

ПРИМЕНЕНИЕ ТРАНСФОРМЕРОВ ДЛЯ АНАЛИЗА ГИСТОПАТОЛОГИЧЕСКИХ ИЗОБРАЖЕНИЙ: АНАЛИЗ ПРЕИМУЩЕСТВ И НЕДОСТАТКОВ, ВОЗМОЖНЫЕ ПЕРСПЕКТИВЫ ДЛЯ ПРИМЕНЕНИЯ В ОБЛАСТИ ГИНЕКОЛОГИЧЕСКОЙ ПАТОМОРФОЛОГИИ

¹Асатурова А.В., ¹Бадлаева А.С., ¹Трегубова А.В., ²Рогожин Д.В.

¹ФГБУ «Национальный медицинский исследовательский центр акушерства, гинекологии и перинатологии имени академика В.И. Кулакова» Министерства здравоохранения Российской Федерации, Москва, e-mail: a.asaturova@gmail.com;

²ФГАОУ ВО «Российский Национальный Исследовательский Медицинский Университет им. Н.И. Пирогова», Москва

Трансформеры в последние годы приобрели широкую популярность в различных задачах компьютерного зрения, демонстрируя превосходство по сравнению со сверточными нейронными сетями. Их ключевое преимущество заключается в способности захватывать долгосрочные контекстные зависимости и моделировать сложные взаимосвязи в данных, что делает их особенно полезными для анализа изображений. Целью обзора было провести анализ публикаций по применению трансформеров для обработки гистопатологических изображений. Проанализированы литературные источники научных баз данных Web of Science, Scopus, Pubmed, Google Scholar, Российского индекса научного цитирования, из которых 52 источника были указаны в списке литературы, за период с 1997 по 2024 г. В статье анализируются новейшие модели трансформеров, их использование в задачах сегментации, классификации, предсказания выживаемости, а также в обнаружении и представлении данных. Особое внимание уделено сравнению эффективности трансформеров и сверточных нейронных сетей в этих задачах на основе недавно опубликованных исследований, преимуществам и недостаткам обоих методов глубокого изучения как с точки зрения перспектив развития каждого из них, так и с точки зрения архитектур, которые могут комбинировать оба подхода к глубокому обучению. Также выделены основные проблемы и ограничения применения трансформеров, такие как необходимость больших объемов данных и вычислительных ресурсов. В заключение предлагаются перспективные направления для будущих исследований, включая применение трансформеров в гинекологической патоморфологии, где они могут улучшить диагностику и прогнозирование злокачественных заболеваний.

Ключевые слова: глубокое обучение, нейросети, искусственный интеллект, трансформеры, гинекологическая патоморфология, цифровая патоморфология

Источник финансирования: государственное задание 122020900122-7.

APPLICATION OF TRANSFORMERS FOR HISTOPATHOLOGICAL IMAGE ANALYSIS: AN EXAMINATION OF ADVANTAGES AND DISADVANTAGES, POTENTIAL PROSPECTS FOR USE IN GYNECOLOGICAL PATHOLOGY

¹Asaturova A.V., ¹Badlaeva A.S., ¹Tregubova A.V., ²Rogozhin D.V.

¹National Medical Research Center for Obstetrics, Gynecology and Perinatology named after Academician V.I. Kulakov of the Ministry of Healthcare of Russian Federation, Moscow, e-mail: a.asaturova@gmail.com;

²Pirogov Russian National Research Medical University, Moscow

In recent years, transformers have gained widespread popularity in various computer vision tasks, demonstrating superiority over convolutional neural networks. Their key advantage lies in their ability to capture long-term contextual dependencies and model complex relationships in data, making them particularly useful for image analysis. The aim of this review was to analyze publications on the application of transformers for processing histopathological images. Materials and Methods: The literature sources from scientific databases such as Web of Science, Scopus, PubMed, Google Scholar, and the Russian Science Citation Index were analyzed, with 52 sources listed in the references, covering the period from 1997 to 2024. The article analyzes the latest transformer models and their use in tasks such as segmentation, classification, survival prediction, as well as data detection and representation. Special attention is given to comparing the effectiveness of transformers and convolutional neural networks in these tasks based on recently published research, highlighting the advantages and disadvantages of both deep learning methods, both in terms of their future development and in terms of architectures that can combine both approaches to deep learning. Key challenges and limitations in the application of transformers, such

as the need for large amounts of data and computational resources, are also outlined. Finally, potential future research directions are suggested, including the application of transformers in gynecological pathology, where they could improve the diagnosis and prognosis of malignant diseases.

Keywords: deep learning, neural networks, artificial intelligence, transformers, gynecological pathology, digital pathology

Source of funding: state assignment 122020900122-7.

Введение

Патоморфологическая диагностика по-прежнему является «золотым стандартом» для большого числа видов рака, имея несомненное преимущество по сравнению с другими методами медицинской визуализации, такими как магнитно-резонансная томография (МРТ), компьютерная томография (КТ), трансректальное ультразвуковое исследование (ТРУЗИ), маммография и многими другими [1]. Гистопатологические изображения являются важным источником первичной информации в клинической практике, они помогают патологоанатомам диагностировать опухолевые и опухолеподобные состояния с помощью идентификации морфологических признаков на клеточном и тканевом уровнях [2]. Кроме того, с развитием цифровой патоморфологии и анализа полноразмерных цифровых копий (whole slide image, WSI) гистологических препаратов, полученных с помощью сканирующих микроскопов, получила развитие новая область знаний, так называемая патомика [3, с. 16–19]. Патомика представляет собой раздел омиксных наук, в основе которой лежит получение дополнительной информации и биологии опухоли на основании оценки данных гистологических препаратов с помощью технологии глубокого обучения [4].

