

doi: 10.17586/2226-1494-2026-26-5-1000-1008

УДК 004.85:616-07

Дообучение мультимодальной языковой модели для многоклассовой диагностики по МРТ-изображениям

Наталья Федоровна Гусарова^{1✉}, Наталья Викторовна Добренко²,
Алексей Александрович Зубаненко³, Иван Вячеславович Томилов⁴,
Павел Геннадьевич Ульянов⁵

^{1,2,3,4,5} Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

¹ natfed@list.ru✉, <https://orcid.org/0000-0002-1361-6037>

² graziokisa@yandex.ru, <https://orcid.org/0000-0001-6206-8033>

³ zubdocmri@gmail.com, <https://orcid.org/0000-0001-6953-5239>

⁴ ivan-tomilov3@yandex.ru, <https://orcid.org/0000-0003-1886-2867>

⁵ pavel.ulyanov@itmo.ru, <https://orcid.org/0009-0008-0641-901X>

Аннотация

Введение. В современной радиологии диагностические ассистенты искусственного интеллекта (ИИ), классифицирующие изображения магнитно-резонансной томографии (МРТ) головного мозга, как правило, основаны на сверточных нейронных сетях (Convolutional Neural Networks, CNN), которые работают исключительно с визуальным сигналом и не способны использовать структурированные диагностические описания патологий. Мультимодальные языковые модели (Vision Language Models, VLM) открывают новые возможности: совместная обработка визуальных и текстовых данных позволяет учитывать семантический контекст патологий, недоступный CNN-классификаторам. Ключевым препятствием для их применения остается высокая ресурсоемкость дообучения, несопоставимая с возможностями обычной клиники. **Метод.** Предложен метод построения ресурсоэффективного диагностического ИИ-ассистента на базе VLM Gemma 3 (4b-it) с применением метода низкоранговой адаптации (Low-Rank Adaptation, LoRA). **Основные результаты.** Дообучение проведено на наборе данных из 91 777 МРТ-снимков головного мозга, полученных в режиме подавления сигнала свободной жидкости (Fluid-Attenuated Inversion Recovery, FLAIR), распределенных по 8 классам патологий. Обучались только низкоранговые адаптеры — 12,5 % параметров модели, что позволило провести обучение на одном графическом процессоре NVIDIA RTX 3090 с 24 ГБ видеопамати без масочной разметки данных. Экспертные знания врача-радиолога интегрированы через системный промпт, задающий модели роль опытного нейрорадиолога с явными критериями анализа. На валидационной выборке (9178 примеров) модель достигла: Accuracy 0,930, F1-score 0,929, Specificity 0,989 и Cohen's Kappa 0,915. **Обсуждение.** Сравнение с базовым классификатором ConvNeXt-Base на идентичном наборе данных показало превосходство VLM+LoRA по всем метрикам: прирост Accuracy составил 7,4 процентных пунктов, Cohen's Kappa — +0,090. Полученные результаты подтверждают практическую применимость предложенного подхода для клиник с ограниченными вычислительными ресурсами.

Ключевые слова

мультимодальная языковая модель, диагностический ИИ-ассистент, МРТ головного мозга, FLAIR-последовательность, низкоранговая адаптация, многоклассовая классификация, патологии головного мозга, ресурсоэффективное обучение, Gemma, LoRA

Благодарности

Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации, в рамках государственного задания FSER-2025-0013.

Ссылка для цитирования: Гусарова Н.Ф., Добренко Н.В., Зубаненко А.А., Томилов И.В., Ульянов П.Г. Дообучение мультимодальной языковой модели для многоклассовой диагностики по МРТ-изображениям // Научно-технический вестник информационных технологий, механики и оптики. 2026. Т. 26, № 5. С. 1000–1008. doi: 10.17586/2226-1494-2026-26-5-1000-1008

Training visual language models for MRI-based diagnostics

Natalia F. Gusarova¹, Natalia V. Dobrenko², Alexey A. Zubanenko³,
Ivan V. Tomilov⁴, Pavel G. Ulyanov⁵

^{1,2,3,4,5} ITMO University, Saint Petersburg, 197101, Russian Federation

¹ natfed@list.ru, <https://orcid.org/0000-0002-1361-6037>

² grazziokisa@yandex.ru, <https://orcid.org/0000-0001-6206-8033>

³ zubdocmri@gmail.com, <https://orcid.org/0000-0001-6953-5239>

⁴ ivan-tomilov3@yandex.ru, <https://orcid.org/0000-0003-1886-2867>

⁵ pavel.ulyanov@itmo.ru, <https://orcid.org/0009-0008-0641-901X>

Abstract

In modern radiology, diagnostic Artificial Intelligence assistants for classifying Magnetic Resonance Imaging (MRI) brain images are typically based on Convolutional Neural Networks (CNNs), which work exclusively with the visual signal and are unable to use structured diagnostic descriptions of pathologies. Vision Language Models (VLMs) open up new possibilities: joint processing of visual and textual data allows for the consideration of the semantic context of pathologies, which is inaccessible to CNN classifiers. The key obstacle to their application remains the high resource intensity of retraining, which is incomparable with the capabilities of a typical clinic. This paper proposes a method for constructing a resource-efficient diagnostic AI assistant based on VLM Gemma3 (4b-it) using low-rank adaptation (LoRA). Retraining was performed on a dataset of 91,777 Fluid-Attenuated Inversion Recovery MRI brain images, distributed across eight pathology classes. Only low-rank adapters were trained — 12.5 % of the model parameters, which allowed training on a single NVIDIA RTX3090 Graphics Processing Unit with 24 GB of Video Random Access Memory without masked data annotation. The radiologist's expert knowledge was integrated via a system prompt, which assigns the model the role of an experienced neuroradiologist with explicit analysis criteria. On the validation set (9178 examples), the model achieved: Accuracy 0.930, F1-score 0.929, Specificity 0.989, and Cohen's Kappa 0.915. Comparison with the baseline ConvNeXt-Base classifier on an identical dataset showed the superiority of VLM+LoRA across all metrics: Accuracy increased by 7.4 percentage points, and Cohen's Kappa by 0.090. The obtained results confirm the practical applicability of the proposed approach for clinics with limited computing resources.

