Объяснительный искусственный интеллект в задачах анализа медицинских изображений: современное состояние и перспективы

Е. Н. Волков[1], А. Н. Аверкин[2]
[1] *Государственный университет «Дубна»*
[2] *ФИЦ «Информатика и управление» РАН*
E-mail envolkoff1998@yandex.ru

***Аннотация*. В преддверии перехода к цифровой системе здравоохранения и персонализированной медицине (Healthcare 5.0) происходит экспоненциальный рост использования систем искусственного интеллекта. Этот рост наиболее заметен в сфере анализа медицинских изображений. Однако, с усложнением алгоритмов работы, возможности пользователя по контролю за принятием решения значительно снижаются, что сказывается на доверии к получаемому результату, которое критически важно при постановке диагноза. Повысить прозрачность в работе систем искусственного интеллекта при анализе медицинских изображений призвано применение методов объяснительного искусственного интеллекта. В нашем исследовании представлен обзор современного состояния применения методов объяснительного искусственного интеллекта для анализа медицинских изображений, а также рассмотрены потенциально перспективные подходы к совершенствованию технологии.**

Ключевые слова: искусственный интеллект, объяснительный искусственный интеллект, объяснимость, нейронные сети, искусственные нейронные сети, цифровое здравоохранение, персонализированная медицина, медицина, здравоохранение, медицинские изображения

I. Введение

В сфере здравоохранения в ближайшем будущем ожидается большой скачок, связанный с переходом к Healthcare 5.0 – новому подходу к медицине, ориентированному на персонализированной медицине. Ожидается, такие технологии как искусственный интеллект, большие данные, интернет вещей станут локомотивом этого перехода. Уже сейчас присутствие технологий искусственного интеллекта в медицине увеличивается в десятки раз с каждым годом. [19]

Однако, несмотря на то, что технологии искусственного интеллекта имеют огромный потенциал, усложнение систем сказывается на доверии к технологии. Особенно это заметно в сферах, где от правильности решения зависит человеческая жизнь. Выходом из этой непростой ситуации становится применение объяснительного искусственного интеллекта (eXplainable Artificial Intelligence (XAI)).

Говоря об XAI, мы подразумеваем «объяснимость» (explainability) как способность модели представить результат работы в виде понятного пользователю интерфейса. В тоже время, под применением искусственного интеллекта в медицине мы понимаем применение систем, основанных на машинном обучении (ML-systems) и глубоком обучении (DL-systems). В основе таких систем, как правило, лежат искусственные нейронные сети, которые не обладают свойством прозрачности (transparency) – не могут быть понятны априори, без применения методов интерпретации. Искусственные нейронные сети искусственного интеллекта, по своей сути, являются «чёрным ящиком» (black-box), то есть процесс принятия ими решения неясен, а прозрачные лишь вход и выход сети. [5]

Попытки объяснить работу сложных систем велись, начиная с 1970-х годов. Согласно [1, 2] выделяют три этапа развития: на первом происходила разработка экспертных систем, использовавших механизм вопросно-ответного интерфейса; на втором (середина 1980-х годов) разрабатывались системы, основанные на знаниях; на третьем (с 2017 г. по н. в.) изучаются глубокие архитектуры искусственный нейронных сетей. Благодаря программе DARPA, стартовавшей в 2017 году возникла новая волна исследований в этом направлении. В настоящее время, тема объяснимости в искусственном интеллекте является одной из самых актуальных.

Применение XAI в медицинских системах имеет огромный потенциал. В нашей работе мы остановимся лишь на одной области – анализе медицинских изображений. К медицинским изображениям относят снимки, полученные как с помощью рентгена, компьютерной томографии, магнитно-резонансной томографии, так и обычные снимки медицинской направленности (снимки образований кожи, снимки глаза и сетчатки, снимки гистологических препаратов). Использование приведённых технологий как неинвазивных методов диагностики и контроля терапии является одной из ключевых черт современной медицины.

II. Материалы и методы

A. Анализ медицинских изображений как задача компьютерного зрения

Основными задачами в анализе медицинских изображений является поиск аномалий, заранее определённых или вновь выявляемых. Наиболее частыми примерами таких задач являются классификация и сегментация. Классификация, как правило бинарная, используется для поиска определённых объектов на снимках. Сегментация используется реже, её задача обозначить границы определённых областей изображения. Например, поиск глаукомы на снимке сетчатки глаза является задачей классификации, а определение границ опухолей головного мозга – задачей сегментации (рис. 1).

