



Диагностическая точность технологий искусственного интеллекта при анализе результатов магнитно-резонансной томографии для выявления дегенеративных изменений пояснично-крестцового отдела

© Тхакахова Р.А.^{1*}, Владзимирский А.В.¹, Арзамасов К.М.^{1,2}

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

² ФГБОУ ВО "Самарский государственный медицинский университет" Минздрава России; 443099 Самара, ул. Чапаевская, д. 89, Российская Федерация

Дегенеративные заболевания позвоночника – одна из ключевых медицинских проблем, являющаяся частой причиной хронических болей, снижения качества жизни и трудоспособности. Современные подходы к терапии, включая консервативные методики и хирургические вмешательства, не всегда демонстрируют достаточную клиническую эффективность, что актуализирует вопросы совершенствования диагностических возможностей.

Магнитно-резонансная томография (МРТ) – основной метод визуализации патологий позвоночника, обеспечивающий детальную оценку структур, включая межпозвонковые диски, спинной мозг и окружающие ткани. Рутинная практика МР-диагностики сопряжена с существенными методическими сложностями, такими как необходимость выполнения многочисленных стандартизированных измерений. Внедрение искусственного интеллекта в радиологию предлагает решение этих проблем: алгоритмы способны автоматизировать обработку изображений, ускоряя анализ и повышая точность диагностики.

Цель исследования: оценить диагностическую точность автоматизированного выявления дегенеративных изменений пояснично-крестцового отдела позвоночника при проведении МРТ с применением технологий искусственного интеллекта.

Материал и методы. Проведено ретроспективное диагностическое исследование, количество исследований – 100. В качестве индекс-теста использованы сервисы искусственного интеллекта (далее называемые сервис ИИ-1, ИИ-2, ИИ-3), интегрированные в Единую радиологическую информационную систему (ЕРИС) в составе Единой медицинской информационно-аналитической системы города Москвы (ЕМИАС).

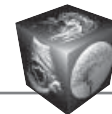
Тестовый набор данных, подготовленный тремя экспертами на основе МРТ пояснично-крестцового отдела позвоночника (T2sag, T2ax, срез 4 мм). Критерии включения: пациенты 18–90 лет без артефактов или сопутствующих заболеваний. Оценка ИИ-сервисов включала анализ дегенеративных изменений: дорсальные выпячивания межпозвонковых дисков; размеры (передне-заднего размера дурального мешка на аксиальных и сагиттальных изображениях, фронтального размера дурального мешка на аксиальных изображениях) и площадь дурального мешка.

Результаты. Сервисы ИИ-1 и ИИ-3 показали высокую диагностическую точность ($AUC > 0,95$), что позволяет рекомендовать их для клинического применения. Сервис ИИ-2 продемонстрировал метрики значительно ниже заявленных ($AUC < 0,95$), что ограничивает практическое использование и требует доработки.

Заключение. Проведенное исследование демонстрирует высокий потенциал технологий искусственного интеллекта для автоматизации МРТ-диагностики дегенеративных изменений пояснично-крестцового отдела позвоночника. Два из трех протестированных ИИ-сервисов (ИИ-1 и ИИ-3) показали высокую диагностическую точность ($AUC > 0,95$), что позволяет рекомендовать их для внедрения в клиническую практику с целью повышения эффективности и стандартизации анализа нейровизуализации. В то же время значительное расхождение заявленных и реальных метрик третьего сервиса (ИИ-2) подчеркивает необходимость строгой валидации алгоритмов перед их использованием в рутинной диагностике.

Ключевые слова: магнитно-резонансная томография; искусственный интеллект; дегенеративные заболевания позвоночника; межпозвонковый диск

Авторы подтверждают отсутствие конфликтов интересов.



Для цитирования: Тхакахова Р.А., Владимирский А.В., Арзамасов К.М. Диагностическая точность технологий искусственного интеллекта при анализе результатов магнитно-резонансной томографии для выявления дегенеративных изменений пояснично-крестцового отдела. *Медицинская визуализация*. 2026; 30 (2): 124–134. <https://doi.org/10.24835/1607-0763-1580>

Поступила в редакцию: 22.06.2025. **Принята к печати:** 12.03.2026. **Опубликована online:** 12.05.2026.

Diagnostic accuracy of artificial intelligence technologies in analyzing magnetic resonance imaging results for detecting degenerative changes in the lumbosacral spine

© Regina A. Tkhakakhova^{1*}, Anton V. Vladzimirsky¹, Kirill M. Arzamasov^{1, 2}

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 Samara State Medical University; 89, Chapaevskaya str., Samara, 443099, Russian Federation

Degenerative spinal diseases are a major medical problem and a common cause of chronic pain, reduced quality of life, and loss of work capacity. Modern treatment approaches, including conservative methods and surgical interventions, do not always demonstrate sufficient clinical efficacy, highlighting the need for improved diagnostic capabilities.

Magnetic resonance imaging is the primary modality for visualizing spinal pathologies, providing a detailed assessment of structures, including intervertebral discs, the spinal cord, and surrounding tissues. However, routine MRI diagnostics face significant methodological challenges, such as the need to perform numerous standardized measurements. The integration of artificial intelligence into radiology offers a solution to these problems: algorithms can automate image processing, speeding up analysis and enhancing diagnostic accuracy.

Objective: To evaluate the diagnostic accuracy of automated detection of degenerative changes in the lumbosacral spine using AI-assisted MRI.

Materials and methods. A retrospective diagnostic study was conducted, comprising 100 examinations. Artificial intelligence services (hereinafter referred to as AI service 1, AI-2, AI-3) integrated into the Unified Radiological Information System (URIS) of the Moscow Unified Medical Information and Analytical System (EMIAS) were used as the index test.

A test dataset was prepared by three experts based on MRI scans of the lumbosacral spine (T2 sagittal, T2 axial, 4 mm slice thickness). Inclusion criteria were patients aged 18–90 years without artifacts or comorbid pathologies. The evaluation of the AI services involved analyzing degenerative changes: dorsal intervertebral disc protrusions; measurements (anteroposterior diameter of the dural sac on axial and sagittal images, transverse diameter of the dural sac on axial images); and the cross-sectional area of the dural sac.

