичность на ROC-кривой, отражающей результат работы алгоритма версии от 15.02.2023

Table 3. Sensitivity and specificity on ROC-curve of the algorithm version dated 02.15.2023

Точка на графике Point on the graph	Чувствительность (Se) Sensitivity	Специфичность (Sp) Specificity
1	0.261	1.000
2	1.000	0.930
3	0.717	0.815

альфа (пороговым значением, которое определяет, насколько необычными должны быть данные, чтобы отвергнуть нулевую гипотезу) 0,258 > 0,05, в данном случае нулевая гипотеза не отвергается, и мы можем говорить о том, что статистически значимые различия между версиями алгоритма от 15.02.2023 г. и от 30.05.2024 г. на самом деле отсутствуют. Таким образом, как первая версия ПО на основе ТИИ, так и вторая версия могут с одинаковой эффективностью использоваться для скрининговой ММГ с известным AUC > 0,8 на заданном пороге принятия решений по индексу Юдена.

Как упоминалось выше, с помощью веб-инструмента возможно осуществить *тонкую настройку* порога принятия решений (cut-off) под необходимые медицинской организации чувствительность или специфичность [20] в зависимости от задачи использования ПО. Как показано в табл. 3, каждой точке на графике соответствует своя чувствительность и специфичность.

В табл. 3 показан пример трех точек, отображающихся на ROC-кривой, построенной с помощью веб-инструмента, найдены точки с максимальной специфичностью, с максимальной чувствительностью и третья точка с метриками диагностической точности ПО на основе ТИИ, определенная на пороге принятия решений, вычисленного с помощью индекса Юдена.

Заключение

Продемонстрирован процесс работы с веб-инструментом для ROC-анализа, который имеет ряд преимуществ перед другими аналогами, например бесплатный доступ, что отличает его от аналогичного сервиса MedCalc, и возможность пользоваться инструментом без установки на персональный компьютер. Оценить производительность СППВР и ПО на основе ТИИ можно также используя языки программирования, такие как Python или R, но это требует специфических навыков, медицинскому работнику будет намного удобнее воспользоваться веб-инструментом для ROC-анализа. Веб-инструмент для ROC-анализа рекомендуется использовать в следующих случаях:

1) для оценки эффективности работы СППВР, включая ПО на основе ТИИ, в частности для приемочного тестирования таких технологий медицинской организацией;

2) для оценки результатов работы СППВР, в том числе ПО на основе ТИИ, при смене версионности;

3) для *тонкой настройки* сервисов-ИИ и, в частности порога принятия решений;

4) для *тонкой настройки* СППВР, в том числе ПО на основе ТИИ, с заданными чувствительностью и специфичностью.

Инструмент для ROC-анализа позволяет производить вышеописанные манипуляции любому пользователю, не требуя специальных навыков программирования, без покупки дополнительного программного обеспечения, а также без предустановки программы.

Участие авторов

Хрустачева М.Ю. – проведение исследования, сбор и обработка данных, написание текста, анализ и интерпретация полученных данных, статистическая обработка данных, ответственность за целостность всех частей статьи, подготовка, создание опубликованной работы, обзор публикаций по теме статьи.

Васильев Ю.А. – концепция и дизайн исследования, подготовка и редактирование текста, участие в научном дизайне, утверждение окончательного варианта статьи.

Памова А.П. – концепция и дизайн исследования, написание текста, подготовка и редактирование текста, сбор и обработка данных, анализ и интерпретация полученных данных, участие в научном дизайне, утверждение окончательного варианта статьи.

Арзамасов К.М. – концепция и дизайн исследования, написание текста, подготовка и редактирование текста, сбор и обработка данных, анализ и интерпретация полученных данных, участие в научном дизайне, утверждение окончательного варианта статьи.

Authors' participation

Khrustacheva M.Yu. – conducting research, collection and analysis of data, writing text, analysis and interpretation

of the obtained data, statistical analysis, responsibility for the integrity of all parts of the article, preparation and creation of the published work, review of publications.

