

SCRIPT: Стратифицированное прогнозирование клинического риска на основе патоморфологических отчетов с использованием больших языковых моделей

Источник: Journal of Pathology Informatics

Оригинал: https://www.sciencedirect.com/science/article/pii/S215335392600132X?dgcid=rss_sd_all

LLM

диагностика

онкология

патоморфология

прогностическая аналитика

Точная стратификация рисков в онкологии имеет решающее значение для принятия решений о лечении, однако современные алгоритмы опираются на узкий набор структурированных переменных и, следовательно, потенциально игнорируют богатый массив данных, содержащийся в описательных патоморфологических отчетах. Эти отчеты содержат нюансированные морфологические описания и экспертные клинические суждения, однако данная повествовательная информация остается в значительной степени неиспользованной в процессе принятия клинических решений, так как она теряется в текстовых отчетах, представленных в виде «прозы». Мы выдвинули гипотезу, что большие языковые модели (LLM — Large Language Models) могут извлекать прогностическую информацию из полных текстовых патоморфологических отчетов и преобразовывать ее в бинарный биомаркер выживаемости.

Методы

Мы использовали модель с открытыми весами **LLaMA 3.3 70B** для генерации показателей риска непосредственно на основе общедоступных патоморфологических отчетов по трем типам рака желудочно-кишечного тракта. Модели был дан промпт (инструкция) на синтез полных описательных отчетов в единый бинарный прогностический показатель. Мы оценивали взаимосвязь между показателями, сгенерированными LLM, и исходами выживаемости, включая общую выживаемость, выживаемость без прогрессирования и специфическую выживаемость при данном заболевании.

Результаты

При колоректальном раке показатели риска, сгенерированные LLM, продемонстрировали значительную прогностическую ценность для общей выживаемости (**отношение рисков (HR — Hazard Ratio) = 2,77, 95% доверительный интервал (CI — Confidence Interval) = 1,92-3,97, $p < 0,001$**), выживаемости без прогрессирования (HR = 2,93, 95% CI = 2,11-4,08, $p < 0,001$) и специфической выживаемости при заболевании (HR = 5,85, 95% CI = 3,66-9,36, $p < 0,001$). Многофакторный анализ подтвердил, что показатель риска, сгенерированный LLM, является независимым прогностическим фактором выживаемости без прогрессирования.

Заключение

LLM способны преобразовывать описательные патоморфологические отчеты в единый независимый биомаркер выживаемости. Этот подход позволяет использовать рутинно доступную текстовую документацию без необходимости проведения дополнительного анализа тканей или увеличения нагрузки на патоморфологов, предоставляя готовый к внедрению метод для улучшения стратификации рисков при принятии решений о лечении.