Conclusion. Calibration testing confirmed the importance of evaluating the effectiveness of AI services in medicine. AI-1 and AI-3 demonstrated high reliability (AUC > 0.95), making them suitable for clinical use. AI-2 requires refinement, as its performance metrics were significantly lower than claimed and further declined in external validation, limiting its practical application.

Keywords: magnetic resonance imaging; artificial intelligence; degenerative spinal diseases; intervertebral disc

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

For citation: Tkhakakhova R.A., Vladzimirsky A.V., Arzamasov K.M. Diagnostic accuracy of artificial intelligence technologies in analyzing magnetic resonance imaging results for detecting degenerative changes in the lumbosacral spine. *Medical Visualization*. 2026; 30 (2): 124–134. <https://doi.org/10.24835/1607-0763-1580>

Received: 22.06.2025.

Accepted for publication: 12.03.2026.

Published online: 12.05.2026.



Введение

Дегенеративные заболевания позвоночника характеризуются прогрессирующими дегенеративно-дистрофическими изменениями тканей позвоночных сегментов –деградацией тканей межпозвонковых дисков, суставов, связочного аппарата [1], сопровождающие их болевые и неврологические симптомы являются одной из ключевых проблем современного здравоохранения. Это связано с постоянным высоким числом пациентов трудоспособного возраста, нередко неудовлетворительными результатами консервативного лечения и частыми рецидивами болевого синдрома после хирургического вмешательства. Боли в спине ограничивают повседневную активность, снижают качество жизни и негативно влияют на психоэмоциональное состояние пациентов [2]. Согласно данным Всемирной организации здравоохранения, дегенеративно-дистрофические заболевания позвоночника (ДДЗП) встречаются у 80% населения и составляют до 90% всех случаев хронических заболеваний [3]. Некоторые исследования показывают, что распространенность ДДЗП и нарушений функции позвоночника составляет 122 случая на каждую тысячу человек взрослого населения, при этом спондилогенные боли испытывают от 80 до 100% людей [4]. ДДЗП с сопутствующим спондилогенным болевым синдромом занимает второе место по частоте обращений к врачам после респираторных заболеваний и третье по частоте госпитализаций [5].

Магнитно-резонансная томография (МРТ) позвоночника является одним из наиболее распространенных методов инструментальной диагностики, позволяющим детально оценить различные патологические изменения в области дегенеративных заболеваний позвоночника. Благодаря высокому контрастному разрешению МРТ предоставляет возможность тщательного изучения паравертебральных мягких тканей, межпозвонковых дисков, суставов, связочного аппарата. Мультипланарные изображения, получаемые при МРТ, значительно расширили диагностические возможности, позволяя точно определять уровень и протяженность патологических изменений, а также степень сужения позвоночного канала и наличие фрагментов тел позвонков, межпозвонковых дисков, гематом и рубцовых образований. Современное техническое оборудование и специализированное программное обеспечение дают возможность точного измерения размеров тел позвонков, межпозвонковых дисков и позвоночного канала. МРТ занимает ведущие позиции в диагностике опухолевых процессов, затрагивающих позвоночник [6].

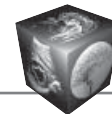
В ежедневной практике врача-рентгенолога кабинета МРТ исследования позвоночника являются наиболее распространенными, при этом большое количество измерений, которые необходимо проводить при оценке позвоночника, занимает значительное время в работе врача-рентгенолога [7]. Трудоемкость обработки изображений и субъективность интерпретации результатов не только увеличивают время проведения исследования, но и повышают риск пропуска патологии.

В последние годы технологии искусственного интеллекта (ИИ) активно внедряются в медицинскую диагностику, предлагая автоматизированные решения для анализа медицинских изображений [8]. ИИ-сервисы способны обрабатывать большие объемы данных, выявлять патологические изменения с высокой точностью и снижать нагрузку на врачей.

Работа А. Jamaludin и соавт. была направлена на автоматизацию процесса оценки межпозвонковых дисков и тел позвонков. Были получены изображения 12 018 отдельных дисков и проведена оценка. Первым этапом выполнялось оконтуривание тел позвонков, затем – дисков. Диски и тела позвонков анализируются по 6 признакам с помощью искусственной нейронной сети:

- классификации дегенеративных изменений межпозвонковых дисков Pfirrmann (I–V);
 - наличию спондилолистеза;
 - степени снижения высоты межпозвонковых дисков;
 - оценке стеноза позвоночного канала;
 - наличию дефекта замыкательных пластин тел позвонков;
 - изменению сигнала от костного мозга тел позвонков (согласно классификации Modic).
- Система достигла точности 95,6% для детекции тел позвонков, ошибки были связаны с наличием переходного позвонка или низким качеством изображения [9].

Инструмент LabelImg для маркировки МРТ на предмет выпуклостей межпозвонкового диска и спондилолистеза использовался для маркировки МРТ-снимков 604 пациентов врачами с клиническим опытом. Для проверки валидности и практической применимости модели PP-YOLOv2 наборы данных МРТ по заболеваниям позвоночника были обучены на трех моделях обнаружения объектов: YOLOv3, YOLOv5 и PP-YOLOv2. Значения mAP (точность, полнота и средняя точность) для трех моделей составили 70,64, 86,66 и 90,08% соответственно. Данное программное обеспечение автоматически устанавливает вспомогательные диагнозы за 14,5 с на стандартном компьютере, значительно быстрее врачей, диагностирующих заболевания



позвоночника человека, на что обычно требуется 10 мин, а его точность, составляющая 98%, сопоставима с точностью опытных врачей при сравнении различных методов диагностики [10].

Развитие ИИ открывает возможности для быстрого, точного и стабильного анализа изображений. ИИ, обученный на обширных наборах данных, может превзойти экспертов-людей в обработке медицинских изображений. Однако эффективность таких технологий требует тщательной оценки на эталонных наборах данных, чтобы гарантировать их надежность и соответствие клиническим требованиям.