В цифровой патоморфологии гистопатологические слайды (WSI) представляют собой иерархическое образование визуальных токенов на различных уровнях разрешения и могут иметь размер пикселей до $160\,000 \times 160\,000$ при увеличении в 20 раз [5]. Цифровая патоморфология, то есть метод оцифровки гистопатологических изображений, представляет собой новый подход к сбору данных изображений для технологий искусственного интеллекта [6]. В последние годы методы искусственного интеллекта, такие как глубокое обучение, особенно сверточные нейронные сети (CNN), стали широко применяться для обработки и анализа гистопатологических изображений как в научных исследованиях, так и в клинической практике. CNN достигли значительных результатов во многих задачах компьютерного зрения [7, 8].

Недавно появилась альтернативная система компьютерной диагностики (CAD), способная моделировать долгосрочные зависимости между пикселями, такие как трансформеры. Трансформеры стали одной из последних технологических разработок в области глубокого обучения, показавшей высокие результаты в решении множества задач компьютерного зрения [9, 10]. Изначально они были разработаны как мощный пример

использования глубокого обучения для решения последовательных задач в области обработки естественного языка (NLP) [11]. В архитектуре Vision Transformer (ViT), предложенной Dosovitskiy и др., трансформеры применялись для задач классификации изображений, показав, что использование CNN для классификации изображений необязательно и что чистый трансформер, применяемый напрямую к последовательностям фрагментов изображения, может давать отличные результаты [12].

Трансформеры, в отличие от подходов на основе CNN, лишены предвзятости, вызванной свертками, что позволяет им захватывать долгосрочные контекстные зависимости и изучать более сложные взаимосвязи данных изображений. Это особенно полезно при анализе гистопатологических изображений, где важно учитывать не только интересующую область, но и окружающие ткани при диагностике определенного заболевания. С другой стороны, трансформеры требуют больших объемов данных и вычислительных мощностей, что может стать серьезной проблемой, особенно в области гистопатологических изображений, где доступ к ресурсам может быть ограничен из-за конфиденциальности данных пациентов [13, 14].

В настоящее время проведено множество исследований в области гистопатологических изображений с использованием подходов на основе трансформеров, включая сегментацию изображений, классификацию, обнаружение, представление, кросс-модальный поиск, генерацию изображений, анализ выживаемости и предсказание выживаемости [15–17]. Однако на основе недавно опубликованных исследований показано, что архитектуры на основе трансформеров способны достичь более высоких показателей производительности по сравнению с предыдущими моделями в различных задачах обработки гистопатологических изображений.

Целью исследования является предоставление всестороннего обзора применения трансформеров в области анализа гистопатологических изображений и демонстрация того, как трансформеры используются для решения различных задач.

Материалы и методы исследования

Для написания данного обзора литературы проанализированы статьи научных баз данных Web of Science, Scopus, Pubmed, Google Scholar, Российского индекса научного цитирования, из которых 52 источника были указаны в списке литературы, за период с 1997 по 2024 г. Из них более 70 % источников опубликованы за последние 10 лет. Для подготовки статьи использовались рекомендуемые принципы подготовки научных обзоров [18].

Результаты исследования и их обсуждение

*Предпосылки для поиска новых решений в области глубокого обучения:
архитектуры с использованием сверточных нейросетей*

На протяжении нескольких лет CNN показывали хорошие результаты при анализе данных изображений и являются наиболее часто используемыми глубокими нейронными сетями для решения медицинских и клинических задач, особенно в области анализа гистопатологических изображений. Это связано с тем, что сверточные операции в CNN накладывают ограничения на веса, что заставляет идентичные веса распределяться по каждому пикселю изображения. Основным преимуществом подходов на основе CNN является их способность автоматически выделять важные признаки на изображении без какого-либо вмешательства со стороны человека [19, 20].

Процесс построения любой архитектуры CNN для анализа гистопатологических изображений – это совместная работа исследователей и медицинских специалистов. Эти новшества в основном движимы большим количеством архитектурных усовершенствований, улучшением функций потерь, доступностью специализированных аппаратных устройств и публично доступных библиотек, созданных для конкретных целей [21, 22].

Несмотря на значительное количество архитектурных улучшений в методах на основе CNN за последние годы, их применение в полной мере к задачам анализа гистопатологических изображений ограничено зависимостью от больших объемов размеченных наборов данных. Исследования в области гистопатологических изображений для различных клинических задач также были улучшены моделями CNN. Однако CNN иногда функционируют как «черный ящик», и обычно их сложнее объяснить [23]. Успех методов на основе CNN в основном объясняется их способностью извлекать полезную информацию из входных изображений, устраняя необходимость в традиционных ручных методах обработки изображений. Несмотря на расширение области восприятия, CNN по-прежнему сталкиваются с многочисленными проблемами в моделировании долгосрочной информации, а также пространственных зависимостей из-за ограниченности сверточных операций [24].

Местная природа сверток в CNN является основной проблемой, поскольку она мешает захватывать долгосрочные семантические зависимости в исходных изображениях. Таким образом, требуется альтернативная система компьютерной диагностики (CAD), такая как трансформеры, которая способна моделировать долгосрочные пиксельные зависимости для достижения более точных результатов по сравнению с предыдущими моделями [25].

Трансформеры: основы

Архитектуры на основе трансформеров представляют собой наиболее продвинутую технику для работы с последовательностями. Они используют механизмы внимания (attention mechanisms) из-за их способности моделировать долгосрочную семантическую информацию. Трансформеры также используют стратегию проектирования на основе энкодера-декодера (encoder-decoder), что позволяет генерировать выходные данные без использования

рекуррентных слоев и сверток. Для начала представим основные идеи, лежащие в основе механизма внимания, а затем подробно объясним работу трансформера [26].