Keywords

multimodal language model, diagnostic AI assistant, brain MRI, FLAIR sequence, low-rank adaptation, multi-class classification, brain pathologies, resource-efficient learning, Gemma, LoRA

Acknowledgements

This work was supported by the Ministry of Science and Higher Education of the Russian Federation, Goszadanie (State Assignment) FSER-2025-0013.

For citation: Gusarova N.F., Dobrenko N.V., Zubanenko A.A., Tomilov I.V., Ulyanov P.G. Training visual language models for MRI-based diagnostics. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2026, vol. 26, no. 5, pp. 1000–1008 (in Russian). doi: 10.17586/2226-1494-2026-26-5-1000-1008

Введение

В современной высокотехнологичной медицине все большую роль приобретают ассистенты искусственного интеллекта (ИИ) — программные средства на основе ИИ, которые играют роль цифрового помощника врача [1]. Для врачей-радиологов особенно востребованы диагностические ИИ-ассистенты (Diagnostic Assistive Artificial Intelligence, DAAI) [2, 3], которые призваны классифицировать изображения магнитно-резонансной томографии (МРТ), выдавать вероятности диагнозов, указывать зоны внимания, сопоставлять находки с клиническими факторами. Задача DAAI — не заменить врача, а предоставить специалисту дополнительные данные для постановки окончательного диагноза [4], в связи с чем особое значение приобретает объяснимость и консистентность данных, формируемых DAAI.

В реальной врачебной практике окончательный диагноз ставится на основе агрегирования всей доступной информации о пациенте, такой как анамнез пациента, записи его медицинской карты, анализ медицинских изображений и результатов лабораторных анализов. Однако большинство известных DAAI радиологического профиля работают в унимодальном режиме, т. е. обрабатывают информацию только одного типа. При

этом в качестве обработчиков визуальной информации используются различные варианты сверточных нейронных сетей (Convolutional Neural Networks, CNN), которые обеспечивают высокие уровни точности классификации (порядка 90 % и более) [5], а обработка текстовой информации все активнее возлагается на большие языковые модели (Large Language Models, LLM) [6]. Мультимодальные языковые модели (Vision Language Models, VLM) принципиально изменяют процесс обработки медицинских изображений: там, где CNN видит только интенсивность пикселей, VLM способна опереться на структурированные диагностические описания, заложенные в промпт, что позволяет врачу получить более согласованные и объяснимые результаты DAAI [7, 8].

Ключевым препятствием для применения крупных VLM как базы для построения DAAI остается высокая ресурсоемкость [9–11]: для обучения используются наборы данных объемом в десятки тысяч изображений, маски для которых размечены специалистами-радиологами, что несопоставимо с возможностями обычной клиники, и серьезные вычислительные ресурсы. Например, в [9] использованы четыре графических процессора NVIDIA A100, что инспирирует недопустимо высокий углеродный след [12]. Для снижения ресурсо-

емкости обучения модели предложен механизм низкоранговой адаптации LLM (Low-Rank Adaptation, LoRA) [13], который на порядки снижает число обучаемых параметров. Тем не менее, проблема построения DAAI на вычислительных мощностях, доступных конечным потребителям, не теряет своей остроты. Особенно это характерно для медицинских клиник, где требуется постоянное дообучение моделей под конкретные заболевания и группы пациентов.

В настоящей работе показано, что, используя LoRA-дообучение VLM на наборе данных, составленном из МРТ-изображений головного мозга без масочной разметки, в режиме многоклассовой классификации, можно получить значения точности, сопоставимые с классификаторами на основе CNN.

Детально описана методика интеграции экспертных знаний врача-радиолога в процесс дообучения через разметку и корректировку промпта. Продemonстрирована эффективность применения LoRA для модели Gemma 3 в специфическом медицинском домене при ограничении объема памяти видеокарты (Video Random Access Memory, VRAM) до 24 Гб.

Обзор работ

Методы классификации патологий МРТ-изображений мозга. Системы автоматической диагностики МРТ-изображений мозга активно развиваются на протяжении последнего десятилетия. Актуально достигаемый уровень эффективности представлен в [5], где мультимодальные входные данные, сегментация 3D U-Net и байесовская интеграция признаков обеспечили точность многоклассовой классификации до 91 % по 19 нозологиям. Превосходство над изолированными CNN здесь достигнуто за счет объединения нескольких последовательностей и пространственного контекста, однако ценой сложного конвейера обучения и высоко-ресурсной инфраструктуры с несколькими графическими процессорами. В [7] предложен мультимодальный подход к диагностике заболеваний мозга, объединяющий визуальные признаки с текстовыми клиническими данными, и показано, что такая интеграция повышает точность дифференциальной диагностики. В [8] на примере рака простаты продемонстрировано, что мультимодальный ИИ, объединяющий изображения и клинические данные, стабильно превосходит модели, обрабатывающие только визуальные признаки.