Цель исследования: оценить диагностическую точность автоматизированного выявления дегенеративных изменений пояснично-крестцового отдела позвоночника при проведении МРТ с использованием искусственного интеллекта.

Материал и методы

Основной тест (исследуемый метод) представляет собой программный продукт, использующий технологии ИИ (ИИ-сервис) и интегрированный в Единый радиологический информационный сервис Единой медицинской информационно-аналитической системы города Москвы (ЕРИС ЕМИАС) в соответствии с процедурами Московского эксперимента (эксперимент по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения этих технологий в системе здравоохранения) [11]. В ходе исследования мы анонимизировали и рандомизировали наименования ИИ-сервисов с последующим использованием условных наименований: ИИ-1, ИИ-2 и ИИ-3.

Тестирование на эталонных наборах данных – это одна из стандартных процедур Московского эксперимента, предназначенная для определения возможности доступа нового ИИ-сервиса к работе с потоком данных в условиях эксперимента,

Референс-тест: эталонный набор данных ($n = 100$) [12], подготовленный по оригинальной методологии путем экспертного пересмотра 3 врачами с опытом работы более 5 лет работы по данному виду исследований. Данные деперсонализированы в соответствии с законодательством Российской Федерации (Закон № 152-ФЗ).

Критерии включения в набор данных:

1) мужчины и женщины в возрасте от 18 до 90 лет;

2) наличие результатов МРТ пояснично-крестцового отдела позвоночника, выполненных при оказании плановой медицинской помощи в амбулаторных условиях;

3) протокол исследования: T2sag (T2-взвешенные изображения в сагиттальной плоскости), T2ax (T2-взвешенные изображения в аксиальной плоскости), толщина среза 4 мм.

Критерии не включения в набор данных:

1) двигательные артефакты, артефакты от инородных объектов на уровне исследования;

2) размеры зон интереса значительно превышают среднестатистические объемы поражения, в частности дорсальные выпячивания размером менее 3 мм;

3) наличие сопутствующих признаков патологий пояснично-крестцового отдела позвоночника (онкологических нозологий, образований, изменений воспалительного характера).

Использована ранее разработанная методология тестирования ИИ-сервисов на этапах жизненного цикла [13]. В процессе тестирования были оценены в соответствии с базовыми диагностическими требованиями (БДТ) [14] критерии, коррелирующие с дегенеративными изменениями в пояснично-крестцовом отделе позвоночника. К числу данных критериев относятся:

– наличие дорсальных выпячиваний межпозвонковых дисков, выходящих за пределы дисковых пространств в области исследования и проникающих в просвет позвоночного канала на каждом уровне с указанием размеров (дифференциальная оценка на протрузии и грыжи не проводилась).

Размеры костного позвоночного канала на сагиттальных и аксиальных срезах, а также площадь дурального мешка, ограниченного твердой мозговой оболочкой:

– передне-заднего размера дурального мешка на аксиальных изображениях,

– фронтального размера дурального мешка на аксиальных изображениях,

– передне-заднего размера дурального мешка на сагиттальных изображениях,

– площади просвета дурального мешка на уровне дисков на аксиальных срезах [15].

Под целевой патологией принимались вышеописанные признаки. Наличие хотя бы одного из перечисленных признаков считалось за патологию, отсутствие всех перечисленных признаков – за норму. В качестве результата обработки исследования ИИ-сервисы возвращали дополнительную графическую серию с маркировкой (рис. 1–4) и дополнительную серию, содержащую текстовое заключение (рис. 5). Набор данных направлялся на анализ каждому ИИ-сервису посредством тестового контура ЕРИС ЕМИАС. Критерии включения в исследование предусматривали наличие ответов от всех ИИ-сервисов. Результаты автоматизированного анализа сопоставлялись с эталонной



Рис. 1. Пример некорректной работы – не определена протрузия L4–L5.

Fig. 1. Example of incorrect operation – L4–L5 disc protrusion not detected.



Рис. 2. Пример некорректной работы – не определена протрузия L5–S1.

Fig. 2. Example of incorrect operation – L5–S1 disc protrusion not detected.



Рис. 3. Пример корректной работы – верно определены дорсальные выпячивания на уровне L4–L5 и L5–S1.

Fig. 3. Example of correct operation – dorsal protrusions at the L4–L5 and L5–S1 levels correctly identified.

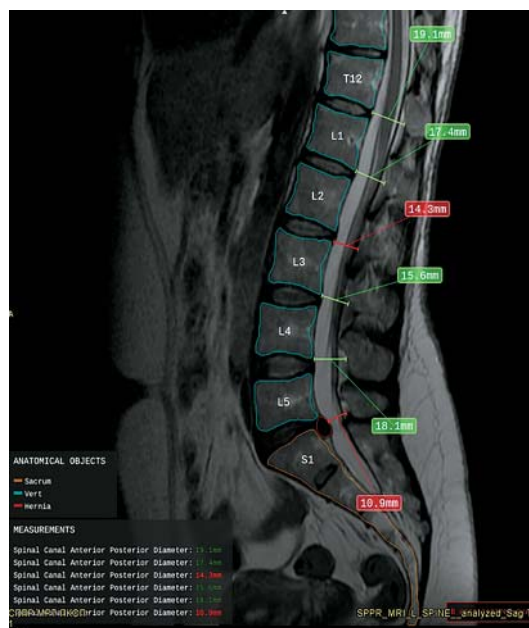
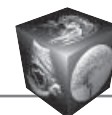


Рис. 4. Пример корректной работы – верно определено дорсальное выпячивание на уровне L5–S1.

Fig. 4. Example of correct operation – dorsal protrusion at the L5–S1 level correctly identified.