Механизм внимания

Механизм внимания естественным образом эволюционировал для решения задач, связанных с последовательностями данных. Сегодня он часто используется для выделения несущественной информации из данных с одновременным акцентом на важные части информации. Механизм внимания может применяться к ряду архитектур глубокого обучения в разных клинических областях и для различных задач. Механизм внимания впервые был разработан для улучшения работы архитектуры энкодера-декодера в задаче машинного перевода. Он был впервые предложен авторами для задачи перевода языков, чтобы решить проблему узкого места, вызванную использованием вектора фиксированной длины при кодировании, когда декодер имел ограниченный доступ к информации, передаваемой входными данными [27].

Архитектура трансформеров

Трансформеры, как правило, проектируются для работы с последовательностями и решения задач, связанных с долгосрочными зависимостями. В ставшей пионерской работе «Attention Is All You Need» была представлена стандартная архитектура трансформера, использующая структуру энкодера-декодера. В этой архитектуре блок энкодера преобразует входную последовательность в серию непрерывных представлений, а декодер генерирует результирующую последовательность на основе этих представлений. Энкодер трансформера состоит из идентичных слоев, которые последовательно извлекают признаки из входной последовательности [28]. Каждый слой включает две подслойные структуры, известные как механизм многоголового самовнимания (MHSA) и полносвязная нейронная сеть (FFN). Кроме того, через каждый подслой проходит остаточная связь, за которой следует нормализация по слоям. Сначала многоголовое внимание рассчитывается в каждом блоке, затем применяется блок нормализации по слоям. Сумма входных и выходных данных многоголового внимания рассчитывается с использованием нормализации по слоям. После применения полносвязного слоя входные и выходные данные MHSA также суммируются с нормализацией по слоям [29].

Декодер трансформера использует извлеченные признаки для генерации выходной последовательности. Он состоит из идентичных слоев с некоторыми модификациями [30]. Дополнительный подслой добавляется поверх закодированного выхода, который выполняет многоголовое внимание по результатам энкодера. Так как предсказание основывается на известном состоянии, в первом блоке самовнимания используется маскирование, чтобы предотвратить добавление новых данных к уже обработанному состоянию. В дополнение к

выходному слою декодера добавлены линейный и softmax слои для генерации окончательного результата.

Vision Transformer (ViT)

Изначально трансформеры были представлены для задач обработки естественного языка, где целью было понимание текста и получение полезных выводов. Архитектуры трансформеров достигли значительных результатов в этой области и стали стандартом в области NLP благодаря их обобщающим способностям и простоте. После успеха в задачах NLP исследователи начали адаптировать архитектуры трансформеров к различным задачам компьютерного зрения [31]. Одной из наиболее известных архитектур на основе трансформеров для задач компьютерного зрения является ViT (Vision Transformer). Было обозначено, что применение CNN больше не требуется и что чистая архитектура трансформера, применяемая непосредственно к последовательностям фрагментов изображений, может давать отличные результаты, особенно в задачах классификации изображений. Входное изображение разбивается на несколько фрагментов, каждый из которых кодируется с использованием пространственного кодирования для передачи пространственной информации. ViT продемонстрировали лучшие или даже превосходящие результаты по сравнению с передовыми (SOTA) CNN в решении многих задач, особенно при предварительном обучении на больших наборах данных [32].

Трансформеры широко используются во многих задачах компьютерного зрения и доказали свою способность давать лучшие результаты по сравнению с другими методами глубокого обучения. Некоторые преимущества трансформеров в задачах компьютерного зрения:

1. Эффективная параллельная обработка – благодаря механизмам внимания, они могут обрабатывать последовательности данных параллельно, что делает их более эффективными и быстрыми по сравнению со стандартными рекуррентными нейронными сетями.
2. Адаптивность к последовательностям переменной длины: архитектуры трансформеров могут работать с последовательностями данных разной длины, что упрощает их применение в различных задачах.
3. Эффективное управление глобальными зависимостями: трансформеры хорошо справляются с глобальными зависимостями благодаря механизмам самовнимания, что делает их подходящими для задач, требующих учета информации с разных частей изображения.
4. Высокая пропускная способность сети: трансформеры обладают большей емкостью для обучения более сложных зависимостей в данных [33, 34].

Однако у трансформеров есть и недостатки, такие как:

1. Высокие вычислительные затраты: трансформеры требуют больше вычислительных ресурсов по сравнению с другими методами глубокого обучения из-за механизма самовнимания, что приводит к значительным затратам времени и ресурсов на обучение.
2. Подверженность переобучению: поскольку трансформеры могут моделировать сложные взаимосвязи, они могут быть склонны к переобучению при недостатке данных.
3. Неэффективность в работе с небольшими объемами данных: трансформеры требуют большого количества данных для эффективного обучения, что становится проблемой в задачах, где доступен ограниченный набор размеченных данных [35, 36].

Сравнение методов трансформеров и CNN

На протяжении многих лет CNN показали выдающиеся результаты в анализе гистопатологических изображений, в то время как трансформеры, такие как ViT, показали более высокие результаты по сравнению с передовыми (SOTA) моделями CNN для многих задач. CNN обладают преимуществом в работе с массивами пикселей и поэтому легче в изучении и применении по сравнению с архитектурами трансформеров. Однако CNN ограничены в захвате дальнобойных корреляций между областями изображения из-за своего узкого поля восприятия. С другой стороны, трансформеры используют механизм самовнимания, который обрабатывает информацию из всего изображения, что позволяет им захватывать более отдаленные и важные части изображения [37–39].