Vasilev Yu.A. – concept and design of the study, participation in scientific design, approval of the final version of the article.

Pamova A.P. – concept and design of the study, writing text, text preparation and editing, collection and analysis of data, analysis and interpretation of the obtained data, participation in scientific design, approval of the final version of the article.

Arzamasov K.M. – concept and design of the study, writing text, text preparation and editing, collection and analysis of data, analysis and interpretation of the obtained data, participation in scientific design, approval of the final version of the article.

Список литературы

1. 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.
2. Говорухина В.Г., Семенов С.С., Гележе П.Б., Диденко В.В., Морозов С.П., Андрейченко А.Е. Роль маммографии в радиомике рака молочной железы. *Digital Diagnostics*. 2021; 2 (2): 185–199. https://doi.org/10.17816/DD70479
3. Алексеева М.Г., Зубов А.И., Новиков М.Ю. Искусственный интеллект в медицине. *Международный научно-исследовательский журнал.* 2022; 7: 10–13. https://doi.org/10.23670/IRJ.2022.121.7.038
4. Ройтберг П.Г., Блинов Д.С., Черемисин В.М. Технологии искусственного интеллекта в автоматизации выполнения стандартных задач врача-рентгенолога. *Проблемы стандартизации в здравоохранении*. 2020; 9 (10): 29–33. https://doi.org/10.26347/1607-2502202009-10029-033
5. Rodriguez-Ruiz A., Lang K., Gubern-Merida A. et al. Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists. *J. Natl. Cancer Inst.* 2019; 111 (9): 916–922. https://doi.org/10.1093/jnci/djy222
6. Шарова Д.Е., Зинченко В.В., Ахмад Е.С., Мокиенко О.А., Владзимирский А.В., Морозов С.П. К вопросу об этических аспектах внедрения систем искусственного интеллекта в здравоохранении. *Digital Diagnostics.* 2021; 3: 356–368.
7. Морозов С.П., Владзимирский А.В., Гомболевский В.А., Кляшторный В.Г., Федулова И.А., Власенков Л.А. Искусственный интеллект в скрининге рака легкого: оценка диагностической точности алгоритма для анализа низкодозовых компьютерных томографий. *Туберкулез и болезни легких*. 2020; 8: 24–31.
8. Васильев Ю.А., Владзимирский А.В., Шарова Д.Е. Клинические испытания систем искусственного интеллекта (лучевая диагностика). В кн.: Лучшие практики лучевой и инструментальной диагностики. 2-е изд., перераб. и доп. М.: ГБУЗ «НПКЦ ДиТ ДЗМ», 2023. 40 с.
9. Park S.H., Goo J.M., Jo C.H. Receiver Operating Characteristic (ROC) Curve: Practical Review for Radiologists. *Korean J. Radiol*. 2004; 5 (1): 11–18. https://doi.org/10.3348/kjr.2004.5.1.11
10. Junge M.R., Dettori J.R. ROC Solid: Receiver Operator Characteristic (ROC) Curves as a Foundation for Better Diagnostic Tests. *Global Spine J.* 2024; 8 (4): 424–429. https://doi.org/10.1177/2192568218778294
11. Мирошниченко И.И. Мониторинг эффективности антипсихотической терапии. *Психиатрия*. 2022; 20: 128–138.
12. Davis J., Goadrich M. The relationship between Precision-Recall and ROC curves. *Proceedings of the 23rd International Conference on Machine Learning*. 2006; 6: 233–240. https://doi.org/10.1145/1143844.1143874
13. Солодкий В.А., Каприн А.Д., Нуднов Н.В. и др. Современные системы поддержки принятия врачебных решений на базе искусственного интеллекта для анализа цифровых маммографических изображений. *Вестник рентгенологии и радиологии*. 2023; 2: 151–162.
14. Kumar R., Indrayan A. Receiver Operating Characteristic (ROC) Curve for Medical Researchers. *Indian Pediatr*. 2011; 48: 277–287. https://doi.org/10.1007/s13312-011-0055-4
15. Fawcett T. ROC Graphs: Notes and Practical Considerations for Researchers. *Machine Learning*. 2004; 31: 1–38.
16. Ковалев А.А., Кузнецов Б.К., Ядченко А.А., Игнатенко В.А. Оценка качества бинарного классификатора в научных исследованиях. *Проблемы здоровья и экологии.* 2020; 4: 105–113.
17. Корнеенков А.А., Рязанцев С.В., Вяземская Е.Э. Вычисление и интерпретация показателей информативности диагностических медицинских технологий. *Медицинский Совет*. 2019; 20: 45–51. https://doi.org/10.21518/2079-701X-2019-20-45-51
18. Electronic resource: Web tool for ROC analysis. https://roc-analysis.mosmed.ai.
19. Efron B., Tibshirani R. The Bootstrap Method for Assessing Statistical Accuracy. *Behaviormetrika*. 1985; 12: 1–35. https://doi.org/10.2333/bhmk.12.17_1\
20. Васильев Ю.А., Тыров И.А., Владзимирский 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; 2: 93–104.

