

Новый анализ вложенного конформного прогнозирования для раскрытия сложности при подтипировании пациентов

Источник: Frontiers in AI — Medicine

Оригинал: <https://www.frontiersin.org/articles/10.3389/frai.2026.1844254>

биоинформатика

диагностика

конформное прогнозирование

машинное обучение

онкология

Подтипирование пациентов значительно осложняется внутриобразцовой гетерогенностью, что ограничивает эффективность традиционных подходов многоклассовой классификации, навязывающих взаимоисключающие метки. Несмотря на недавние многообещающие результаты перехода от многоклассовой к многометковой классификации, этот процесс не является прямолинейным и оказывается трудносистематизируемым, особенно при работе с малоразмерными наборами данных.

В данной работе мы разрабатываем новый подход, реализованный в рамках вычислительной платформы, который использует модели машинного обучения и оригинальную стратегию вложенного конформного прогнозирования (**nested conformal prediction**) для малых наборов данных, чтобы обеспечить строгий переход к многометковой классификации. Конформное прогнозирование действительно может предложить статистически обоснованный подход для получения надежных прогнозов подтипов; однако до сих пор его использование ограничивалось только большими наборами данных, которые часто недоступны в реальных биомедицинских сценариях.

Мы оцениваем наш подход, используя данные секвенирования РНК (**RNA-seq**) пациенток с раком молочной железы из проекта «Атлас ракового генома» (**The Cancer Genome Atlas**, TCGA), где подтипы изначально были определены с помощью классификатора **PAM50**. С методологической точки зрения наша инновационная вложенная стратегия позволяет успешно применять конформное прогнозирование, которое обычно требует больших наборов данных, к относительно малому объему выборки.

В частности, конформное прогнозирование адаптируется и применяется как к эталонной классификации **PAM50**, так и к нескольким моделям машинного обучения для проверки согласованности и сравнения полученных многометковых прогнозов. Наши результаты выявляют сложность, специфичную для конкретных подтипов и образцов, и подчеркивают устойчивые паттерны неоднозначности в различных моделях.

Что немаловажно, мы разделяем присвоения на надежные, внутренне гетерогенные и слабые, что дает более глубокое понимание как четких подтверждений подтипов и сложности, так и критически несогласованных прогнозов. Таким образом, в целом, данная работа демонстрирует полезность нашей платформы для подтипирования образцов с учетом неопределенности, особенно в условиях малых и сильно несбалансированных выборок.

Перевод выполнен: 20.07.2026 | ai4med.ru

Машинный перевод. Рекомендуем сверять с оригиналом при клиническом использовании.