

Гибридная трансформерная сеть с механизмом внимания для классификации плоскоклеточного рака полости рта с использованием гистопатологических изображений

Источник: Frontiers in AI — Medicine

Оригинал: <https://www.frontiersin.org/articles/10.3389/frai.2026.1751520>

гистопатология

глубокое обучение

диагностика

компьютерное зрение

онкология

трансформеры

Введение

В последние годы рак полости рта стал одной из наиболее распространенных злокачественных опухолей. Ранняя диагностика рака полости рта по гистопатологическим изображениям, безусловно, позволит снизить тяжесть заболевания и уменьшить уровень смертности. В литературе представлено несколько алгоритмов глубокого обучения (Deep Learning), варьирующихся от моделей сверточных нейронных сетей (Convolutional Neural Network, CNN) до моделей трансформеров компьютерного зрения (Vision Transformer, ViT), предназначенных для классификации нормальных тканей и тканей с плоскоклеточным раком полости рта (Oral Squamous Cell Carcinoma, OSCC).

Методы

В данном исследовании предлагается сеть CBA-TransNet (Convolutional Block Attention-aided Transformer Network), которая объединяет ResNet50 с ViT. ResNet50 выступает в качестве базовой архитектуры (backbone) для извлечения локальных признаков с помощью сверточных слоев, в то время

как ViT улавливает глобальный контекст и дальние зависимости через механизм самовнимания (self-attention) на гистопатологических изображениях. Для дальнейшего улучшения извлеченных признаков после слоя сети прямого распространения (Feed-Forward Network) в блоке энкодера ViT применяется механизм внимания сверточного блока (Convolutional Block Attention Mechanism, CBAM). CBAM обладает канальным и пространственным вниманием, что помогает трансформеру более эффективно фокусироваться на релевантных областях изображений.

Результаты

Для проведения экспериментов использовался общедоступный набор данных, состоящий из 5192 гистопатологических изображений. Результаты экспериментов и анализ показывают, что предложенная гибридная модель обеспечила точность 98,97% при сравнении с базовой предобученной моделью ResNet50, ViT, CNN и современными передовыми методами (state-of-the-art).

Обсуждение

Результаты экспериментов показывают, что предложенная сеть CBA-TransNet обладает гибкостью в сочетании архитектур на основе сверток и трансформеров вместе с механизмами внимания, такими как CBAM, для извлечения как локальных, так и глобальных признаков. Эта гибридная архитектура позволяет модели концентрироваться на диагностически значимых областях, что приводит к более качественной классификации.