

Систематический анализ метрик оценки для глубокого обучения в распознавании этапов хирургических вмешательств при лапароскопической холецистэктомии

Источник: Frontiers in AI — Medicine

Дата публикации: 2025

Оригинал: <https://www.frontiersin.org/articles/10.3389/frai.2026.1873219>

анализ метрик

глубокое обучение

компьютерное зрение

лапароскопия

хирургия

Метрики оценки имеют важное значение для оценки эффективности моделей искусственного интеллекта (AI) и обеспечения сопоставимости различных подходов. Однако в области распознавания фаз хирургических вмешательств их применение остается непоследовательным. Данное исследование представляет собой систематический анализ метрик оценки, используемых в этой области, с целью содействия стандартизации и стимулирования более четкого обоснования выбора метрик.

Был проведен систематический обзор литературы в соответствии с протоколом PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), охватывающий публикации с 2016 по 2025 год. В общей сложности было выявлено и проанализировано 50 исследований с точки зрения применяемых метрик оценки, при этом особое внимание уделялось «смягченным границам» (relaxed boundaries) и различиям между онлайн- и офлайн-моделями.

Результаты показывают, что наиболее часто используемыми метриками являются **точность** (accuracy), **прецизионность** (precision) и **полнота** (recall), за которыми следует **индекс Жаккара** (Jaccard index). Начиная примерно с 2023 года наблюдается рост использования метрик, основанных на сегментации, что отражает растущий акцент на временной динамике.

Значимых различий в выборе метрик между онлайн- и офлайн-подходами выявлено не было. Более того, во многих исследованиях отсутствует явное объяснение выбора метрик оценки. Вместо этого авторы часто полагаются на общепринятые метрики или практики, заимствованные из предыдущих работ.

Кроме того, результаты, полученные с использованием «смягченных границ», имеют тенденцию демонстрировать более высокие стандартные отклонения по сравнению с результатами без использования «смягченных границ».

На основании этих результатов рекомендуется более явно приводить метрики оценки в соответствии с целями модели и продолжить изучение различий между онлайн- и офлайн-режимами. Адаптации **F1-меры**, такие как **Fβ-мера**, могут обеспечить более гибкую структуру оценки. Кроме того, **коэффициент корреляции Мэтьюса** (Matthews Correlation Coefficient) представляет собой перспективную дополнительную метрику, особенно для многоклассовых задач, при условии его надлежащей нормализации.

Будущие работы должны быть сосредоточены на четком обосновании выбора метрик и последовательном представлении данных с «смягченными границами» для повышения прозрачности, сопоставимости и интерпретируемости в задачах распознавания фаз хирургических вмешательств.