Среди CNN-архитектур для классификации МРТ-изображений особое место занимает ConvNeXt: систематическое включение принципов трансформеров в сверточную архитектуру позволяет ConvNeXt конкурировать с Vision Transformer при значительно меньших вычислительных затратах [14]. Тем не менее, будучи сугубо одномодовой архитектурой, ConvNeXt не использует семантические описания патологий — ограничение, которое VLM преодолевает принципиально.

Мультимодальные языковые модели в медицинской визуализации. Применение VLM для анализа медицинских изображений привлекает все большее внимание. В [15] показана эффективность VLM для

классификации медицинских изображений в режиме few-shot даже при малых наборах данных. В [16] предложена методика дообучения VLM для медицинских изображений: LoRA-адаптеры, внедренные во все линейные слои, обеспечивают лучшую адаптацию, чем дообучение только классификационных слоев. В [17] показано, что проблемно-ориентированное предобучение LLM на парах «изображение-описание» существенно улучшает качество классификации медицинских изображений. В работе [18] систематизированы приложения мультимодальных LLM в здравоохранении, при этом выделена диагностика, генерация отчетов и вопросно-ответные системы с поддержкой визуального входа как ключевые направления.

Анализ мультимодальных LLM [11] показал, что современные VLM превосходят унимодальные модели на задачах с семантически сложными визуальными паттернами, к которым, в частности, относятся задачи интерпретации МРТ-изображений головного мозга. В [19] продемонстрирована возможность развертывания эффективных VLM уровня GPT-4V на устройствах с ограниченными ресурсами, тем самым обосновано применение компактных моделей класса Gemma-3-4b-it в клинических условиях.

Параметрически эффективное дообучение. В [13] предложен механизм низкоранговой адаптации LoRA, в котором вместо обновления всех весов LLM к линейным слоям добавляются пары низкоранговых матриц, снижая число обучаемых параметров на два порядка при сохранении качества адаптации. В [20] предложена методика мультимодальной генерации медицинских изображений и показана ее эффективность при ограниченных данных. В то же время в [12] изучен вопрос об энергетической стоимости обучения крупных нейросетей и обоснована приоритетность ресурсоэффективных методов — аргумент, напрямую релевантный для клинических учреждений.

Совокупность перечисленных работ формирует базу для используемого в настоящей работе подхода к дообучению мультимодальной языковой модели на основе LoRA на одном графическом процессоре как конкурентоспособной альтернативы тяжелым CNN-архитектурам в задаче диагностики патологий по МРТ-изображениям головного мозга.

Метод и материалы

Схема и гиперпараметры обучающей процедуры. Общий пайплайн для дообучения модели представлен на рис. 1.

В качестве базовой модели использована открытая мультимодальная языковая модель Gemma-3-4b-it с поддержкой изображений. Для адаптации модели к задаче мультиклассовой диагностики патологий головного мозга по последовательностям с подавлением сигнала свободной жидкости (Fluid-Attenuated Inversion Recovery, FLAIR) применен метод LoRA с параметрами: ранг низкоранговых матриц r равен 16; параметр масштабирования α равен 16; вероятность случайного отключения входных активаций в ветви адаптера dropout равна 0,05. Низкоранговые адаптеры внедря-

лись во все линейные слои модели. При дообучении параметры базовой модели оставались замороженными, обновлялись только параметры адаптеров.

Для дополнительного снижения потребления памяти обучение выполнялось с использованием смешанной точности bfloat16 и градиентного чекпоинтинга.