References

1. Chetverikov S.F., Arzamasov K.M., Andreichenko A.E. et al. Approaches to sampling for quality control of artificial intelligence in biomedical research. *Sovremennye Tehnologii v Medicine*. 2023; 2: 19–27. https://doi.org/10.17691/stm2023.15.2.02 (In Russian)
2. Govorukhina V.G., Semenov S.S., Gelezhe P.B. et al. The role of mammography in breast cancer radiomics. *Digital Diagnostics.* 2021; 2 (2): 185–199. https://doi.org/10.17816/DD70479 (In Russian)

3. Alekseeva M.G., Zubov A.I., Novikov M.Yu. Artificial intelligence in medicine. *International Research Journal*. 2022; 7: 10–13. https://doi.org/10.23670/IRJ.2022.121.7.038 (In Russian)
4. Roitberg PG, Blinov DS, Cheremisin VM. Integrating AI technologies in radiology workflows. *Health Care Standardization Problems*. 2020; 9 (10): 29–33. https://doi.org/10.26347/1607-2502202009-10029-033 (In Russian)
5. Rodriguez-Ruiz A., Lang K., Gubern-Merida A. et al. Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists. *J. Natl. Cancer Inst.* 2019; 111 (9): 916–922. https://doi.org/10.1093/jnci/djy222
6. Sharova D.S., Zinchenko V.V, Akhmad E.S. et al. On the issue of ethical aspects of the artificial intelligence systems implementation in healthcare. *Digital Diagnostics.* 2021; 3: 356–368. https://doi.org/10.17816/DD77446 (In Russian)
7. Morozov S.P., Vladzimirskiy A.V., Gombolevskiy V.A. et al. Artificial intelligence in lung cancer screening: assessment of the diagnostic accuracy of the algorithm analyzing low-dose computed tomography. *Tuberculosis and Lung Diseases.* 2020; 8: 24–31. https://doi.org/10.21292/2075-1230-2020-98-8-24-31 (In Russian)
8. Vasilev Yu.A., Vladzimirskij A.V., Sharova D.E.. Klinicheskie ispytaniya sistem iskusstvennogo intellekta (luchevaya diagnostika). In: Luchshie praktiki luchevoj i instrumental'noj diagnostiki. 2-e izd., pererab. i dop. M.: "Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department", 2023. 40 p. (In Russian)
9. Park S.H., Goo J.M., Jo C.H. Receiver Operating Characteristic (ROC) Curve: Practical Review for Radiologists. *Korean J. Radiol*. 2004; 5 (1): 11–18. https://doi.org/10.3348/kjr.2004.5.1.11
10. Junge M.R., Dettori J.R. ROC Solid: Receiver Operator Characteristic (ROC) Curves as a Foundation for Better Diagnostic Tests. *Global Spine J.* 2024; 8 (4): 424–429. https://doi.org/10.1177/2192568218778294
11. Miroshnichenko I.I. Monitoring of the Antipsychotic Therapy Effectiveness. *Psychiatry*. 2022; 20: 128–138. https://doi.org/10.30629/2618-6667-2022-20-4-128-138 (In Russian)
12. Davis J., Goadrich M. The relationship between Precision-Recall and ROC curves. *Proceedings of the 23rd International Conference on Machine Learning*. 2006; 6: 233–240. https://doi.org/10.1145/1143844.1143874
13. Solodkiy V.A., Kaprin A.D., Nudnov N.V. et al. Contemporary Medical Decision Support Systems Based on Artificial Intelligence for the Analysis of Digital Mammographic Images. *Journal of Radiology and Nuclear Medicine*. 2023; 2: 151–162. https://doi.org/10.20862/0042-4676-2023-104-2-151-162 (In Russian)
14. Kumar R., Indrayan A. Receiver Operating Characteristic (ROC) Curve for Medical Researchers. *Indian Pediatr*. 2011; 48: 277–287. https://doi.org/10.1007/s13312-011-0055-4
15. Fawcett T. ROC Graphs: Notes and Practical Considerations for Researchers. *Machine Learning*. 2004; 31: 1–38.
16. Kovalev A.A., Kuznetsov B.K., Yadchenko A.A., Ignatenko V.A. Assessment of the quality of a binary classifier in research. *Health and Ecology Issues*. 2020; 4: 105–113. (In Russian)