Описание	Описание
Обнаружены признаки, коррелирующие с наличием дегенеративно-дистрофических изменений пояснично-крестцового отдела позвоночника. Дорзальное выпячивание диска: L5-S1, L4-L5. Измерения передне-заднего размера выпяченных дорзальных выпячиваний дисков на сагитальных изображениях: L4-L5 размер — 4.75 мм, L5-S1 размер — 4.70 мм. Измерения передне-заднего и фронтального размеров дурального мешка на аксиальных изображениях: L1-L2 передне-задний размер — 15.95 мм, L1-L2 фронтальный размер — 22.00 мм, L2-L3 передне-задний размер — 13.75 мм, L2-L3 фронтальный размер — 22.30 мм, L3-L4 передне-задний размер — 11.15 мм, L3-L4 фронтальный размер — 18.20 мм, L4-L5 передне-задний размер — 11.15 мм, L4-L5 фронтальный размер — 14.95 мм, L5-S1 передне-задний размер — 10.15 мм, L5-S1 фронтальный размер — 11.65 мм. Измерения передне-заднего размера дурального канала на сагитальных изображениях: L1-L2 передне-задний размер — 17.10 мм, L2-L3 передне-задний размер — 17.85 мм, L3-L4 передне-задний размер — 15.80 мм, L4-L5 передне-задний размер — 13.35 мм, L5-S1 передне-задний размер — 14.00 мм. Площадь просвета дурального мешка: L1-L2 площадь — 260.70 мм², L2-L3 площадь — 226.55 мм², L3-L4 площадь — 164.55 мм², L4-L5 площадь — 143.75 мм², L5-S1 площадь — 101.20 мм².	Обнаружены признаки, коррелирующие с наличием дегенеративно-дистрофических изменений пояснично-крестцового отдела позвоночника. Дорзальное выпячивание диска: L4-L5, L3-L4, L1-L2. Измерения передне-заднего размера выпяченных дорзальных выпячиваний дисков на сагитальных изображениях: L1-L2 размер — 3.85 мм, L3-L4 размер — 3.40 мм, L4-L5 размер — 4.25 мм. Измерения передне-заднего и фронтального размеров дурального мешка на аксиальных изображениях: L1-L2 передне-задний размер — 11.65 мм, L1-L2 фронтальный размер — 21.90 мм, L2-L3 передне-задний размер — 10.90 мм, L2-L3 фронтальный размер — 19.90 мм, L3-L4 передне-задний размер — 11.25 мм, L3-L4 фронтальный размер — 16.90 мм, L4-L5 передне-задний размер — 7.15 мм, L4-L5 фронтальный размер — 13.90 мм, L5-S1 передне-задний размер — 11.25 мм, L5-S1 фронтальный размер — 16.30 мм. Измерения передне-заднего размера дурального канала на сагитальных изображениях: L1-L2 передне-задний размер — 16.40 мм, L2-L3 передне-задний размер — 15.50 мм, L3-L4 передне-задний размер — 8.95 мм, L4-L5 передне-задний размер — 6.35 мм, L5-S1 передне-задний размер — 11.90 мм. Площадь просвета дурального мешка: L1-L2 площадь — 207.55 мм², L2-L3 площадь — 168.35 мм², L3-L4 площадь — 140.75 мм², L4-L5 площадь — 82.55 мм², L5-S1 площадь — 151.45 мм².
Заключение Обнаружены признаки, коррелирующие с наличием дегенеративно-дистрофических изменений пояснично-крестцового отдела позвоночника. Дорзальное выпячивание диска: L5-S1, L4-L5.	Заключение Обнаружены признаки, коррелирующие с наличием дегенеративно-дистрофических изменений пояснично-крестцового отдела позвоночника. Дорзальное выпячивание диска: L4-L5, L3-L4, L1-L2.

Рис. 5. Примеры текстового заключения.

Fig. 5. Examples of text conclusions.

разметкой и оценивались как истинно положительные или отрицательные, ложноположительные или отрицательные.

Статистическая обработка данных

Оценка диагностической точности алгоритмов ИИ проводилась с использованием ROC-анализа. Порог активации определялся по индексу Юдена [16]. Площадь под характеристической кривой была рассчитана с 95% доверительным интервалом (ДИ) методом Делонга [17]. Для определения оптимального порогового значения использовался максимум индекса Юдена, на основе которого были определены такие метрики, как чувствительность, специфичность и точность [18]. Сравнительный анализ AUC проводился с использованием перестановочного теста [19]. Проверялась нулевая гипотеза H_0 об отсутствии различий между ROC-кривыми против альтернативной гипотезы H_1 , утверждающей наличие различий, при уровне значимости $p < 0,05$.

При тестировании программного обеспечения, основанного на технологии ИИ, было учтено, что число исследований, показывающих наибольшую неоднородность значений AUCROC (максимальное отклонение от среднего), варьирует в зависимости от соотношения классов. Баланс классов устанавливается в соответствии с имеющимся в наборе данных классов, так, при доле патологии, 10% минимальный объем выборки составляет 190, при повышении доли патологии, минимальный объем выборки уменьшается и при 50% соот-

ветствует 70. В используемом наборе данных баланс классов был 1:1, соответственно мы взяли 100 исследований с запасом (на случай возникновения технических дефектов в обработке исследований) [20].

Этическая экспертиза

Настоящее исследование выполнено в рамках Эксперимента по внедрению передовых технологий компьютерного зрения для обработки медицинских изображений с последующим использованием в московской системе здравоохранения. Проект получил одобрение этического комитета (протокол №2 НЭК МРО РОПР от 20.02.2020) и зарегистрирован в международной базе клинических исследований ClinicalTrials под идентификационным номером NCT04489992.

Результаты исследования

ИИ-сервисы: ИИ-1, ИИ-2, ИИ-3 осуществили обработку эталонного набора данных, полученные значения показателей точности приведены в табл. 1.

