

ОПТИМИЗАЦИЯ ПОСТФАКТУМНЫХ МЕТОДОВ ОБЪЯСНИМОГО ИИ ДЛЯ АВТОМАТИЗИРОВАННОГО АНАЛИЗА МЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ

Сорокина С. А.¹

Научный руководитель – д-р. техн. наук, доцент Кузяков О. Н.¹

¹Тюменский индустриальный университет

SorokinaSA.Mail@yandex.ru

Введение

В условиях цифровой трансформации автоматизация процесса лучевой диагностики в области объективной и точной интерпретации результатов врачом-рентгенологом остается одним из наиболее актуальных направлений развития здравоохранения [1]. В свою очередь, компьютерная томография (КТ) является одним из наиболее информативных методов лучевой диагностики, обеспечивающим высокий уровень детализации и точности выявления патологий. Однако традиционный метод анализа КТ-изображений, основанный на визуальной оценке врачом-рентгенологом, сталкивается с рядом проблем, возникающих по причине субъективности и высокой нагрузки на медицинских специалистов, что повышает риск диагностических ошибок. В связи с этим, анализ медицинских изображений при помощи моделей глубокого обучения на основе свёрточных нейронных сетей (CNN) становится ключевым направлением развития лучевой диагностики. Такие автоматизированные системы демонстрируют высокую эффективность в задачах обнаружения, классификации и сегментации патологий, открывая тем самым перспективы для повышения эффективности, точности и прозрачности постановки диагноза [2]. Однако широкого внедрения систем на основе алгоритмов глубокого обучения в клиническую практику на данный момент не происходит. Причиной этого является эффект «чёрного ящика», при котором пользователю не предоставляется информация о том, на основании чего было принято то или иное решение системой [3]. Для решения этой проблемы предлагается использовать постфактумные методы объяснимого искусственного интеллекта (Explainable AI, XAI), которые позволят анализировать решения уже обученных моделей глубокого обучения без непосредственного изменения их архитектуры.

Среди таких методов особое место занимает метод отображения активации класса с взвешенным градиентом (Gradient-weighted Class Activation Mapping, Grad-CAM), визуализирующий области изображения, наиболее значимые для принятия решения моделью [4]. Однако стандартная реализация метода Grad-CAM имеет ограничения, связанные с выбором слоя CNN для генерации карт активации, что может приводить к неточной локализации патологий и снижению клинической ценности результатов. Целью данного исследования является оптимизация постфактумных методов объяснимого ИИ (методов визуализации на основе Grad-CAM) для повышения точности и интерпретируемости автоматического анализа медицинских изображений.

Основная часть

В ходе исследования для проверки эффективности метода Grad-CAM визуализировать наиболее значимые регионы (патологии) карт активации слоев CNN при анализе КТ-изображений легких, было проведено тестирование на наборе данных NIH ChestX-ray. Набор содержал КТ-изображения грудной клетки формата DICOM с классами «Норма» (1500 изображений) и «Патология» (1350 изображений). Для автоматизированного анализа КТ-изображений для диагностики патологий легких была выбрана CNN архитектуры ResNet18, обученная на задаче бинарной классификации.

В ходе применения метода визуализации Grad-CAM и анализа карт активации CNN архитектуры ResNet18 было выявлено ограничение, которое заключается в использовании активации последнего свёрточного слоя сети, что может приводить к генерации неспецифичных карт и становиться критичным фактором при решении задачи точной локализации патологий на изображении. Для решения этой проблемы была разработана методика автоматического выбора оптимального слоя CNN для генерации карт активации Grad-CAM на основе количественных критериев: семантической связности (оценки целостности выделенного региона активации); пространственной точности (соответствия карт активации экспертной разметке с использованием метрики Intersection over Union); контрастности (чёткости перехода между областью интереса и фоном изображения). Методика включила два этапа: количественную оценку качества карт активации для каждого слоя CNN и выбор оптимального слоя на основе взвешенная суммы критериев. В результате исследования, было установлено, что промежуточные слои CNN обеспечивают более высокую точность локализации патологий по сравнению со стандартным методом Grad-CAM. При количественном анализе слоёв ResNet18 были выявлены следующие закономерности: первый слой характеризовался высоким уровнем шума и низкой семантической связностью, на втором слое сохранялась высокая фрагментированность активаций, на третьем слое были зафиксированы наилучшие результаты по критериям семантической связности и пространственной точности, четвёртый слой содержал излишнюю абстракцию. Таким образом, третий слой был выбран в качестве оптимального для генерации карт активации Grad-CAM, что позволило повысить точность локализации патологий на 42% по сравнению со стандартным подходом.

Выводы

Таким образом, предложенная методика автоматического выбора оптимального слоя CNN для Grad-CAM позволит адаптивно настраивать механизм объяснения работы автоматизированной системы под конкретную архитектуру и клиническую задачу, что позволит повысить интерпретируемость и точность систем анализа медицинских изображений при их внедрении непосредственно в клиническую практику и при дальнейшей оптимизации постфактумных методов ХАИ в других областях лучевой диагностики и радиологии при решении задач медицинской визуализации.

Литература

1. Ганичев П. А., Тихомирова А. А., Дохов М. А. Перспективы использования искусственного интеллекта в радиологии. Краткий обзор // Визуализация в медицине. 2022. Т 4. №4. С. 7–14.
2. Combi C., Amico B., Bellazzi R., Holzinger A., Moore J.H., Zitnik M., Holmes J.H., A manifesto on explainability for artificial intelligence in medicine // Artificial Intelligence in Medicine. 2022. Vol. 133, no. 102423. P. 1–13. <https://doi:10.1016/j.artmed.2022.102423>.
3. De Vries B.M., Zwezerijnen G.J.C., Burchell G.L., van Velden F.H.P., Menke-van der Houven van Oordt C.W., Boellaard R. Explainable artificial intelligence (XAI) in radiology and nuclear medicine: a literature review // Front Med. 2023. Vol. 10, no. 1180773. P. 1–14. <https://doi:10.3389/fmed.2023.1180773>.
4. Niranjana K., Shankar Kumar S., Vedanth S., Chitrakala D.S. An Explainable AI driven Decision Support System for COVID-19 Diagnosis using Fused Classification and Segmentation // Procedia Comput Sci. 2023. No. 218. P. 1915–1925. <https://doi:10.1016/j.procs.2023.01.168>.