УДК 004.8

ОБЪЯСНИМЫЙ ИНКРЕМЕНТНЫЙ ПОДХОД В ДИАГНОСТИКЕ ПНЕВМОНИИ: ОТ СВЁРТОЧНЫХ НЕЙРОСЕТЕЙ ДО ГЕНЕРАТИВНЫХ ТЕКСТОВЫХ ЗАКЛЮЧЕНИЙ

Кузнецов Е.М.[1], Муравьев И.П.[1], Трофимов Ю.В.[1]
Научный руководитель – кандидат физико-математических наук, доцент Аверкин А.Н.[1]
1 – Государственный университет «Дубна»

Работа выполнена в рамках государственного задания Министерства науки и высшего образования Российской Федерации (тема № 124112200072-2).

Введение. Автоматизированная диагностика на основе рентгеновских (X-ray) изображений грудной клетки является одним из перспективных направлений применения методов Deep Learning (DL) и компьютерного зрения (CV) в современной медицине. Однако высокой точности компьютерных моделей недостаточно для практического внедрения, поскольку врачи требуют прозрачности и осмысленной интерпретации итоговых решений. Именно поэтому всё большее внимание уделяется методам eXplainable AI (XAI), позволяющим повысить объяснимость работы нейросетевых алгоритмов и преодолеть барьер «чёрного ящика» [1]. При этом настоящая работа дополняет визуальные объяснения (Grad-CAM, LIME, SHAP) автоматической генерацией диагностических заключений на естественном языке, ориентированной на удобство интеграции в клинические информационные системы (веб-приложения, PACS/EMR).

Основная часть. На первом этапе проведён углублённый обзор публичных X-ray-датасетов (Kaggle Pneumonia Dataset [2], NIH ChestXRay [3], MIMIC-CXR [4] и др.) с целью выявления наиболее репрезентативной выборки, содержащей метки «пневмония» различных типов (бактериальная, вирусная). По итогам сравнительного анализа (учитывая объём, разнообразие патологий, качество этикеток/аннотаций) выбран датасет *Kaggle Chest X-Ray Images (Pneumonia)* в качестве опорного. Дополнительно составлен вспомогательный набор тестовых снимков (30–50 образцов) для ручной валидации и отладки методов объяснимости.

На втором этапе изучены и переобучены (fine-tuned) предобученные CNN-архитектуры: ResNet, VGG, DenseNet и EfficientNet, а также более компактные варианты (например, SqueezeNet), используя фреймворки PyTorch и TensorFlow. Произведена первичная кросс-валидация (k-fold) с метриками *Accuracy, Precision, Recall, F1*, что позволило исключить менее перспективные модели. По итогам экспериментов наилучшее соотношение «качество–скорость–стабильность» продемонстрировали ResNet34 и DenseNet121. По окончанию тестирования ResNet34 была выбрана для последующей интеграции в складывающийся инкрементальный XAI подход.

В качестве третьего этапа для повышения обобщающей способности (generalization) применены расширенные техники Data Augmentation (случайные повороты, горизонтальные/вертикальные отражения, вариации контраста/яркости), а также регуляризация (Dropout, Weight Decay). Разработан инкрементный (incremental) подход: первоначально модель формирует исходную классификацию (пороговое разделение «пневмония / норма»), второй частью идёт уточнение подтипов (вирусная/бактериальная), учитывая результаты первой. На данном этапе обеспечены надёжные результаты (80–90% *Accuracy*), но ключевой задачей оставалось обоснование принятых решений.

Четвертый этап, Применены классические техники XAI: Grad-CAM (Gradient-weighted Class Activation Mapping), LIME (Local Interpretable Model-agnostic Explanations), SHAP

(SHapley Additive exPlanations) [5]. Полученные тепловые карты (heatmaps) и локальные объяснения позволяют визуально оценить зону локализации потенциальных патологических участков. Данный инструментарий XAI существенно повышает прозрачность, однако для практикующего врача, не владеющего тонкостями ML, важно иметь более удобный формат объяснения.

Пятый предпоследний этап, в работе реализован модуль «автоописания», который из активированных областей (выделенных XAI-методами) формирует текстовую интерпретацию, используя «мягкие дескрипторы» (soft descriptors) и онтологические термины (например, «признак инфильтрации», «уплотнение лёгочной ткани»). Разработанный алгоритм генеративного описания (на базе предобученных Language Models) сопоставляет тепловую карту модели с медико-семантическими шаблонами. В результате врач получает краткий, но содержательный текст, подтверждающий или опровергающий патологию. Такой многоступенчатый подход (CNN-выделение → XAI-визуализация → генеративное описание) представляет собой один из ключевых компонентов Human-Centric XAI.

На финальном шестом этапе, итоговое решение оформлено как веб-приложение с возможностью интеграции в PACS/EMR. Пользователь (врач) загружает снимок, получает классификацию (есть/нет пневмонии), визуализацию с Grad-CAM/LIME/SHAP, а также текстовое обоснование результата. Возможность адаптации формата вывода позволяет внедрять данный инструмент как вспомогательную систему поддержки принятия решений (Clinical Decision Support, CDS) для рутинных протоколов.

Выводы. Предложенный XAI-ориентированный подход (состоит из двухуровневой классификации, визуальной интерпретации и генеративного текстового описания) показывает, что при диагностике пневмонии на рентген-снимках не только повышается точность классификации, но и достигается высокая степень объяснимости (transparency) решений. Применение Grad-CAM, LIME, SHAP совместно с человечески понятной (human-readable) генерацией заключений формирует «мост» между глубинными нейросетями и реальными клиническими сценариями. Разработанное веб-приложение даёт врачу возможность интуитивно оценивать обоснованность предсказаний модели, а также упрощает дальнейшую интеграцию системы в инфраструктуру медицинских учреждений.

Список использованных источников:

1. E. Tjoa and C. Guan, "A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 11, pp. 4793-4813, Nov. 2021, doi: 10.1109/TNNLS.2020.3027314.
2. Kermany, Daniel; Zhang, Kang; Goldbaum, Michael (2018), “Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification”, Mendeley Data, V2, doi: 10.17632/rscbjbr9sj.2
3. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri and R. M. Summers, "ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 3462-3471, doi: 10.1109/CVPR.2017.369.
4. Johnson, A.E.W., Pollard, T.J., Berkowitz, S.J. *et al.* MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. *Sci Data* **6**, 317 (2019). https://doi.org/10.1038/s41597-019-0322-0
5. Molnar C. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. Leanpub, 2020. [Электронный ресурс]. URL: https://christophmolnar.com/books/interpretable-machine-learning/ (дата обращения: 10.10.2024).