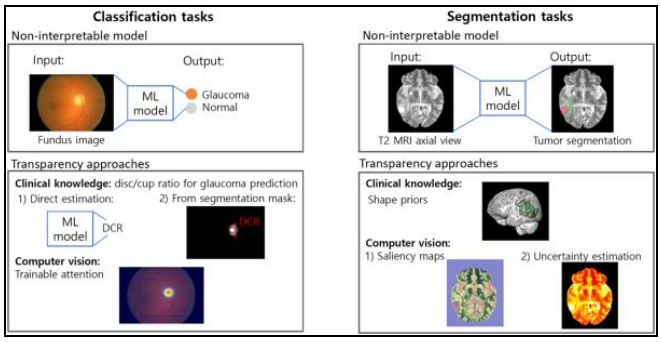

Рис. 1. Примеры задач и методов, используемых в обычных и прозрачных ML-системах. Из исследования Chen H. и др. [8]

B. Методы XAI

Стоит отметить, что существует большое количество таксономий методов XAI, каждая из которых предлагает свою систематизацию. В нашем исследовании мы будем использовать определения терминов XAI и таксономию методов объяснения, введенную Arrieta A. в работе [5].

Методы объяснения результатов работы искусственной нейронной сети искусственного интеллекта – это методы, использующие внешние интерпретаторы (model-agnostic) или вычислительный граф (model-specific) для представления результата работы модели в виде понятного интерфейса. Объяснения также могут быть локальными или глобальными, то есть описывающими поведение всей модели или модели относительно отдельного объекта. Наиболее часто используемыми объяснениями являются объяснения по итогу результатов работы или пост-фактом объяснения (post-hoc).

Наиболее часто в работах, посвящённых анализу изображений, применяются два метода объяснения: CAM (Class Activation Maps) [25] и Grad-CAM (Gradient-Class Activation Maps) [20], что объясняется возможностями применения этих методов за счёт встраивания в нейронные сети различных архитектур, оптимизированным вычислительным затратам, понятности представления результата. Редко применяемыми, но встречающимися в отдельных работах, являются следующие методы LIME (Local Interpretable Model-agnostic Explanations) [17], LRP (Layer-wise Relevance Propagation) [6], SHAP (SHapley Additive exPlanations) [14].

III. Современное состояние

A. Методы лучевой диагностики

Наибольший опыт применения XAI при анализе изображений накоплен в сфере интерпретации результатов лучевой диагностики. Использование методов лучевой диагностики, таких как магнитно-резонансная томография (МРТ) и компьютерная томография (КТ), рентгена требует высокой квалификации врача лучевой диагностики и лаборантов при анализе снимков. В этом случае, использование сочетания нейросетевых методов и методов XAI даст не только значительный выигрыш во времени анализа изображения, но и дополнительную уверенность в принадлежности искомых классов объектов, что особенно актуально, так как МРТ, КТ и рентген снимки содержат значительное количество артефактов.

В последние несколько лет компьютерная томография стала самым известным среди обывателей и наиболее используемым методом лучевой диагностики вследствие появления пандемии новой коронавирусной инфекции. Компьютерная томография проводится как при подозрении на наличие COVID-19, так и для диагностики наличия пневмонии или степени поражения лёгких ею. Количество исследований применения методов машинного обучения для анализа снимков грудной клетки (chest CT/MRI/X-ray) многократно возросло, начиная с 2020 года. Публикации, посвященные применению XAI в этой сфере, также стали довольно частыми [4, 15, 16]. Возможности использования методов XAI в диагностике ковидной пневмонии представлены на рис. 2.

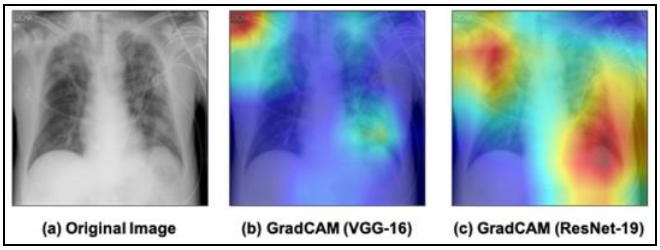

Рис. 2. Визуализация очагов поражения лёгких с использованием метода объяснения Grad-CAM по результатам распознавания сетями VGG-16 и ResNet-19. Из исследования Giuste F. и др. [9]

Проведя анализ обзоров литературы [4, 22], посвящённых применению XAI в диагностике заболеваний ковидной пневмонией и других заболеваний лёгких нами было отобрано 21 и 33 публикации соответственно. В качестве метода выбора для объяснения результата применялся CAM (n=19) и Grad-CAM (n=13), в отдельных случаях использовался LIME (n=7) и LRP (n=5).