ИИ-1 и ИИ-3 продемонстрировали высокие и стабильные показатели точности, чувствительности, специфичности и AUCROC, что делает их наиболее надежными для использования в практическом здравоохранении. ИИ-2 требует доработки, так как его показатели значительно ниже заявленных, что может ограничивать его применение в реальных условиях. У ИИ-2 отмечено отклонение от заявленного значения чувствительности

**Таблица 1.** Оценка диагностической точности и качества ИИ-сервисов для анализа результатов МРТ пояснично-крестцового отдела с целью дегенеративных изменений**Table 1.** Assessment of the diagnostic accuracy and quality of AI services for analyzing lumbar spine MRI results for degenerative changes

ИИ-сервис / Показатели точности AI Service / Accuracy Metrics	ИИ-1 AI-1	ИИ-2 AI-2	ИИ-3 AI-3			
Заявленные значения / Declared values						
Точность / Accuracy	0.930	0.980	0.900			
Чувствительность/ Sensitivity	0.972	0.980	0.960			
Специфичность / Specificity	0.940	0.890	0.870			
AUCROC	0.920	0.970	0.950			
Результаты калибровочных тестирований / Calibration test results						
	Полученные значения Obtained values	95% ДИ 95% CI	Полученные значения Obtained values	95% ДИ 95% CI	Полученные значения Obtained values	95% ДИ 95% CI
Точность / Accuracy	0.975	0.966–1.000	0.865	0.833–0.955	0.970	0.975–1.000
Чувствительность/ Sensitivity	0.970	0.940–1.000	0.890	0.830–0.950	0.980	0.950–1.000
Специфичность / Specificity	0.980	0.940–1.000	0.840	0.740–0.940	0.960	0.910–1.000
AUCROC	0.960	0.910–1.000	0.940	0.870–1.000	1.000	1.000–1.000

на 10%. Таким образом, результаты исследования подчеркивают важность калибровочного тестирования для оценки реальной эффективности ИИ-сервисов в медицинской диагностике.

Примеры корректной и некорректной работы ИИ-сервисов представлены на рис. 1–4.

Обсуждение

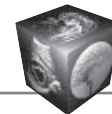
Дегенеративно-дистрофические заболевания позвоночника остаются одной из ключевых проблем современной медицины, приводя к хроническому болевому синдрому, снижению качества жизни и значительным экономическим затратам на лечение. В настоящее время МРТ – это лучший доступный инструмент визуализации для просмотра заболеваний мягких тканей, играющий роль в ведении болей в спине и планировании хирургического лечения [21]. Использование МРТ в медицине растет из-за клинических преимуществ и повышения безопасности пациентов.

По данным многопрофильных исследований оценки хронометража среднее время интерпретации МРТ в ряде работ составляет порядка 15–20 мин и может увеличиваться при анализе более сложных исследований, тогда как для компьютерной томографии и рентгенографии оно, как правило, меньше и нередко ограничивается несколькими минутами [22]. Увеличение сложности протоколов и числа последовательностей в стандартных

МРТ-исследованиях приводит к дополнительной когнитивной нагрузке и повышает риск диагностических ошибок при дефиците времени.

Проведенное ретроспективное диагностическое исследование трех ИИ-сервисов, в котором мы самостоятельно рассчитывали оптимальный порог, выявило существенные различия в их диагностической точности при выявлении дегенеративных изменений пояснично-крестцового отдела позвоночника. ИИ-1 продемонстрировал высокую точность (0,975), превысив заявленные значения (0,93). Чувствительность (0,97) и специфичность (0,98) также оказались на высоком уровне, что подтверждает его надежность. Однако AUCROC (0,96) несколько уступает ИИ-3, что может указывать на менее стабильную работу при различной степени выраженности патологий. ИИ-2 показал снижение точности (0,865 против заявленных 0,98), а также чувствительности (0,89 против 0,98). ИИ-3 проявил себя как наиболее стабильный и точный алгоритм: точность 0,97 (при заявленных 0,9); чувствительность 0,98 (практически идеальная); специфичность 0,96 (высокая способность исключать ложноположительные случаи); AUCROC 1,0 (максимальное возможное значение, что свидетельствует о выдающейся диагностической эффективности).

Сервис ИИ-3 может рассматриваться как наиболее надежный инструмент для автоматизированного анализа МРТ позвоночника, демон-



стрируя превосходство не только над другими ИИ-сервисами, но и над заявленными характеристиками. ИИ-1 также показал высокие результаты, но требует дополнительной валидации на более разнообразных клинических данных. ИИ-2 оказался менее воспроизводимым в реальных условиях, что подчеркивает необходимость доработки алгоритма. Различия в точности между сервисами указывают на важность тщательного тестирования.

Научные группы, как в России, так и за рубежом, работают над созданием ИИ-систем для диагностики ДДЗП по МРТ-изображениям. Как показано в систематическом обзоре R. Compte и соавт. (2023), существующие разработки используют методы машинного обучения для анализа: дегенеративных изменений межпозвонковых дисков, грыж различного типа, протрузий, патологических изменений Modic. Были включены модели, использующие глубокое обучение, опорные векторные машины. Из 27 включенных исследований 22 были направлены на разработку алгоритма, 2/27 сообщили о разработке и внешней проверке [23] и 3/27 были сосредоточены исключительно на внешней проверке [24]. Большинство исследований (24/27) использовали ретроспективные, уже существующие наборы данных из больничных или университетских коллекций; однако 4/27 исследований (24/27) использовали проспективно изучали сканы пациентов. 16 исследований с наборами данных изображений использовали алгоритмы глубокого обучения. Модели глубокого обучения в среднем имели точность 87,0% (SD 7,0%), специфичность 90,4% (SD 6,3%) и чувствительность 88,2% (SD 7,2%). 8/25 исследований сравнивали производительность нескольких классических алгоритмов машинного обучения, основанных на глубоких нейронных сетях (DL). Другие исследования использовали неспецифические алгоритмы, которые были указаны как пользовательские модели (2/25). Большинство исследований, включенных в этот систематический обзор, использовали модели глубокого обучения для классификации или градации дегенеративных изменений межпозвонкового диска по критериям Пфиррмана или МК. Можно ожидать, что модели глубокого обучения, состоящие из сложных многослойных входных сетей, превзойдут другие алгоритмы для чтения МРТ позвоночника. Это ожидание было отражено в выборе многих авторов исследований в этом обзоре в пользу использования моделей глубокого обучения. Однако результаты исследования не показывают различий в производительности DL по сравнению с другими подходами МО. Алгоритмы глубокого обучения, как правило, имеют высокий уро-

вень ложноположительных ошибок и их трудно калибровать [25].