Для корреспонденции*: Хрустачева Маргарита Юрьевна – e-mail: myukhrustacheva@mail.ru

Хрустачева Маргарита Юрьевна – аспирант ГБУЗ "Научно-практический клинический центр диагностики и телемедицинских технологий ДЗ города Москвы"; врач-рентгенолог ГБУЗ "Московский многопрофильный клинический центр «Коммунарка»" ДЗ города Москвы, Москва. http://orcid.org/0009-0001-1381-0809

Васильев Юрий Александрович – канд. мед. наук, главный врач ГБУЗ "Научно-практический клинический центр диагностики и телемедицинских технологий ДЗ города Москвы"; главный внештатный специалист по лучевой диагностике, Москва. http://orcid.org/0000-0002-5283-5961

Памова Анастасия Петровна – канд. мед. наук, начальник сектора разработки систем внедрения медицинских интеллектуальных технологий отдела медицинской информатики, радиомики и радиогеномики ГБУЗ "Научно-практический клинический центр диагностики и телемедицинских технологий ДЗ города Москвы", Москва. http://orcid.org/0000-0002-0041-3281

Арзамасов Кирилл Михайлович – канд. мед. наук, руководитель отдела медицинской информатики, радиомики и радиогеномики ГБУЗ "Научно-практический клинический центр диагностики и телемедицинских технологий ДЗ города Москвы"; доцент кафедры технологий искусственного интеллекта ФГБОУ ВО "МИРЭА – Российский технологический университет", Москва. http://orcid.org/0000-0001-7786-0349

Contact*: Margarita Yu. Khrustacheva – e-mail: myukhrustacheva@mail.ru

Margarita Yu. Khrustacheva – graduate student, Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of Moscow Healthcare Department; radiologist of Moscow Clinical Center "Kommunarka" of Moscow Healthcare Department, Moscow. http://orcid.org/0009-0001-1381-0809

Yuri A. Vasiliev – Cand. of Sci. (Med.), Chief Medical Officer of the Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Healthcare Department; Chief specialist in radiation diagnostics, Moscow. http://orcid.org/0000-0002-5283-5961

Anastasia P. Pamova – Cand. of Sci. (Med.), Head of Medical Intelligent Technologies Integration Sector of the Department of Medical Informatics, Radiomics, and Radiogenomics, the Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department, Moscow. http://orcid.org/0000-0002-0041-3281

Kirill M. Arzamasov – Cand. of Sci. (Med.), Head of the Department of Medical Informatics, Radiomics, and Radiogenomics of the Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department; Associate Professor of the Department of Artificial Intelligence Technologies, MIREA – Russian Technological University, Moscow. http://orcid.org/0000-0001-7786-0349