Методы XAI также с успехом нашли своё применение в диагностике рака молочной железы у женщин (breast cancer) (рис. 3). В обзоре [22] приводится 17 работ, посвященных данной тематике. В большинстве из них методом объяснения служит CAM (n=8).

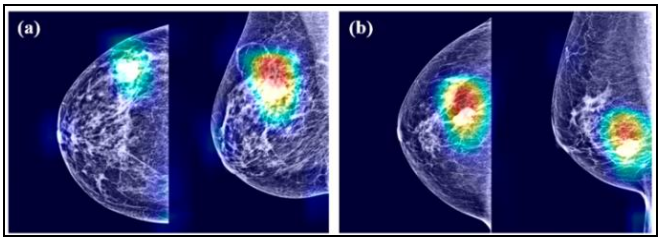

Рис. 3. Визуализация опухоли молочной железы с использованием Grad-CAM по результатам распознавания сетями DenseNet-169 и EfficientNet-B5. Из исследования Suh Y. и др. [21]

В анализе КТ, МРТ снимков головного мозга методы объяснения также нашли своё применение. Они используются при диагностике опухолевых (глиома, глиобластома) и неопухолевых заболеваний (Болезнь Альцгеймера и т. д.) головного мозга (рис. 4). Обзор литературы [22] содержит упоминания о 42 исследованиях по данной теме. Наиболее часто в анализе снимков головного мозга применялись методы CAM (n=12), Grad-CAM (n=11), LRP (n=5).

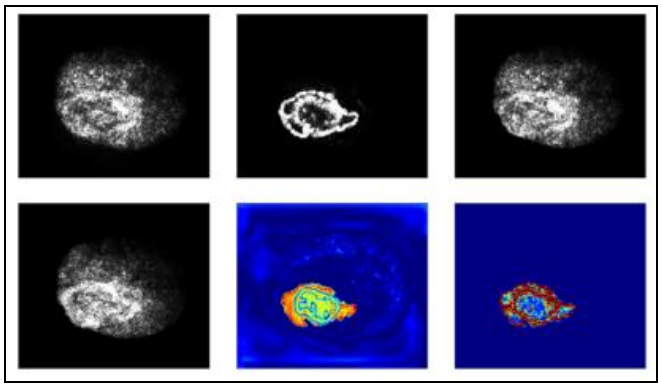

Рис. 4. Визуализация опухоли головного мозга с помощью методов XAI. Из исследования R. Zeineldin и др. [24]

B. Снимки кожных покровов

Отдельное место занимает использование XAI в диагностике заболеваний кожных покровов. Особого внимания заслуживают исследования применения методов XAI в дерматоонкологии при диагностике меланомы и рака кожи. В систематическом обзоре [22] приведено 6 публикаций подобной тематики. Пример использования метода объяснения для диагностики меланомы представлен на рис. 5.

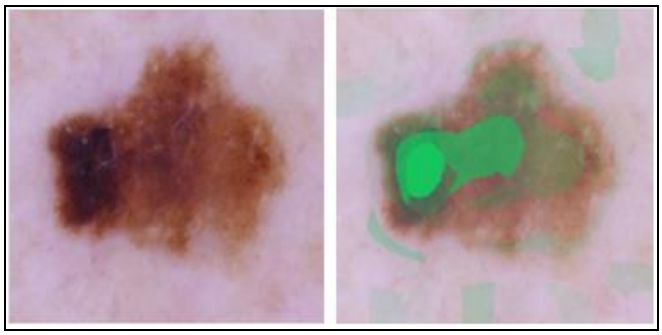

Рис. 5. Использование XAI при диагностике меланомы. Слева – оригинальное изображение. Справа – объяснение с помощью метода KernelSHAP. Из исследования Young K. и др. [23]

C. Снимки глаза

Методы XAI нашли своё применение в диагностике болезней глаза, таких как глаукома, диабетическая ретинопатия и других менее распространённых заболеваний. Использование методов XAI в диагностике этих заболеваний возможно через анализ снимков радужной оболочки, сетчатки и глазного дна. В систематических обзорах [8, 18, 22] представлено 31 публикация по данной теме. Наиболее часто методами объяснения становились CAM (n=12) и Grad-CAM (n=8).

D. Цифровая патология и гистология

Использование методов машинного обучения в цифровую патологию и гистологию пришло сравнительно недавно по сравнению с другими областями медицины, где используется анализ изображений. Однако, несмотря на это, существует ряд исследований, представленных в обзоре [22], изучающих возможности XAI в этой области. Всего проиндексировано 16 таких исследований, наиболее часто применяемым методом объяснения, в данном случае, стал Grad-CAM (n=11). Пример использования XAI в цифровой патологии и гистологии приведён на рис. 6.