Проведенное нами исследование можно классифицировать как внешнюю независимую валидацию на ограниченном наборе данных. Заявленные разработчиком значения являются внутренней валидацией компании разработчика. Исследование подтвердило высокую обобщающую способность разработанных алгоритмов: их диагностическая точность не снизилась при внешней валидации, причем для ИИ-3 показатели оказались выше верхней границы 95% ДИ. Важно понимать, что подобные испытания на эталонных данных представляют собой только начальную стадию внедрения ИИ-решений в систему здравоохранения.

Подавляющее большинство опубликованных исследований не включает апробацию алгоритмов ИИ в реальной клинической практике. Разработчики программного обеспечения, основанного на технологиях ИИ, часто не раскрывают метрики диагностической точности своих продуктов. Существующие классификации программного обеспечения на основе ИИ, разработанные другими авторами, сосредоточены на технических характеристиках систем, в то время как клиническая применимость остается вне их внимания [26]. В связи с этим остаются нерешенными ключевые вопросы, такие как воспроизводимость результатов ИИ на новых данных, возможность масштабирования внедрения, клиническая значимость полученных показателей. В качестве примера возьмем данные нашего исследования: у одного из ИИ-сервисов расхождение между заявленной и фактической точностью около 10%. Это свидетельствует о том, что при работе с эталонным набором, не входившими в обучающую выборку, диагностическая эффективность алгоритма может существенно снижаться. В связи с этим основной задачей дальнейших исследований мы считаем оценки стабильности и диагностической точности ИИ-сервисов в условиях рутинной работы отделений лучевой диагностики, в частности при выявлении дегенеративных изменений пояснично-крестцового отдела позвоночника. Только после решения этой задачи можно переходить к анализу результативности и эффективности подобных технологий в клинической практике. Дальнейшее развитие исследований должно быть направлено на расширение выборки пациентов, оценку точности проводимых измерений и воспроизводимости алгоритмов при потоковой обработке МР-исследований и создание систем постоянного мониторинга качества работы ИИ-сервисов в условиях реального клинического процесса.



Заключение

Проведенное исследование автоматизированного выявления дегенеративных изменений пояснично-крестцового отдела позвоночника при проведении МРТ с применением ИИ на основе результатов калибровочного тестирования выявило, что этап внешней оценки ИИ-сервисов играет решающую роль в объективном подтверждении их диагностической эффективности. Сопоставление заявленных и фактических метрик показало расхождение до 10%, демонстрируя необходимость системной процедуры валидации, обеспечивающей надежную оценку качества медицинского программного обеспечения. Полученные значения метрик диагностической точности подтверждают значительный потенциал ИИ-технологий для повышения эффективности диагностики дегенеративных заболеваний позвоночника, что может способствовать сокращению времени анализа изображений, повышению точности интерпретации и снижению влияния человеческого фактора при рутинной работе рентгенологов.

Источник финансирования

Данная статья подготовлена авторским коллективом в рамках НИР "Научные методологии устойчивого развития технологий искусственного интеллекта в медицинской диагностике" (№ ЕГИСУ: № 123031500004-5) в соответствии с Приказом от 17.12.2024 № 1184 "Об утверждении государственных заданий, финансовое обеспечение которых осуществляется за счет средств бюджета города Москвы государственным бюджетным (автономным) учреждениям подведомственным Департаменту здравоохранения города Москвы, на 2025 год и плановый период 2026 и 2027 годов" Департамента здравоохранения города Москвы.

Конфликт интересов

Авторы заявляют об отсутствии как явных, так и потенциальных конфликтов интересов, которые могли бы повлиять на содержание данной публикации.

Вклад авторов

Все указанные авторы соответствуют критериям Международного комитета редакторов медицинских журналов (ICMJE): участвовали в разработке концепции исследования, внесли значимый вклад в проведение работы, принимали активное участие в подготовке рукописи. Каждый автор ознакомился с финальной версией статьи и дал согласие на ее публикацию.

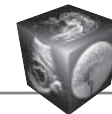
Authors' participation

All listed authors meet the criteria of the International Committee of Medical Journal Editors (ICMJE): they

participated in the study conception, made a significant contribution to the conduct of the work, and actively participated in the preparation of the manuscript. Each author reviewed the final version of the article and consented to its publication.

Список литературы [References]

1. Тюрников В.М. Дегенеративно-дистрофические поражения позвоночника: диагностика, клиника и лечение. *РМЖ*. 2008; 26: 1739.
Tyurnikov V.M. Degenerative-dystrophic lesions of the spine: diagnosis, clinical features, and treatment. *RMJ*. 2008; 26: 1739. (In Russian)
2. Подчуфарова Е.В., Яхно Н.Н., Алексеев В.В., Аведисова А.С., Чахава К.О., Ершова Е.М., Протасенко Т.В. Хронические болевые синдромы пояснично-крестцовой локализации: значение структурных скелетно-мышечных расстройств и психологических факторов. *Боль*. 2003; 1: 38–43.
Podchufarova E.V., Yakhno N.N., Alekseev V.V. et al. Chronic pain syndromes of lumbosacral localization: the role of structural musculoskeletal disorders and psychological factors. *Pain*. 2003; 1: 38–43. (In Russian)
3. Кренева Ю.А., Панов В.П., Авдеева М.В., Богомолова К.А. Восстановительный потенциал комплексных реабилитационных мероприятий при оказании первичной медико-санитарной помощи больным с дегенеративными заболеваниями позвоночника. *Профилактическая и клиническая медицина*. 2018; 2 (67): 76–82.
Kreneva Yu.A., Panov V.P., Avdeeva M.V., Bogomolova K.A. Restorative potential of complex rehabilitation in the primary health care provision for patients with degenerative disc diseases. *Preventive and Clinical Medicine*. 2018; 2 (67): 76–82. (In Russian)
4. Орел А.М., Семенова О.К. Частота дегенеративно-дистрофических поражений суставов позвоночника у пациентов молодого и среднего возраста с дорсопатией. *Российский остеопатический журнал*. 2019; 3–4: 30–36. <https://doi.org/10.32885/2220-0975-2019-3-4-30-36>
Orel A.M., Semenova O.K. The frequency of degenerative-dystrophic lesions of the joints of the spine in young and middle-aged patients with dorsopathy. *Russian Osteopathic Journal*. 2019; 3–4: 30–36. <https://doi.org/10.32885/2220-0975-2019-3-4-30-36> (In Russian)
5. Вышлова И.А., Карпов С.М., Стародубцев А.И. Вертеброгенные болевые синдромы поясничного уровня: эпидемиология, клинические проявления. *Медицинский вестник Северного Кавказа*. 2016; 4: 586–587. <https://doi.org/10.14300/mnnc.2016.00142>
Vishlova I., Karpov S., Starodubtsev A. Low back pain syndrome: epidemiology and clinical manifestations. *Medical News of North Caucasus*. 2016; 4: 586–587. <https://doi.org/10.14300/mnnc.2016.00142> (In Russian)
6. Gholamrezanezhad A., Basques K., Batouli A. et al. Clinical Nononcologic Applications of PET/CT and PET/MRI in Musculoskeletal, Orthopedic, and Rheumatologic Imaging. *Am. J. Roentgenol*. 2018; 210 (6): W245–W263. <https://doi.org/10.2214/AJR.17.18523>
7. Методические рекомендации по разработке инструкций по охране труда для персонала кабинета (отделения) магнитно-резонансной томографии: Методические рекомендации. М.: Государственное бюджет-



