

Сравнительная оценка объяснимых моделей компьютерного зрения для обнаружения костных опухолей на рентгенограммах

Источник: Frontiers in Digital Health

Оригинал: <https://www.frontiersin.org/articles/10.3389/fdgth.2026.1926330>

классификация

компьютерное зрение

объяснимый ИИ (XAI)

онкология

радиология

рентгенография

сегментация изображений

Введение

Точное выявление костных опухолей на рентгенограммах представляет собой сложную задачу, поскольку внешний вид поражений варьируется, а их интерпретация требует экспертных знаний в узкой специализации. Данное исследование оценивает возможности моделей глубокого обучения (Deep Learning, DL) для локализации, сегментации, классификации и обеспечения объяснимости (explainability) костных опухолей с использованием общедоступного набора данных «Рентгенограммы костных опухолей» (Bone Tumour X-Ray Radiograph Dataset, BTXRD), который включает 3 746 рентгенограмм опорно-двигательного аппарата.

Методы

Модель YOLOv8s-seg была оценена совместно с классификаторами MobileNetV3-Large, EfficientNet-B3 и ConvNeXt-Tiny, обученными в рамках протокола единого разрешения. Карты активации Grad-CAM, Grad-CAM++ и Layer-CAM сравнивались с экспертными масками опухолей с использованием среднего значения пересечения (Intersection over Union, IoU) и пороговой меры перекрытия поражений. Для ретроспективного анализа согласия

EfficientNet-B3 был объединен с YOLOv8s-seg. Также было проведено исследование небольшой внешней серии, состоящей из 11 рентгенограмм с подтвержденными опухолями.

Результаты

Модель ConvNeXt-Tiny показала наивысшую точность бинарной классификации (0,879), тогда как YOLOv8s-seg достигла общей средней точности (mAP50) для ограничивающих рамок (box) на уровне 0,676 и для масок (mask) на уровне 0,668. Компоненты системы дали согласованные бинарные предсказания наличия опухоли для 605 из 750 тестовых рентгенограмм, что соответствует покрытию в 80,7% (95% доверительный интервал [ДИ]: 77,7%–83,3%). Точность бинарной классификации в пределах согласованной подвыборки составила 93,6% (566/605; 95% ДИ: 91,3%–95,2%), тогда как в оставшихся 145 случаях наблюдалось несоответствие, и консенсусное решение не было выдано.

Обсуждение

Анализ объяснимости искусственного интеллекта (XAI) количественно оценивает пространственное соответствие, а не достоверность объяснений или клиническую надежность. Поскольку в исследуемой внешней серии содержалось семь рентгенограмм с доброкачественными и четыре с злокачественными опухолями, но не было нормальных рентгенограмм, данная выборка не позволяет оценить специфичность или поведение ложноположительных результатов и не представляется как репрезентативная внешняя валидация.