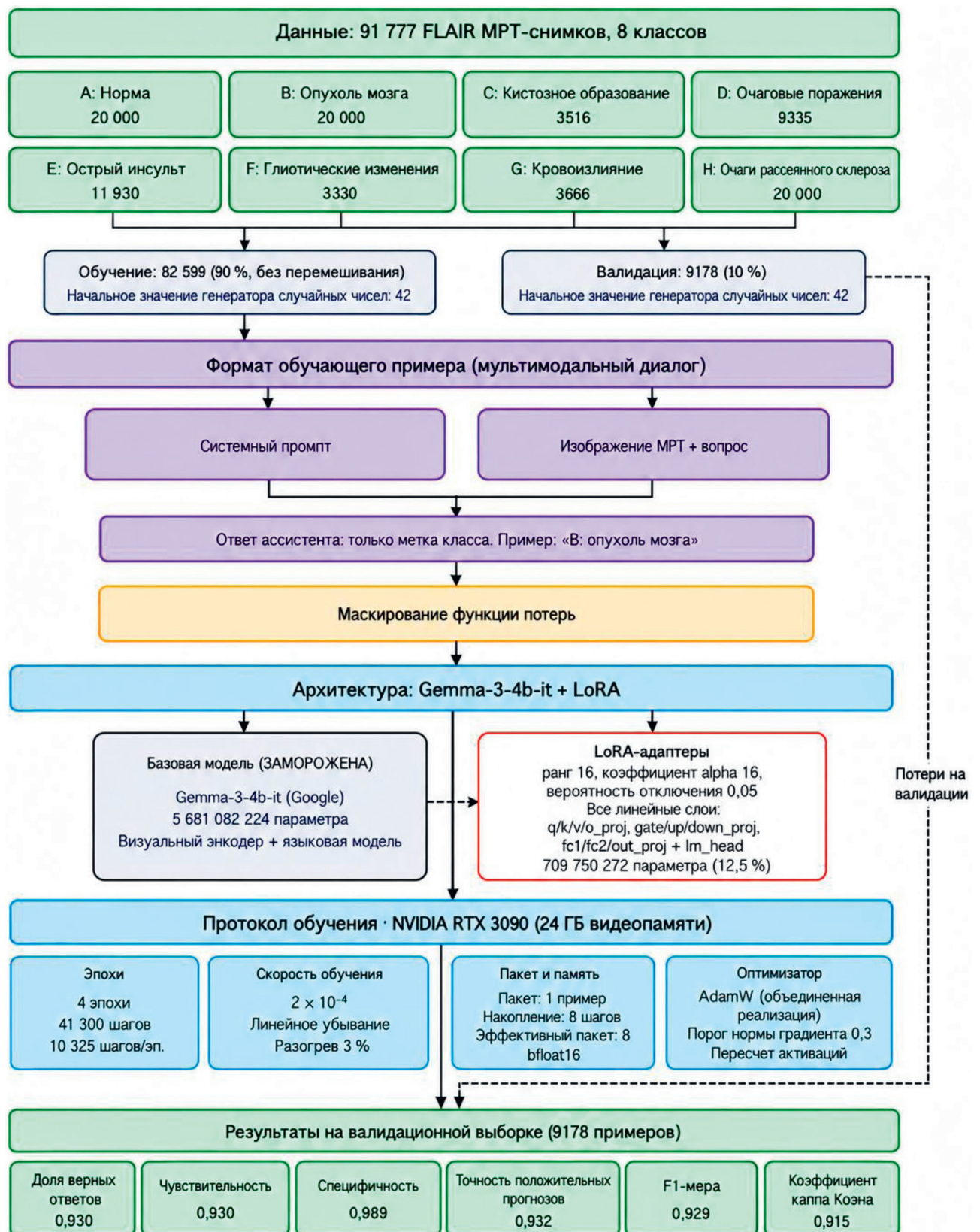


Рис. 1. Пайплайн дообучения модели

Fig. 1. Fine-tuning pipeline of the model

Число обучаемых параметров составило 709 750 272 из 5 681 082 224 (12,5 %), что позволило проводить обучение на одном графическом процессоре (Graphics Processing Unit, GPU) NVIDIA RTX 3090.

Использованные гиперпараметры обучения представлены в табл. 1.

Набор данных. Для дообучения модели в рамках настоящей работы был подготовлен специализированный набор данных, содержащий 91 777 МРТ-снимков головного мозга FLAIR-типа, распределенных по 8 классам. Распределение классов представлено в табл. 2, а примеры изображений из отдельных классов — на рис. 2.

Набор данных построен на основании деперсонализированных снимков из клинического архива. Отметим, что сформированный набор данных отражает дисбаланс данных, характерный для реальной клинической практики. А именно, классы с наибольшей представленностью (А, В, Н) насчитывают по 20 000 изображений, в то время как редкие классы (С, F, G) — от 3330 до 3666 изображений.

Набор данных был разбит в соотношении 90/10 (начальное значение генератора случайных чисел 42, без перемешивания), из которых 82 599 сэмплов использованы для обучения, 9178 — для валидации.

Метрики. Для оценки эффективности классификации использованы типовые метрики: Accuracy, Sensitivity, Specificity, Precision, F1-score и Cohen's

Kappa. F1-score (F1-мера) — гармоническое среднее точности положительных предсказаний (Precision) и полноты (Sensitivity); Cohen's Kappa (коэффициент каппа Коэна, κ) — мера согласованности предсказаний с эталонной разметкой с поправкой на случайное совпадение. В табл. 4 и 6 используется единое обозначение F1-score.

Результаты и обсуждение

Интеграция экспертных знаний врача-радиолога в процесс дообучения модели. Производилась через разметку и корректировку запроса (промпта). Для оценки роли промпта в эффективности обучения модели были построены варианты промптов, которые формировались путем последовательного добавления в базовый промпт клинических критериев, характерных для работы опытного врача-радиолога. Тем самым методика дообучения модели целенаправленно фокусирует модель на принятие диагностического решения.

В табл. 3 представлены характеристики и тексты промптов, а в табл. 4 — метрики эффективности обучения модели с применением каждого из них, полученные после прохождения 40 % первой эпохи обучения.

Как видно из сопоставления табл. 3 и 4, последовательное добавление в промпт уточняющих признаков, характерных для действий врача-радиолога, уже на

Таблица 1. Гиперпараметры обучения

Table 1. Training hyperparameters

Гиперпараметр	Значение	Гиперпараметр	Значение
Число эпох	4	Расписание скорости обучения	Линейное убывание
Размер пакета на устройство, примеров	1	Доля шагов разогрева, %	3 (около 1238 шагов)
Число шагов накопления градиента	8	Максимальная норма градиента	0,3
Эффективный размер пакета, примеров	8	Точность	bf16
Шагов на эпоху	10 325	Оптимизатор	AdamW (объединенная реализация)
Всего шагов	41 300	Пересчет активаций при обратном проходе	Включен
Скорость обучения	2×10^{-4}	—	—

Таблица 2. Распределение данных по классам

Table 2. Class distribution of the dataset

Класс	Описание	Количество изображений
А	Норма	20 000
В	Опухоль мозга	20 000
С	Кистозное образование	3516
Д	Очаговые поражения	9335
Е	Острый инсульт	11 930
F	Глиотические изменения	3330
G	Кровоизлияние	3666
Н	Очаги рассеянного склероза	20 000
Итого		91 777