Различные способы использования трансформеров для гистопатологических изображений

Недавно было проведено множество исследований по применению трансформеров для анализа гистопатологических изображений. Некоторые исследования пытались использовать чисто трансформеры (то есть трансформеры без сверточных блоков), в то время как другие интегрировали преимущества трансформеров (например, DETR, ViT, DeiT, BEiT, Swin-transformer) и CNN (например, EfficientNet, Unet, ResNet) для различных задач. В этом разделе они классифицированы на три типа:

1. **Чистые трансформеры:** к чистым трансформерам относятся архитектуры на основе ViT, которые не включают значительных структурных изменений. Они превосходят традиционные модели CNN по масштабируемости и эффективности как для малых, так и для больших вычислительных данных [40].
2. **Графовые методы на основе трансформеров:** эти сети вводят графы в традиционные Vision Transformers. Графы представляют собой типы данных, которые часто используются для представления биологических или социальных сетей, что делает их применимыми для анализа гистопатологических изображений. Примером такой модели является графовая сеть трансформеров (Graph Transformer Network, GTN) [41].

3. **Гибридные трансформеры и CNN:** в области анализа гистопатологических изображений существует множество способов комбинирования трансформеров с CNN для создания гибридной модели. Одним из самых простых способов является замена части сети на трансформер или интеграция трансформера в структуру сети с использованием CNN в качестве основы [42].

Классификация гистопатологических изображений

Vision Transformer (ViT) продемонстрировал выдающиеся результаты в задачах классификации естественных изображений с момента его появления. Согласно предыдущим исследованиям, подходы на основе трансформеров, использованные для изучения и прогнозирования рака, часто называют задачами классификации и разделяют на три категории. Во-первых, это прямое применение архитектур трансформеров к гистопатологическим изображениям. Во-вторых, это использование архитектур трансформеров в сочетании с конволюциями для лучшего изучения локальных признаков. В-третьих, это использование трансформеров вместе с графовыми представлениями для более эффективного управления данными сложных размеров [43, 44].

Таким образом, авторы могут обобщить текущие применения трансформеров для классификации гистопатологических изображений следующим образом:

1. Архитектуры трансформеров достигли равных или лучших результатов по сравнению с моделями на основе CNN в задачах классификации.
2. Трансформеры в настоящее время несколько ограничены в их применении, особенно в области гистопатологической визуализации, из-за необходимости в большом количестве аннотированных данных. Предобучение может стать альтернативным подходом для решения этой проблемы.
3. Обучение трансформеров на гигапиксельных изображениях требует значительных вычислительных затрат. Поэтому важно снизить вычислительные затраты моделей и разрабатывать облегченные архитектуры.
4. Большинство существующих архитектур трансформеров ориентированы на 2D-анализ гистопатологических изображений, но растет интерес к разработке моделей для работы с 3D-данными.
5. Гибридные трансформеры, которые объединяют в себе преимущества как трансформеров, так и традиционных сетей, таких как CNN и графовые нейронные сети (GNN), получают все большее внимание [45, 46].

Возможности трансформеров для гинекологической патоморфологии

Применение ИИ в области гинекологической патоморфологии не столь распространено, как при раке молочной железы, раке толстой кишки и простаты, однако с

каждым годом появляется все больше и больше решений в данной области. В этом разделе мы рассмотрим представленные решения по различным локализациям органов женской репродуктивной системы.

Шейка матки

Применение цитологических изображений для диагностики рака и предрака шейки матки давно используется в рамках коммерческих решений, однако почти все они основаны на применении CNN. Однако в работе Pascal представлена усовершенствованная архитектурная структура Multi-Axis Vision Transformer (MaxViT), разработанная для решения таких проблем, как вариабельность заключений в зависимости от уровня подготовки цитолога, а также качества преаналитической подготовки мазков. Адаптация MaxViT для данных Пап-теста позволила создать легковесную модель, обеспечивающую высокую точность и скорость анализа. Всего для каждого из включенных в исследование набора данных было использовано 106 моделей глубокого обучения, включая 53 модели CNN и 53 модели трансформеров для обработки изображений. В сравнении с экспериментальными и современными методами предложенный подход показал выдающуюся точность, превзойдя существующие решения (диагностическая точность составила 99,02–99,48 %) [47].

Тело матки

В одном из недавних исследований была представлена модель глубокого обучения EndoNet, которая может точно классифицировать препараты с раком эндометрия по категориям низкой и высокой степени злокачественности, что помогает патологоанатомам в диагностике и планировании лечения. Главная идея данного исследования состоит в том, что EndoNet, использующая сверточные нейронные сети (CNN) и Vision Transformer (ViT), демонстрирует высокую точность в классификации слайдов с раком эндометрия, что было подтверждено результатами тестирования как на внутренних, так и на внешних наборах данных [48].

Кроме того, в исследовании применения трансформеров для предикции статуса микросателлитной нестабильности были протестированы различные алгоритмы глубокого обучения, включая CNN и модели Vision Transformers. Исследование показало, что модели CNN, такие как ResNet, продемонстрировали более высокую точность в предсказании статуса MMR по сравнению с Vision Transformers. Тем не менее ViT оказались полезными, так как они предложили новый способ анализа изображений, фокусируясь на конкретных областях интереса с помощью механизма внимания [49].