- ное учреждение здравоохранения города Москвы "Научно-практический центр медицинской радиологии Департамента здравоохранения города Москвы", 2017. 32 с. (Лучшие практики лучевой и инструментальной диагностики). https://telemedai.ru/media/uploads/2023/02/21/44-_.pdf
- Guidelines for the development of occupational safety instructions for personnel of magnetic resonance imaging (MRI) units (departments) [Internet]. Moscow: Research and Practical Center of Medical Radiology, Moscow Healthcare Department; 2017. 32 p. https://telemedai.ru/media/uploads/2023/02/21/44-_.pdf (In Russian)
8. Hajiheydari N., Delgosha M.S., Saheb T. AI in medical diagnosis: A contextualised study of patient motivations and concerns. *Soc. Sci. Med.* 2025; 371: 117850. <https://doi.org/10.1016/j.socscimed.2025.117850>
 9. Xuan J., Ke B., Ma W. et al. Spinal disease diagnosis assistant based on MRI images using deep transfer learning methods. *Front. Public Health.* 2023; 11: 1044525. <https://doi.org/10.3389/fpubh.2023.1044525>
 10. Su Z.H., Liu J., Yang M.S. et al. Automatic Grading of Disc Herniation, Central Canal Stenosis and Nerve Roots Compression in Lumbar Magnetic Resonance Image Diagnosis. *Front. Endocrinol. (Lausanne).* 2022; 13: 890371. <https://doi.org/10.3389/fendo.2022.890371>
 11. Эксперимент по использованию технологий искусственного интеллекта для анализа медицинских изображений [Электронный ресурс]. *MosMed.AI.* ГБУЗ "Научно-практический клинический центр диагностики и телемедицинских технологий ДЗМ". <https://mosmed.ai/ai/>
Experiment on the use of artificial intelligence technologies for the analysis of medical images [Electronic resource]. *MosMed.AI.* State Budgetary Healthcare Institution 'Scientific and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Department of Health'. URL: <https://mosmed.ai/ai/> (In Russian)
 12. Бобровская Т.М., Васильев Ю.А., Никитин Н.Ю., Арзамасов К.М. Подходы к формированию наборов данных в лучевой диагностике. *Врач и информационные технологии.* 2023; 4: 14–23. https://doi.org/10.25881/18110193_2023_4_14
Bobrovskaya T.M., Vasiliev Yu.A., Nikitin N.Yu., Arzamasov K.M. Approaches to building radiology datasets. *Medical Doctor and Information Technologies.* 2023; 4: 14–23. https://doi.org/10.25881/18110193_2023_4_14 (In Russian)
 13. Васильев Ю.А., Владимирский А.В., Омелянская О.В., Арзамасов К.М., Четвериков С.Ф., Румянцев Д.А., Зеленова М.А. Методология тестирования и мониторинга программного обеспечения на основе технологий искусственного интеллекта для медицинской диагностики. *Digital Diagnostics.* 2023; 4 (3): 252–267. <https://doi.org/10.17816/DD321971>
Vasiliev Yu.A., Vladzimirsky A.V., Omelyanskaya O.V. et al. Methodology for testing and monitoring artificial intelligence-based software for medical diagnostics. *Digital Diagnostics.* 2023; 4 (3): 252–267. <https://doi.org/10.17816/DD321971> (In Russian)
 14. Документы эксперимента по использованию технологий искусственного интеллекта для анализа медицинских изображений [Электронный ресурс]. *MosMed.AI.* ГБУЗ "Научно-практический клинический центр диагностики и телемедицинских технологий ДЗМ". URL: <https://mosmed.ai/ai/docs/>
Documents of the experiment on the use of artificial intelligence technologies for the analysis of medical images [Electronic resource]. *MosMed.AI.* State Budgetary Healthcare Institution 'Scientific and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Department of Health'. URL: <https://mosmed.ai/ai/docs/> (In Russian)
 15. Свидетельство о государственной регистрации базы данных № 2023621223 Российская Федерация. *MosMedData: MPT с признаками патологий пояснично-крестцового отдела позвоночника: № 2023620789: заявл. 24.03.2023; опублик. 17.04.2023 / Ю.А. Васильев, Е.В. Туравилова, А.В. Владимирский и др.; заявитель Государственное бюджетное учреждение здравоохранения города Москвы "Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы".* <https://mosmed.ai/en/datasets/otkritie-dlya-self-testov/mosmeddata-mrt-s-priznakami-patologii-poyasnichnokresttsovogo-otdela-pozvonochnika-tip-iii/> Certificate of State Registration of a Database No. 2023621223, Russian Federation. *MosMedData: MRI with signs of pathologies of the lumbosacral spine: No. 2023620789; application dated March 24, 2023; published April 17, 2023 / Yu.A. Vasiliev, E.V. Turavilova, A.V. Vladzimirsky et al.; applicant: State Budgetary Healthcare Institution of the City of Moscow "Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Healthcare Department."* (In Russian)
 16. Клинические испытания систем искусственного интеллекта (лучевая диагностика). М.: Государственное бюджетное учреждение здравоохранения города Москвы "Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы", 2023. 40 с. https://mosmed.ai/documents/294/web_Клинические_испытания_систем_искусственного_интеллекта_Pxp0V2i.pdf
Clinical Trials of Artificial Intelligence Systems (Radiology). Moscow: State Budgetary Healthcare Institution of Moscow "Scientific and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Department of Healthcare", 2023. 40 p. https://mosmed.ai/documents/294/web_Клинические_испытания_систем_искусственного_интеллекта_Pxp0V2i.pdf (In Russian)
 17. Sun X., Xu W. Fast implementation of DeLong's algorithm for comparing the areas under correlated receiver operating characteristic curves. *IEEE Transactions on Signal Processing.* 2014; 21 (11): 1389–1393. <https://doi.org/10.1109/LSP.2014.2337313>
 18. Pauly M., Asendorf T., Konietzschke F. Permutation-based inference for the AUC: A unified approach for continuous and discontinuous data. *Biom. J.* 2016; 58 (6): 1319–1337. <https://doi.org/10.1002/bimj.201500105>
 19. Бобровская Т.М., Васильев Ю.А., Никитин Н.Ю., Владимирский А.В., Омелянская О.В., Четвериков С.Ф., Арзамасов К.М. Объем выборки для оценки диагностической точности программного обеспечения на основе технологий искусственного интеллекта в лучевой диагностике. *Сибирский журнал*