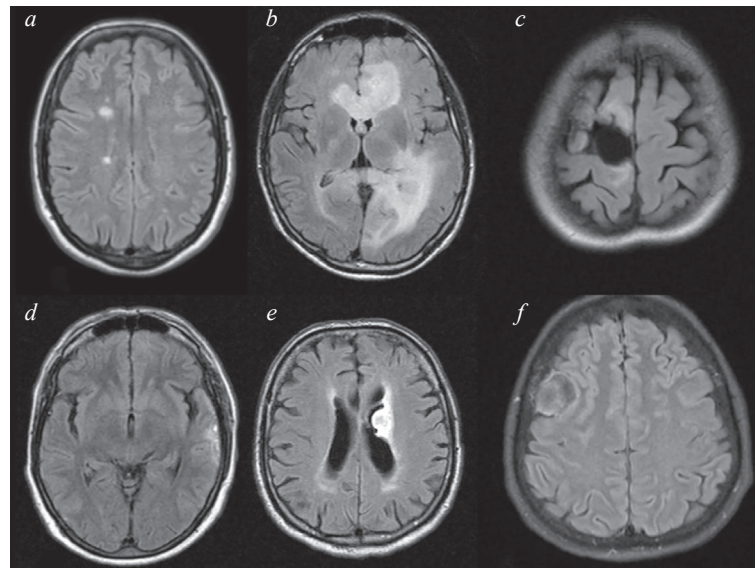


Рис. 2. Примеры патологий на изображениях из набора данных: рассеянный склероз (a); опухоль головного мозга (b); кистозная полость (c); острый инсульт (d); гематома (кровь) (e); менингиома (f)

Fig. 2. Examples of pathologies in dataset images: multiple sclerosis (a); brain tumor (b); cystic formation (c); acute stroke (d); hematoma (blood) (e); meningioma (f)

Таблица 3. Тексты промптов

Table 3. Prompt texts

Обозначение	Характеристика	Текст
P1	Базовый промпт	What pathology is shown in this MRI image? Select the most likely diagnosis from the options
P2	Добавлена роль радиолога	You are a radiologist. Look at this MRI image and identify the most likely diagnosis from the options provided
P3	Добавлены клинические критерии	You are an experienced neuroradiologist. Analyze the provided FLAIR MRI brain image carefully and determine the most likely diagnosis from the given options. Base your decision on signal intensity, lesion morphology, location, and distribution patterns

Таблица 4. Оценки эффективности обучения модели по разным промптам

Table 4. Model training performance evaluation for different prompts

Тип промпта	Потери на валидации	Accuracy	F1-score	Cohen's Kappa
P1	0,0381	0,821	0,809	0,781
P2	0,0247	0,863	0,851	0,827
P3	0,0088	0,930	0,929	0,915

ранних шагах обучения улучшает все метрики эффективности классификации, выполняемой моделью.

Экспериментальные оценки эффективности модели. Для оценки динамики обучения модели были замерены потери на обучении и на валидации в течение четырех эпох обучения по 10 325 шагов каждая. Результаты замеров по ключевым контрольным точкам приведены в табл. 5.

Как видно из табл. 5, модель продемонстрировала быструю сходимость: потери на обучении снизились с 0,086 на первых шагах до 0,005 за первую эпоху, после чего стабилизировались на уровне около 0,001.

Потери на валидации минимальны на шаге 4130, затем колеблются в диапазоне 0,02–0,06. Расхождение между потерями на обучении и потерями на валидации на поздних эпохах свидетельствует о небольшом переобучении, что характерно для задач с сильным дисбалансом классов и относительно небольшим числом редких примеров.

Для сравнительной оценки эффективности предложенного подхода в качестве бейзлайна был использован ConvNeXt-Base [14] — CNN-классификатор изображений, не использующий ни языковую модель, ни мультимодальный контекст. Важно подчеркнуть, что

Таблица 5. Экспериментальная оценка динамики обучения модели

Table 5. Experimental evaluation of model training dynamics

Контрольная точка, число пройденных эпох	Номер шага	Потери на обучении, отн. ед.	Потери на валидации, отн. ед.
0,4	4130	0,0860	0,0088
0,8	8260	0,0049	0,0223
1,2	12 390	0,0019	0,0311
2,0	20 650	0,0015	0,0202
3,2	33 040	0,0010	0,0620
4,0	41 300	0,0009	0,0259

ConvNeXt является сильным бейзлайном: предобученные веса ImageNet обеспечивают хорошую инициализацию признаков, а архитектура ConvNeXt-Base конкурентоспособна с трансформерами на задачах классификации изображений. Дообучение бейзлайна проводилось на том же специализированном наборе данных, что и для предлагаемой в настоящей работе модели (Gemma-3-4b-it + LoRA), что обеспечило методическую корректность сопоставления.

Обучение проводилось в течение четырех эпох. Для сравнительной оценки использована сохраненная версия модели с минимальными потерями на валидационной выборке, достигнутыми на шаге 4130. Оценка выполнена на валидационной выборке из 9178 примеров.

Результаты экспериментальной оценки эффективности сравниваемых классификаторов представлены в табл. 6.

Как видно из табл. 6, реализованная модель Gemma-3-4b-it с LoRA-дообучением превосходит ConvNeXt-Base по всем представленным метрикам. Наиболее выраженный прирост наблюдается по Cohen's Kappa (+0,090) и F1-score (+0,075), что особенно значимо при наличии несбалансированных классов в наборе данных, что характерно для клинической практики. Прирост Accurasy составил 7,4 процентных пункта (0,930 против 0,856).