В исследовании Volinsky-Fremont и соавт. глубокое обучение на основе трансформеров было применено для прогнозирования отдаленного рецидива рака эндометрия, что имеет решающее значение для персонализированного подбора адъювантной терапии. Была

разработана система HECTOR (Histopathology-based Endometrial Cancer Tailored Outcome Risk) – мультимодальная прогностическая модель глубокого обучения, использующая окрашенные гематоксилином и эозином изображения гистологических препаратов и стадию опухоли в качестве входных данных. Модель была обучена на данных 2072 пациентов из 8 когорт РЭ, включая рандомизированные исследования PORTEC-1/-2/-3. HECTOR продемонстрировал индекс согласия (С-индекс) 0,789; 0,828 и 0,815 на внутренних ($n = 353$) и двух внешних ($n = 160$ и $n = 151$) тестовых наборах данных соответственно, превзойдя современный «золотой стандарт». Модель также идентифицировала группы пациентов с резко различающимися прогнозами: вероятность 10-летнего безрецидивного течения для групп низкого, среднего и высокого риска по HECTOR составила 97,0 %, 77,7 % и 58,1 % соответственно (по анализу Каплана – Мейера). Было показано, что HECTOR также лучше, чем текущие методы, прогнозировал пользу от адъювантной химиотерапии. Экстракция морфологических и геномных характеристик выявила корреляции с группами риска по HECTOR, некоторые из которых имеют терапевтический потенциал [50].

Яичники

В исследовании Alahmadi представлен передовой фреймворк компьютерной диагностики для классификации рака яичников, объединяющий модели Vision Transformer (ViT) и метод локальных интерпретируемых объяснений, независимых от модели (Local Interpretable Model-agnostic Explanations, LIME). Модели ViT, включая ViT-Base-P16-224-In21K, ViT-Base-P16-224, ViT-Base-P32-384 и ViT-Large-P32-384, продемонстрировали исключительную точность, полноту, чувствительность и общую надежность по разнообразным метрикам оценки. Использование составной (stacked) модели дополнительно повысило общую производительность. Результаты экспериментов, проведенных на тренировочных и тестовых наборах данных в данном исследовании, подчеркивают высокую эффективность моделей ViT в точной классификации подтипов рака яичников на основе гистопатологических изображений. Модель ViT-Large-P32-384 была выделена как наиболее эффективная, достигнув точности 98,79 % на тренировочном этапе и 97,37 % на этапе тестирования. Визуализация, включая ROC-кривые и метод LIME, показала важные данные о дискриминативных способностях моделей, что свидетельствует о потенциальном повышении интерпретируемости результатов. Таким образом, предложенный фреймворк CAD представляет собой значительный прорыв в диагностике рака яичников, предлагая перспективный подход для точной и прозрачной многоклассовой классификации гистопатологических изображений [51].

В работе, посвященной индексу цифровой патологии рака яичников (OCDPI), авторы использовали графовую модель глубокого обучения, разработанную для прогноза

выживаемости и реакции на адъювантную терапию у пациентов с раком яичников с использованием полноразмерных сканированных копий срезов, окрашенных гематоксилином и эозином. Модель была разработана на основе данных из проекта TCGA-OV и прошла внешнюю валидацию на двух независимых когортах. OCDPI продемонстрировала значительные прогностические способности в отношении общей выживаемости в обеих валидационных когортах и стабильно выступала в качестве независимого прогностического фактора, даже с учетом клинико-патологических факторов. Модель также эффективно прогнозировала лучшую выживаемость и более низкие показатели рецидивов у пациентов с низкими значениями OCDPI после адъювантной терапии. Кроме того, модель показала хорошие результаты в группах пациентов с низкодифференцированными опухолями или с дефицитом гомологичной рекомбинации (HRD), что указывает на ее потенциал в персонализации лечения [52].

Таким образом, можно заключить, что в области гинекологической патоморфологии, так же как и в диагностике злокачественных заболеваний других локализаций, трансформеры обычно имеют преимущества в решении задач классификации грейдинга опухоли, предикции мутационного статуса и прогноза выживаемости благодаря своим механизмам внимания, которые позволяют модели фокусироваться на ключевых особенностях внутри WSI. Это может усилить способность таких алгоритмов улавливать как локальные, так и глобальные гистологические паттерны, улучшая точность прогноза и ответа на лечение, особенно в случае сложных и гетерогенных опухолей, таких как рак яичников. Эффективность трансформеров в обработке изображений всего слайда на большом масштабе также делает алгоритмы с их использованием инструментами, экономящими трудозатраты, что способствует более информированному и персонализированному принятию решений в лечении пациентов с гинекологическими раками разных локализаций.

Выводы

На основе проведенного обзора можно сделать следующие выводы:

- 1) трансформеры продемонстрировали выдающиеся результаты в задачах классификации и сегментации гистопатологических изображений;
- 2) существующие архитектуры трансформеров имеют определенные ограничения, связанные с высокими вычислительными затратами и потребностью в больших наборах данных для обучения;
- 3) будущие исследования должны быть направлены на разработку более легких и эффективных моделей, которые смогут работать с меньшими объемами данных и меньшими затратами на вычисления;

- 4) активное развитие гибридных архитектур, объединяющих преимущества как трансформеров, так и CNN, является перспективным направлением;
- 5) в будущем возможна интеграция трехмерных трансформеров для работы с объемными данными, что может значительно улучшить результаты в задачах анализа медицинских изображений.

Заключение

Архитектуры на основе трансформеров открыли новые горизонты для анализа гистопатологических изображений. Эти методы доказали свою эффективность, особенно в задачах сегментации и прогнозирования выживаемости. Несмотря на текущие успехи, дальнейшие исследования необходимы для решения проблем, связанных с высокой сложностью моделей и потребностью в большом количестве данных. Гибридные подходы, сочетающие CNN и трансформеры, предоставляют исследователям уникальные возможности для разработки новых методов анализа медицинских данных. Первые успехи в использовании таких архитектур в решении задач гинекологической патоморфологии открывают новые перспективы для разработки эффективных инструментов для диагностики, прогнозирования и предикции злокачественных новообразований матки и яичников, а также в области репродуктивной патологии.