- клинической и экспериментальной медицины. 2024; 39 (3): 188–198. <https://doi.org/10.29001/2073-8552-2024-39-3-188-198>
- Bobrovskaya T.M., Vasiliev Yu.A., Nikitin N.Yu. et al. Sample Size for Assessing the Diagnostic Accuracy of Artificial Intelligence-Based Software in Radiology. *Siberian Journal of Clinical and Experimental Medicine*. 2024; 39 (3): 188–198. <https://doi.org/10.29001/2073-8552-2024-39-3-188-198> (In Russian)
20. Nahm F.S. Receiver operating characteristic curve: overview and practical use for clinicians. *Korean J. Anesthesiol.* 2022; 75 (1): 25–36. <https://doi.org/10.4097/kja.21209>
21. Low back pain and sciatica in over 16s: assessment and management. London: National Institute for Health and Care Excellence (NICE); 2020 Dec 11. <https://www.ncbi.nlm.nih.gov/books/NBK562933/>
22. Kwee R.M., Amasha A.A.H., Kwee T.C. Reading Times of Common Musculoskeletal MRI Examinations: A Survey Study. *Tomography*. 2024; 10 (9): 1527–1533. <https://doi.org/10.3390/tomography10090112>
23. Zheng H.D., Sun Y.L., Kong D.W. et al. Deep learning-based high-accuracy quantitation for lumbar intervertebral disc degeneration from MRI. *Nat. Commun.* 2022; 13 (1): 841. <https://doi.org/10.1038/s41467-022-28387-5>
24. Grob A., Loibl M., Jamaludin A. et al. External validation of the deep learning system “SpineNet” for grading radiological features of degeneration on MRIs of the lumbar spine. *Eur. Spine J.* 2022; 31 (8): 2137–2148. <https://doi.org/10.1007/s00586-022-07311-x>
25. Jamaludin A., Kadir T., Zisserman A. SpineNet: Automated classification and evidence visualization in spinal MRIs. *Med. Image Anal.* 2017; 41: 63–73. <https://doi.org/10.1016/j.media.2017.07.002>
26. Bichu Y.M., Hansa I., Bichu A.Y. et al. Applications of artificial intelligence and machine learning in orthodontics: a scoping review. *Prog. Orthod.* 2021; 22 (1): 18. <https://doi.org/10.1186/s40510-021-00361-9>

Для корреспонденции*: Тхакахова Регина Альбертовна – e-mail: TkhakakhovaRA@zdrav.mos.ru

Тхакахова Регина Альбертовна – аспирант, врач-рентгенолог ГБУЗ “НПКЦ ДиТ ДЗМ”, Москва. <https://orcid.org/0009-0006-0943-2223>

Владимирский Антон Вячеславович – доктор мед. наук, заместитель директора по научной работе ГБУЗ “НПКЦ ДиТ ДЗМ”, Москва. <https://orcid.org/0000-0002-2990-7736>

Арзамасов Кирилл Михайлович – доктор мед. наук, руководитель отдела медицинской информатики, радиомики и радиогеномики, ГБУЗ “НПКЦ ДиТ ДЗМ”, Москва; доцент кафедры лучевой диагностики и лучевой терапии ФГБОУ ВО “Самарский государственный медицинский университет” Минздрава России, Самара. <https://orcid.org/0000-0001-7786-0349>

Contact*: Regina A. Tkhakakhova – e-mail: TkhakakhovaRA@zdrav.mos.ru

Regina A. Tkhakakhova – Radiologist, Postgraduate Student, Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department, Moscow. <https://orcid.org/0009-0006-0943-2223>

Anton V. Vladzmyrskyy – Doct. of Sci. (Med.), Deputy Director for Research, Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department, Moscow. <https://orcid.org/0000-0002-2990-7736>

Kirill M. Arzamasov – Doct. of Sci. (Med.), Functional Diagnostics Physician, Cybernetics Physician, Head of the Department of Medical Informatics, Radiomics, and Radiogenomics at the Moscow Center for Diagnostics and Telemedicine, Moscow; Associate Professor, Department of Radiology and Radiation Therapy, Samara State Medical University, Samara. <https://orcid.org/0000-0001-7786-0349>