Высокое значение Specificity (0,989) свидетельствует о минимальном числе ложноположительных срабатываний, что критически важно в клинической

практике. Specificity ConvNeXt (0,974) также высока и лишь незначительно уступает LoRA-модели (0,989), что указывает на низкий уровень ложноположительных срабатываний у обоих подходов.

Значение коэффициента каппа Коэна, равное 0,915, соответствует «почти совершенному» согласию по шкале Лэндиса и Коха [21] и указывает на высокую согласованность предсказаний модели с экспертной разметкой.

Заключение

В работе представлен метод контролируемого дообучения мультимодальной модели Gemma-3-4b-it с использованием механизма низкоранговой адаптации Low-Rank Adaptation для мультиклассовой диагностики патологий головного мозга по FLAIR-последовательностям изображений магнитно-резонансной томографии (МРТ). Дообучение 12,5 % параметров модели (709 млн из 5,68 млрд) на 91 777 снимках позволило достичь значений Accuracy 0,930, F1-score 0,929 и Cohen's Kappa 0,915 на валидационной выборке при обучении на одной GPU NVIDIA RTX 3090.

Ключевыми достоинствами метода являются: ресурсоэффективность — обучение на одной GPU, что доступно в реальной клинической практике; высокая Specificity (0,989), минимизирующая ложные срабатывания; легкая расширяемость на новые классы патологий путем дальнейшего дообучения адаптеров. Дальнейшие исследования будут направлены на интеграцию мультипоследовательного ввода (T1- и T2-взвешенные изображения, контраст которых определяется соответственно временем продольной и поперечной релаксации; диффузионно-взвешенные изображения (Diffusion-Weighted Imaging, DWI)) и клинических метаданных (возраст, пол, симптомы); инкрементальное добавление редких классов (как, например, концентрический склероз Бало) с оценкой по метрикам качества PR-AUC и ROC-AUC; использование других механизмов низкоранговой адаптации, таких как квантованная низкоранговая адаптация (Quantized Low-Rank Adaptation, QLoRA). В перспективе планируется автоматизация формирования клинических заключений по МРТ-снимкам, что будет способствовать развитию персонализированной высокотехнологичной медицинской помощи в России и за рубежом.

Таблица 6. Экспериментальная оценка эффективности классификаторов

Table 6. Experimental evaluation of classifier performance

Метрика	Gemma-3-4b-it + + LoRA	ConvNeXt-Base
Accuracy	0,930	0,856
Sensitivity	0,930	0,856
Specificity	0,989	0,974
Precision	0,932	0,855
F1-score	0,929	0,854
Cohen's Kappa	0,915	0,825