Список литературы

1. Rorke LB. Pathologic diagnosis as the gold standard // Cancer. 1997. № 79 (4). P. 665–667. DOI: 10.1002/(SICI)1097-0142(19970215)79:4<665::AID-CNCR1>3.0.CO;2-D.
2. Funkhouser WK. Pathology: The Clinical Description of Human Disease // Molecular Pathology. 2009. № 10. P. 197–207. DOI: 10.1016/B978-0-12-374419-7.00011-1.
3. Парвани А.В. Цифровая патоморфология / пер. с англ. под ред. А.В. Асатуровой, Д.Д. Проценко М.: Практическая медицина, 2024. 272 с.
4. Bülow R.D., Hölscher D.L., Costa I.G., Boor P. Extending the landscape of omics technologies by pathomics // NPJ Syst. Biol. Appl. 2023. № 9(1). P. 38. DOI: 10.1038/s41540-023-00301-9.
5. Atabansi C.C., Nie J., Liu H., Song Q., Yan L., Zhou X. A survey of Transformer applications for histopathological image analysis: New developments and future directions // Biomed Eng Online. 2023. № 22 (1). P. 96. DOI: 10.1186/s12938-023-01157-0.
6. Maleki D., Rahnamayan S., Tizhoosh H.R. A self-supervised framework for cross-modal search in histopathology archives using scale harmonization // Sci Rep. 2024. № 14. P. 9724. DOI: 10.1038/s41598-024-60256-7.

7. Yamashita R., Nishio M., Do R.K.G., Togashi K. Convolutional neural networks: an overview and application in radiology // *Insights Imaging*. 2018. № 9 (4). P. 611–629.
DOI: 10.1007/s13244-018-0639-9.
8. Li Z., Liu F., Yang W., Peng S., Zhou J. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects // *IEEE Trans Neural Netw Learn Syst*. 2022. № 33 (12). P. 6999–7019.
DOI: 10.1109/TNNLS.2021.3084827.
9. Choi S.R., Lee M. Transformer Architecture and Attention Mechanisms in Genome Data Analysis: A Comprehensive Review // *Biology (Basel)*. 2023. № 12 (7). P. 1033.
DOI: 10.3390/biology12071033.
10. Jia Y., Wu R., Lu X., Duan Y., Zhu Y., Ma Y., Nie F. Deep Learning with Transformer or Convolutional Neural Network in the Assessment of Tumor-Infiltrating Lymphocytes (TILs) in Breast Cancer Based on US Images: A Dual-Center Retrospective Study // *Cancers (Basel)*. 2023. № 15 (3). P. 838. DOI: 10.3390/cancers15030838.
11. Hommel B.E., Wollang F.M., Kotova V., Zacher H., Schmukle S.C. Transformer-Based Deep Neural Language Modeling for Construct-Specific Automatic Item Generation // *Psychometrika*. 2022. № 87 (2). P. 749–772. DOI: 10.1007/s11336-021-09823-9.
12. Li J., Chen J., Tang Y., Wang C., Landman B.A., Zhou S.K. Transforming medical imaging with Transformers? A comparative review of key properties, current progresses, and future perspectives // *Med Image Anal*. 2023. № 85. P.102762. DOI: 10.1016/j.media.2023.102762.
13. Wu J.H., Koseoglu N.D., Jones C., Liu T.Y.A. Vision transformers: The next frontier for deep learning-based ophthalmic image analysis // *Saudi J Ophthalmol*. 2023. № 37 (3). P. 173–178.
DOI: 10.4103/sjopt.sjopt_91_23.
14. Moutik O., Sekkat H., Tigani S., Chehri A., Saadane R., Tchakoucht T.A., Paul A. Convolutional Neural Networks or Vision Transformers: Who Will Win the Race for Action Recognitions in Visual Data? // *Sensors (Basel)*. 2023. № 23 (2). P. 734. DOI: 10.3390/s23020734.
15. Khan R.F., Lee B.D., Lee M.S. Transformers in medical image segmentation: a narrative review // *Quant Imaging Med Surg*. 2023. № 13 (12). P. 8747–8767. DOI: 10.21037/qims-23-542.
16. Shinohara H., Kodera S., Nagae Y., Hiruma T., Kobayashi A., Sato M., Sawano S., Kamon T., Narita K., Hirose K., Kiriya H., Saito A., Miura M., Minatsuki S., Kikuchi H., Takeda N., Akazawa H., Morita H., Komuro I. The potential of the transformer-based survival analysis model, SurvTrace, for predicting recurrent cardiovascular events and stratifying high-risk patients with ischemic heart disease // *PLoS One*. 2024. № 19 (6). P. e0304423. DOI: 10.1371/journal.pone.0304423.
17. Zheng Y., Qiu B., Liu S., Song R., Yang X., Wu L., Chen Z., Tuersun A., Yang X., Wang W., Liu Z. A transformer-based deep learning model for early prediction of lymph node metastasis in

locally advanced gastric cancer after neoadjuvant chemotherapy using pretreatment CT images // *EClinicalMedicine*. 2024. № 75. P. 102805. DOI: 10.1016/j.eclinm.2024.102805.

18. Белобородов В.А., Воробьев В.А., Семинский И.Ж., Калягин А.Н. Порядок выполнения систематического обзора и мета-анализа по протоколу PRISMA // *Система менеджмента качества: опыт и перспективы*. 2023. № 12. С. 5–9.

19. Gertych A., Swiderska-Chadaj Z., Ma Z., Ing N., Markiewicz T., Cierniak S., Salemi H., Guzman S., Walts A.E., Knudsen B.S. Convolutional neural networks can accurately distinguish four histologic growth patterns of lung adenocarcinoma in digital slides // *Sci Rep*. 2019. № 9. P. 1483. DOI: 10.1038/s41598-018-37638-9.