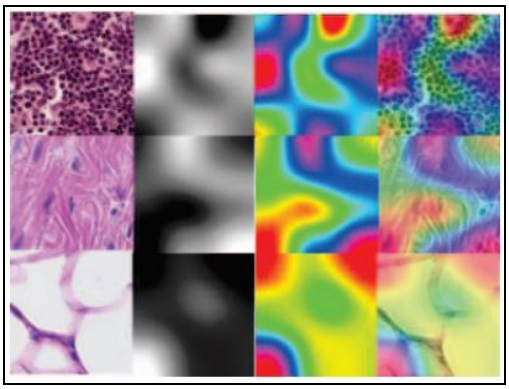

Рис. 6. Использование метода объяснения Grad-CAM в диагностике наличия метастатических клеток в ткани лимфатического узла. Из исследования J. Ji [12]

IV. Перспективы

A. Человекоцентричность как основной принцип при проектировании систем анализа медицинских изображений с применением XAI

В систематическом обзоре литературы Chen H. [8] было проанализировано 68 публикаций из электронных ресурсов PubMed, EMBASE, Compendex за период с 2012 по 2021 год, посвящённых созданию решений на основе XAI для анализа медицинских изображений. Ни в одной из публикаций автором исследования не было найдено упоминаний об ориентированности решения на конечного пользователя – клинициста. Наоборот, в большинстве публикаций предпочтение отдается оптимизации алгоритмов. По мнению авторов другого обзора [7], выходом из такой ситуации может стать создание принципов проектирования систем, ориентированных на удобстве клиницистов, а также соответствующих требованиям: простоты использования, высокой достоверности предсказания и его объяснения, надёжности, вычислительных затрат, возможности точной настройки, открытого кода продукта.

B. Совместное применение XAI с элементами нечёткой логики и нейро-нечёткими системами

В работе [10] показаны возможности использования нейро-нечёткой искусственной нейронной сети в сочетании с методом XAI CAM для анализа изображений и госпитальных записей больных COVID-19. Подобное решение также было представлено для ранней диагностики ковидной пневмонии в исследовании [11]. Анализ случаев возникновения глаукомы и прогнозирования факторов, влияющих на процесс лечения также возможен с применением данных методов. В работе [13] был применён метод XAI LIME и адаптивная нейро-нечеткая система логического вывода ANFIS для анализа изображений и записей пациентов, страдающих глаукомой.

В целом совместное применение XAI и решений на основе нечёткой логики расширяет границы использования систем анализа медицинских изображений в реальной клинической практике, благодаря появлению возможности следования стратегии мультимодальности при принятии решения клиницистами. Совместное применение решений на основе машинного обучения, XAI и элементов нечёткой логики говорит о возможностях создания медицинских

экспертных систем с применением гибридного искусственного интеллекта. [3]

V. Заключение

В исследовании представлены примеры применения методов XAI для анализа медицинских изображений, а также современное состояние литературы по исследуемой тематике, оценена частота использования отдельных методов объяснения. Показаны два перспективных направления развития для анализа медицинских изображений с использованием XAI, основанных на примерах реальных исследований.

Список литературы

[1] Аверкин А.Н. Исследование развития систем объяснительного искусственного интеллекта / А.Н. Аверкин, С.А. Ярушев // Интегрированные модели и мягкие вычисления в искусственном интеллекте ИММВ-2022: Сборник научных трудов XI Международной научно-практической конференции. В 2-х томах, Коломна, 16–19 мая 2022 года. Коломна: Общероссийская общественная организация «Российская ассоциация искусственного интеллекта», 2022. С. 127-134.

[2] Аверкин А.Н. Обзор исследований в области разработки методов извлечения правил из искусственных нейронных сетей / А.Н. Аверкин, С.А. Ярушев // Известия Российской академии наук. Теория и системы управления. 2021. Т. 6. № 6. С. 106-121. DOI 10.31857/S0002338821060044

[3] Аверкин А.Н. Перспективы применения объяснительного искусственного интеллекта в задачах персонализированной медицины / А.Н. Аверкин, С.А. Ярушев // Двадцатая Национальная конференция по искусственному интеллекту с международным участием : труды конференции, Москва, 21–23 декабря 2022 года. Том 1. Москва: Издательский дом МЭИ, 2022. С. 248-256.