Литература

1. Chew B.-H., Ngiam K.Y. Artificial intelligence tool development: what clinicians need to know? // *BMC Medicine*. 2025. V. 23. N 1. P. 244. doi: 10.1186/s12916-025-04076-0
2. Najjar R. Redefining radiology: a review of artificial intelligence integration in medical imaging // *Diagnostics*. 2023. V. 13. N 17. P. 2760. doi: 10.3390/diagnostics13172760
3. Sima D.M., Phan T.V., Van Eyndhoven S., Vercruyssen S., Magalhães R., Liseune A., et al. Artificial intelligence assistive software tool for automated detection and quantification of amyloid-related imaging abnormalities // *JAMA Network Open*. 2024. V. 7. N 2. P. e2355800. doi: 10.1001/jamanetworkopen.2023.55800
4. Harada T., Shimizu T., Kaji Y., Suyama Y., Matsumoto T., Kosaka C., et al. A Perspective from a case conference on comparing the diagnostic process: human diagnostic thinking vs. Artificial Intelligence (AI) decision support tools // *International Journal of Environmental Research and Public Health*. 2020. V. 17. N 17. P. 6110. doi: 10.3390/ijerph17176110
5. Rauschecker A.M., Rudie J.D., Xie L., Wang J., Duong M.T., Botzakis E.J., et al. Artificial intelligence system approaching neuroradiologist-level differential diagnosis accuracy at brain MRI // *Radiology*. 2020. V. 295. N 3. P. 626–637. doi: 10.1148/radiol.2020190283
6. Shi Y., Shu P., Liu Z., Wu Z., Li Q., Liu T., et al. MGH Radiology Llama: A Llama 3 70B // *arXiv*. 2024. arXiv:2408.11848v2. doi: 10.48550/arXiv.2408.11848
7. Zhang G., Gao Z., Duan C., Liu J., Lizhu Y., Liu Y., et al. A multimodal vision-text AI copilot for brain disease diagnosis and medical imaging // *medRxiv*. 2024. doi: 10.1101/2025.01.09.25320293
8. Roest C., Yakar D., Sitar D.I.R., Bosma J.S., Rouw D.B., Fransen S.J., et al. Multimodal AI combining clinical and imaging inputs improves prostate cancer detection // *Investigative Radiology*. 2024. V. 59. N 12. P. 854–860. doi: 10.1097/RLI.0000000000001102
9. Liu Z., Li Y., Shu P., Zhong A., Jiang H., Pan Y., et al. Radiology-GPT: A large language model for radiology // *Meta-Radiology*. 2025. V. 3. N 2. P. 100153. doi: 10.1016/j.metrad.2025.100153
10. Lee S., Youn J., Kim H., Kim M., Yoon S.H. CXR-LLAVA: a multimodal large language model for interpreting chest X-ray images // *European Radiology*. 2025. V. 35. N 7. P. 4374–4386. doi: 10.1007/s00330-024-11339-6
11. Wang M.L., Zhang R.P., Wu W.J., Lu Y., Wei X.-E., Sun Z., et al. Evaluation of large language models for diagnostic impression generation from brain MRI report findings: a multicenter benchmark and reader study // *npj Digital Medicine*. 2026. V. 9. N 1. P. 187. doi: 10.1038/s41746-026-02380-4
12. Strubell E., Ganesh A., McCallum A. Energy and policy considerations for deep learning in NLP // *Proc. of the 57th Annual Meeting of the Association for Computational Linguistics*. 2019. doi: 10.18653/v1/p19-1355
13. Hu E.J., Shen Y., Wallis P., Allen-Zhu Z., Li Y., Wang S., et al. LoRA: Low-Rank Adaptation of Large Language Models // *International Conference on Learning Representations*. 2022.
14. Wang J., Jiang H., Liu Y., Ma C., Zhang X., Pan M., et al. A comprehensive review of multimodal Large Language Models: performance and challenges across different tasks // *arXiv*. 2024. arXiv:2408.01319v1. doi: 10.48550/arXiv.2408.01319
15. Moor M., Huang Q., Wu S., Yasunaga M., Dalmia Y., Leskovec J., et al. Med-Flamingo: a multimodal medical few-shot learner // *Proceedings of Machine Learning Research*. 2023. V. 225. P. 353–367.
16. He J., Li P., Liu G., He G., Chen Z., Zhong S. PeFoMed: Parameter efficient fine-tuning of multimodal Large Language Models for medical imaging // *arXiv*. 2025. arXiv:2401.02797. doi: 10.48550/arXiv.2401.02797
17. Chen J., Gui C., Ouyang R., Gao A., Chen S., Chen G.H., et al. Towards injecting medical visual knowledge into multimodal LLMs at scale // *Proc. of the Conference on Empirical Methods in Natural Language Processing*. 2024. P. 7346–7370. doi: 10.18653/v1/2024.emnlp-main.418
18. AlSaad R., Abd-alrazaq A., Boughorbel S., Ahmed A., Renault M.A., Damsheh R., et al. Multimodal Large Language Models in health care: applications, challenges, and future outlook // *Journal of Medical Internet Research*. 2024. V. 26. P. e59505. doi: 10.2196/59505
19. Yao Y., Yu T., Zhang A., Wang C., Cui J., Zhu H., et al. Efficient GPT-4V level multimodal large language model for deployment on edge

References

1. Chew B.-H., Ngiam K.Y. Artificial intelligence tool development: what clinicians need to know? *BMC Medicine*, 2025, vol. 23, no. 1, pp. 244. doi: 10.1186/s12916-025-04076-0
2. Najjar R. Redefining radiology: a review of artificial intelligence integration in medical imaging. *Diagnostics*, 2023, vol. 13, no. 17, pp. 2760. doi: 10.3390/diagnostics13172760
3. Sima D.M., Phan T.V., Van Eyndhoven S., Vercruyssen S., Magalhães R., Liseune A., et al. Artificial intelligence assistive software tool for automated detection and quantification of amyloid-related imaging abnormalities. *JAMA Network Open*, 2024, vol. 7, no. 2, pp. e2355800. doi: 10.1001/jamanetworkopen.2023.55800
4. Harada T., Shimizu T., Kaji Y., Suyama Y., Matsumoto T., Kosaka C., et al. A Perspective from a case conference on comparing the diagnostic process: human diagnostic thinking vs. Artificial Intelligence (AI) decision support tools. *International Journal of Environmental Research and Public Health*, 2020, vol. 17, no. 17, pp. 6110. doi: 10.3390/ijerph17176110
5. Rauschecker A.M., Rudie J.D., Xie L., Wang J., Duong M.T., Botzakis E.J., et al. Artificial intelligence system approaching neuroradiologist-level differential diagnosis accuracy at brain MRI. *Radiology*, 2020, vol. 295, no. 3, pp. 626–637. doi: 10.1148/radiol.2020190283
6. Shi Y., Shu P., Liu Z., Wu Z., Li Q., Liu T., et al. MGH Radiology Llama: A Llama 3 70B. *arXiv*, 2024. arXiv:2408.11848v2. doi: 10.48550/arXiv.2408.11848
7. Zhang G., Gao Z., Duan C., Liu J., Lizhu Y., Liu Y., et al. A multimodal vision-text AI copilot for brain disease diagnosis and medical imaging. *medRxiv*, 2024. doi: 10.1101/2025.01.09.25320293
8. Roest C., Yakar D., Sitar D.I.R., Bosma J.S., Rouw D.B., Fransen S.J., et al. Multimodal AI combining clinical and imaging inputs improves prostate cancer detection. *Investigative Radiology*, 2024, vol. 59, no. 12, pp. 854–860. doi: 10.1097/RLI.0000000000001102
9. Liu Z., Li Y., Shu P., Zhong A., Jiang H., Pan Y., et al. Radiology-GPT: A large language model for radiology. *Meta-Radiology*, 2025, vol. 3, no. 2, pp. 100153. doi: 10.1016/j.metrad.2025.100153
10. Lee S., Youn J., Kim H., Kim M., Yoon S.H. CXR-LLAVA: a multimodal large language model for interpreting chest X-ray images. *European Radiology*, 2025, vol. 35, no. 7, pp. 4374–4386. doi: 10.1007/s00330-024-11339-6
11. Wang M.L., Zhang R.P., Wu W.J., Lu Y., Wei X.-E., Sun Z., et al. Evaluation of large language models for diagnostic impression generation from brain MRI report findings: a multicenter benchmark and reader study. *npj Digital Medicine*, 2026, vol. 9, no. 1, pp. 187. doi: 10.1038/s41746-026-02380-4
12. Strubell E., Ganesh A., McCallum A. Energy and policy considerations for deep learning in NLP. *Proc. of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, doi: 10.18653/v1/p19-1355
13. Hu E.J., Shen Y., Wallis P., Allen-Zhu Z., Li Y., Wang S., et al. LoRA: Low-Rank Adaptation of Large Language Models. *International Conference on Learning Representations*, 2022.
14. Wang J., Jiang H., Liu Y., Ma C., Zhang X., Pan M., et al. A comprehensive review of multimodal Large Language Models: performance and challenges across different tasks. *arXiv*, 2024. arXiv:2408.01319v1. doi: 10.48550/arXiv.2408.01319
15. Moor M., Huang Q., Wu S., Yasunaga M., Dalmia Y., Leskovec J., et al. Med-Flamingo: a multimodal medical few-shot learner. *Proceedings of Machine Learning Research*, 2023, vol. 225, pp. 353–367.
16. He J., Li P., Liu G., He G., Chen Z., Zhong S. PeFoMed: Parameter efficient fine-tuning of multimodal Large Language Models for medical imaging. *arXiv*, 2025. arXiv:2401.02797. doi: 10.48550/arXiv.2401.02797
17. Chen J., Gui C., Ouyang R., Gao A., Chen S., Chen G.H., et al. Towards injecting medical visual knowledge into multimodal LLMs at scale. *Proc. of the Conference on Empirical Methods in Natural Language Processing*, 2024, pp. 7346–7370. doi: 10.18653/v1/2024.emnlp-main.418
18. AlSaad R., Abd-alrazaq A., Boughorbel S., Ahmed A., Renault M.A., Damsheh R., et al. Multimodal Large Language Models in health care: applications, challenges, and future outlook. *Journal of Medical Internet Research*, 2024, vol. 26, pp. e59505. doi: 10.2196/59505
19. Yao Y., Yu T., Zhang A., Wang C., Cui J., Zhu H., et al. Efficient GPT-4V level multimodal large language model for deployment on edge