20. Shafi S., Parwani A.V. Artificial intelligence in diagnostic pathology // *Diagn Pathol*. 2023. № 18. P. 109. DOI: 10.1186/s13000-023-01375-z.

21. Meirelles A.L.S., Kurc T., Kong J., Ferreira R., Saltz J.H., Teodoro G. Building Efficient CNN Architectures for Histopathology Images Analysis: A Case-Study in Tumor-Infiltrating Lymphocytes Classification // *Front Med (Lausanne)*. 2022. № 9. P. 894430. DOI: 10.3389/fmed.2022.894430.

22. Springenberg M., Frommholz A., Wenzel M., Weicken E., Ma J., Strodthoff N. From modern CNNs to vision transformers: Assessing the performance, robustness, and classification strategies of deep learning models in histopathology // *Med Image Anal*. 2023. № 87. P. 102809. DOI: 10.1016/j.media.2023.102809.

23. Prijs J., Liao Z., To M.S., Verjans J., Jutte P.C., Stirler V., Olczak J., Gordon M., Guss D., DiGiovanni C., Jaarsma R.L., Ijpma F., Doornberg J. Development and external validation of automated detection, classification, and localization of ankle fractures: inside the black box of a convolutional neural network (CNN) // *Eur. J. Trauma Emerg Surg*. 2023 № 49 (2). P. 1057–1069. DOI: 10.1007/s00068-022-02136-1.

24. Alzubaidi L., Zhang J., Humaidi A.J., Al-Dujaili A., Duan Y., Al-Shamma O., Santamaría J., Fadhel M.A., Al-Amidie M., Farhan L. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions // *J. Big Data*. 2021. № 8 (1). P. 53. DOI: 10.1186/s40537-021-00444-8.

25. Alharthi A.G., Alzahrani S.M. Do it the transformer way: A comprehensive review of brain and vision transformers for autism spectrum disorder diagnosis and classification // *Comput. Biol. Med*. 2023. № 167. P. 107667. DOI: 10.1016/j.combiomed.2023.107667.

26. Zhang Y., Wang J., Gorriz J.M., Wang S. Deep Learning and Vision Transformer for Medical Image Analysis // *J. Imaging*. 2023. № 9 (7). P. 147. DOI: 10.3390/jimaging9070147.

27. Bahdanau D., Cho K., Bengio Y. Neural machine translation by jointly learning to align and translate // *arXiv preprint arXiv*. 2014. № 1409. P. 0473.

28. Choi S.R., Lee M. Transformer Architecture and Attention Mechanisms in Genome Data Analysis: A Comprehensive Review // *Biology (Basel)*. 2023. № 12 (7). P. 1033. DOI: 10.3390/biology12071033.
29. Sibilano E., Brunetti A., Buongiorno D., Lassi M., Grippo A., Bessi V., Micera S., Mazzoni A., Bevilacqua V. An attention-based deep learning approach for the classification of subjective cognitive decline and mild cognitive impairment using resting-state EEG // *J Neural Eng*. 2023 № 20 (1). DOI: 10.1088/1741-2552/acb96e.
30. Yang Z., Mitra A., Liu W., Berlowitz D., Yu H. TransformEHR: transformer-based encoder-decoder generative model to enhance prediction of disease outcomes using electronic health records // *Nat Commun*. 2023. № 14 (1). P. 7857. DOI: 10.1038/s41467-023-43715-z.
31. Courant R., Edberg M., Dufour N., Kalogeiton V. Transformers and Visual Transformers. 2023 Jul 23. In: Colliot O., editor. *Machine Learning for Brain Disorders* [Internet]. NY: Humana, 2023. Chapter 6. P. 133–233.
32. Liu Z., Qian S., Xia C., Wang C. Are transformer-based models more robust than CNN-based models? // *Neural Netw*. 2024. № 172. P. 106091. DOI: 10.1016/j.neunet.2023.12.045.
33. Denecke K., May R., Rivera-Romero O. Transformer Models in Healthcare: A Survey and Thematic Analysis of Potentials., Shortcomings and Risks // *J Med Syst*. 2024. № 48 (1). P. 23. DOI: 10.1007/s10916-024-02043-5.
34. Holderried F., Stegemann-Philipps C., Herschbach L., Moldt J.A., Nevins A., Griewatz J., Holderried M., Herrmann-Werner A., Festl-Wietek T., Mahling M. A Generative Pretrained Transformer (GPT)-Powered Chatbot as a Simulated Patient to Practice History Taking: Prospective, Mixed Methods Study // *JMIR Med Educ*. 2024. № 10. P. e53961. DOI: 10.2196/53961.
35. Gao S., Alawad M., Young M.T., Gounley J., Schaefferkoetter N., Yoon H.J., Wu X.C., Durbin E.B., Doherty J., Stroup A., Coyle L., Tourassi G. Limitations of Transformers on Clinical Text Classification // *IEEE J Biomed Health Inform*. 2021. № 25 (9). P. 3596–3607. DOI: 10.1109/JBHI.2021.3062322.
36. Hong Z., Chen M., Hu W., Yan S., Qu A., Chen L., Chen J. Dual encoder network with transformer-CNN for multi-organ segmentation // *Med Biol Eng Comput*. 2023. № 61 (3). P. 661–671. DOI: 10.1007/s11517-022-02723-9.
37. Fanizzi A., Fadda F., Comes M.C., Bove S., Catino A., Di Benedetto E., Milella A., Montrone M., Nardone A., Soranno C., Rizzo A., Guven D.C., Galetta D., Massafra R. Comparison between vision transformers and convolutional neural networks to predict non-small lung cancer recurrence // *Sci Rep*. 2023. № 13 (1). P. 20605. DOI: 10.1038/s41598-023-48004-9.
38. Takahashi S., Sakaguchi Y., Kouno N., Takasawa K., Ishizu K., Akagi Y., Aoyama R., Teraya N., Bolatkan A., Shinkai N., Machino H., Kobayashi K., Asada K., Komatsu M., Kaneko S.,