[4] Ahmed S.B., Solis-Oba R., Ilie L. Explainable-AI in Automated Medical Report Generation Using Chest X-ray Images //Applied Sciences. 2022. Т. 12. №. 22. С. 11750. DOI 10.3390/app122211750

[5] Arrieta A.B. et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI //Information fusion. 2020. Т. 58. С. 82-115. DOI 10.1016/j.inffus.2019.12.012

[6] Bach S. et al. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation //PloS one. 2015. Т. 10. №. 7. С. e0130140. DOI 10.1371/journal.pone.0130140

[7] Chaddad A. et al. Survey of Explainable AI Techniques in Healthcare //Sensors. 2023. Т. 23. №. 2. С. 634. DOI 10.3390/s23020634

[8] Chen H. et al. Explainable medical imaging AI needs human-centered design: guidelines and evidence from a systematic review //npj Digital Medicine. 2022. Т. 5. №. 1. С. 156. DOI 10.1038/s41746-022-00699-2

[9] Giuste F. et al. Explainable artificial intelligence methods in combating pandemics: A systematic review //IEEE Reviews in Biomedical Engineering. 2022. DOI 10.1109/RBME.2022.3185953

[10] Hu Q. et al. Explainable artificial intelligence-based edge fuzzy images for COVID-19 detection and identification //Applied Soft Computing. 2022. Т. 123. С. 108966. DOI 10.1016/j.asoc.2022.108966

[11] Ieracitano C. et al. A fuzzy-enhanced deep learning approach for early detection of Covid-19 pneumonia from portable chest X-ray images //Neurocomputing. 2022. Т. 481. С. 202-215. DOI 10.1016/j.neucom.2022.01.055

[12] Ji J. Gradient-based interpretation on convolutional neural network for classification of pathological images //2019 International Conference on Information Technology and Computer Application (ITCA). – IEEE, 2019. С. 83-86. DOI 10.1109/ITCA49981.2019.00026

[13] Kamal M.S. et al. Explainable AI for glaucoma prediction analysis to understand risk factors in treatment planning //IEEE Transactions on Instrumentation and Measurement. 2022. Т. 71. С. 1-9. DOI 10.1109/TIM.2022.3171613

[14] Lundberg S.M., Lee S. I. A unified approach to interpreting model predictions //Advances in neural information processing systems. 2017. Т. 30.

[15] Matsuyama E., Watanabe H., Takahashi N. Explainable analysis of deep learning models for coronavirus disease (COVID-19) classification with chest X-Ray images: Towards practical applications //Open Journal of Medical Imaging. 2022. Т. 12. №. 3. С. 83-102. DOI 10.4236/ojmi.2022.123009

[16] Moura L.V. et al. Explainable machine learning for COVID-19 pneumonia classification with texture-based features extraction in chest radiography //Frontiers in digital health. 2022. Т. 3. С. 662343. – DOI 10.3389/fdgth.2021.662343

[17] Ribeiro M.T., Singh S., Guestrin C. " Why should i trust you?" Explaining the predictions of any classifier //Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. 2016. С. 1135-1144. DOI 10.1145/2939672.2939778

[18] Salahuddin Z. et al. Transparency of deep neural networks for medical image analysis: A review of interpretability methods //Computers in biology and medicine. 2022. Т. 140. С. 105111. DOI 10.1016/j.compbiomed.2021.105111

[19] Saraswat D. et al. Explainable AI for healthcare 5.0: opportunities and challenges // IEEE Access. 2022. DOI 10.1109/ACCESS.2022.3197671.

[20] Selvaraju R. R. et al. Grad-cam: Visual explanations from deep networks via gradient-based localization //Proceedings of the IEEE international conference on computer vision. 2017. С. 618-626.

[21] Suh Y.J., Jung J., Cho B.J. Automated breast cancer detection in digital mammograms of various densities via deep learning //Journal of personalized medicine. 2020. Т. 10. №. 4. С. 211. DOI 10.3390/jpm10040211

[22] Van der Velden B. H. M. et al. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis //Medical Image Analysis. 2022. С. 102470. DOI 10.1016/j.media.2022.102470

[23] Young K. et al. Deep neural network or dermatologist? //Interpretability of Machine Intelligence in Medical Image Computing and Multimodal Learning for Clinical Decision Support: Second International Workshop, iMIMIC 2019, and 9th International Workshop, ML-CDS 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Proceedings 9. Springer International Publishing, 2019. С. 48-55. DOI 10.1007/978-3-030-33850-3_6

[24] Zeineldin R.A. et al. Explainability of deep neural networks for MRI analysis of brain tumors //International journal of computer assisted radiology and surgery. 2022. Т. 17. №. 9. С. 1673-1683. DOI 10.1007/s11548-022-02619-x

[25] Zhou B. et al. Learning deep features for discriminative localization //Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. С. 2921-2929.