- devices // *Nature Communications*. 2025. V. 16. N 1. P. 5509. doi: 10.1038/s41467-025-61040-5
20. Zhan C., Lin Y., Wang G., Wang H., Wu J. MedM2G: Unifying medical multi-modal generation via cross-guided diffusion with visual invariant // *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2024. P. 11502–11512. doi: 10.1109/cvpr52733.2024.01093
 21. Landis J.R., Koch G.G. The Measurement of observer agreement for categorical data // *Biometrics*. 1977. V. 33. N 1. P. 159–174. doi: 10.2307/2529310

Авторы

Гусарова Наталия Федоровна — кандидат технических наук, старший научный сотрудник, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 57162764200](https://orcid.org/0000-0002-1361-6037), [natfed@list.ru](https://orcid.org/0000-0002-1361-6037)

Добренко Наталья Викторовна — кандидат технических наук, доцент, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 56499375200](https://orcid.org/0000-0001-6206-8033), [graziokisa@yandex.ru](https://orcid.org/0000-0001-6206-8033)

Зубаненко Алексей Александрович — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 57215436184](https://orcid.org/0000-0001-6953-5239), [zubdocmri@gmail.com](https://orcid.org/0000-0001-6953-5239)

Томилев Иван Вячеславович — студент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 57772599000](https://orcid.org/0000-0003-1886-2867), [ivan-tomilov3@yandex.ru](https://orcid.org/0000-0003-1886-2867)

Ульянов Павел Геннадьевич — кандидат физико-математических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 42762414900](https://orcid.org/0009-0008-0641-901X), [pavel.ulyanov@itmo.ru](https://orcid.org/0009-0008-0641-901X)

Статья поступила в редакцию 07.04.2026

Одобрена после рецензирования 20.08.2026

Принята к печати 21.09.2026

Authors

Natalia F. Gusarova — PhD, Senior Researcher, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 57162764200](https://orcid.org/0000-0002-1361-6037), [natfed@list.ru](https://orcid.org/0000-0002-1361-6037)

Natalia V. Dobrenko — PhD, Associate Professor, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 56499375200](https://orcid.org/0000-0001-6206-8033), [graziokisa@yandex.ru](https://orcid.org/0000-0001-6206-8033)

Aleksey A. Zubanenko — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 57215436184](https://orcid.org/0000-0001-6953-5239), [zubdocmri@gmail.com](https://orcid.org/0000-0001-6953-5239)

Ivan V. Tomilov — Student, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 57772599000](https://orcid.org/0000-0003-1886-2867), [ivan-tomilov3@yandex.ru](https://orcid.org/0000-0003-1886-2867)

Pavel G. Ulyanov — PhD (Physics & Mathematics), Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 42762414900](https://orcid.org/0009-0008-0641-901X), [pavel.ulyanov@itmo.ru](https://orcid.org/0009-0008-0641-901X)

Received 07.04.2026

Approved after reviewing 20.08.2026

Accepted 21.09.2026



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»