- Sugiyama M., Hamamoto R. Comparison of Vision Transformers and Convolutional Neural Networks in Medical Image Analysis: A Systematic Review // J Med Syst. 2024. № 48 (1). P. 84. DOI: 10.1007/s10916-024-02105-8.
39. Moutik O., Sekkat H., Tigani S., Chehri A., Saadane R., Tchakoucht T.A., Paul A. Convolutional Neural Networks or Vision Transformers: Who Will Win the Race for Action Recognitions in Visual Data? // Sensors (Basel). 2023. № 23 (2). P. 734. DOI: 10.3390/s23020734.
40. Azad R., Kazerouni A., Heidari M., Aghdam E.K., Molaei A., Jia Y., Jose A., Roy R., Merhof D. Advances in medical image analysis with vision Transformers: A comprehensive review // Med Image Anal. 2024. № 91. P. 103000. DOI: 10.1016/j.media.2023.103000
41. Zheng Y., Gindra R.H., Green E.J., Burks E.J., Betke M., Beane J.E., Kolachalama V.B. A Graph-Transformer for Whole Slide Image Classification // IEEE Trans Med Imaging. 2022 № 41 (11). P. 3003–3015. DOI: 10.1109/TMI.2022.3176598.
42. Li X., Pang S., Zhang R., Zhu J., Fu X., Tian Y., Gao J. ATTransUNet: An enhanced hybrid transformer architecture for ultrasound and histopathology image segmentation // Comput Biol Med. 2023. № 152. P.106365. DOI: 10.1016/j.compbiomed.2022.106365.
43. Xu H., Xu Q., Cong F., Kang J., Han C., Liu Z., Madabhushi A., Lu C. Vision Transformers for Computational Histopathology // IEEE Rev Biomed Eng. 2024. № 17. P. 63–79. DOI: 10.1109/RBME.2023.3297604.
44. Zhang J., Lu J.D., Chen B., Pan S., Jin L., Zheng Y., Pan M. Vision transformer introduces a new vitality to the classification of renal pathology // BMC Nephrol. 2024. № 9. P. 25 (1): 337. DOI: 10.1186/s12882-024-03800-x.
45. Halder A., Gharami S., Sadhu P., Singh P.K., Woźniak M., Ijaz M.F. Implementing vision transformer for classifying 2D biomedical images // Sci Rep. 2024. № 14 (1). P. 12567. DOI: 10.1038/s41598-024-63094-9.
46. Baroni G.L., Rasotto L., Roitero K., Tullisso A., Di Loreto C., Della Mea V. Optimizing Vision Transformers for Histopathology: Pretraining and Normalization in Breast Cancer Classification // J. Imaging. 2024. № 10 (5). P. 108. DOI: 10.3390/jimaging10050108.
47. Pacal. I. Maxcervixt: A novel lightweight vision transformer-based approach for precise cervical cancer detection // Knowl.-Based Syst. 2024. № 289. P. 111482. DOI: 10.1016/j.knosys.2024.111482.
48. Goyal M., Tafe L.J., Feng J.X., Muller K.E., Hondelink L., Bentz J.L., Hassanpour S. Deep Learning for Grading Endometrial Cancer // Am J Pathol. 2024. № 194 (9). P. 1701–1711. DOI: 10.1016/j.ajpath.2024.05.003.
49. Umemoto M., Mariya T., Nambu Y., Nagata M., Horimai T., Sugita S., Kanaseki T., Takenaka Y., Shinkai S., Matsuura M., Iwasaki M., Hirohashi Y., Hasegawa T., Torigoe T., Fujino Y., Saito T.

Prediction of Mismatch Repair Status in Endometrial Cancer from Histological Slide Images Using Various Deep Learning-Based Algorithms // *Cancers (Basel)*. 2024. № 16 (10). P. 1810.

DOI: 10.3390/cancers16101810.

50. Volinsky-Fremont S., Horeweg N., Andani S., Barkey Wolf J., Lafarge M.W., de Kroon C.D., Ørtoft G., Høgdall E., Dijkstra J., Jobsen J.J., Lutgens L.C.H.W., Powell M.E., Mileshekin L.R., Mackay H., Leary A., Katsaros D., Nijman H.W., de Boer S.M., Nout R.A., de Bruyn M., Church D., Smit V.T.H.B.M., Creutzberg C.L., Koelzer V.H., Bosse T. Prediction of recurrence risk in endometrial cancer with multimodal deep learning // *Nat Med*. 2024. № 30 (7). P. 1962–1973.

DOI: 10.1038/s41591-024-02993-w.

51. A. Alahmadi. Towards ovarian cancer diagnostics: A vision transformer-based computer-aided diagnosis framework with enhanced interpretability *Results in Engineering*. 2024. № 23 (1). P. 102651. DOI: 10.1016/j.rineng.2024.102651.

52. Yang Z., Zhang Y., Zhuo L., Sun K., Meng F., Zhou M., Sun J. Prediction of prognosis and treatment response in ovarian cancer patients from histopathology images using graph deep learning: a multicenter retrospective study // *Eur. J. Cancer*. 2024. № 199. P. 113532.

DOI: 10.1016/j.ejca.2024.113532.