

Пан-раковый мультиомический SuperLearner для топологий выживаемости, связанных с регулируемой клеточной гибелью

Источник: Frontiers in AI — Medicine

Оригинал: <https://www.frontiersin.org/articles/10.3389/frai.2026.1895164>

SuperLearner

выживаемость

машинное обучение

мультиомика

онкология

персонализированная медицина

Введение

Пути регулируемой клеточной гибели (РКГ) влияют на прогрессию опухоли и иммунную модуляцию. Ранее мы создали базу данных сигнатур, картировавшую 25 форм РКГ на семи мультиомических уровнях и 33 типах опухолей (CancerRCDShiny). Несмотря на способность этих сигнатур выявлять группы риска, их перевод в персонализированные клинические рабочие процессы требует перехода от стратификации когорт к индивидуальному картированию риска путем моделирования риска пациента (топологий выживаемости) для учета нелинейной динамики сигнатур РКГ.

Методы

Мы разработали пан-раковый мультиомический конвейер SuperLearner для 33 типов рака. Фаза I выполняла гармонизацию без утечки данных и групповую импутацию для предотвращения смещения между когортами. Фаза II применяла Сох-регрессию с регуляризацией Elastic Net в качестве диагностического инструмента CANARY для выявления нарушений пропорциональных рисков. Страты с порогом пропущенных данных 35% переходили в Фазу III, где применялся квадрипартитный ансамбль:

случайные леса выживаемости, XGBoost, Survival-Boruta и мультизадачная логистическая регрессия, объединенные в мультимодовой мета-обучающий механизм (MVL) на основе Elastic Net, с последующей интерпретацией через TreeSHAP и LIME.

Результаты

Диагностический инструмент CANARY продемонстрировал структурную несостоятельность пан-раковой геометрической модели пропорциональных рисков. В 96 допустимых стратах Фаза III выполнила алгоритмическое смещение: непрерывные мультиомиксные топологии подавили статические геномные мутации и вариации числа копий (85,7% против 0,0% удержания вершины). MVL стабилизировал прогнозы против экстремальной дисперсии; суррогатные валидации LIME ($R^2 < 0,10$) подтвердили систематическую несостоятельность линейных интерпретационных прокси-моделей. N-мерное картирование взаимодействий TreeSHAP выявило синергетические и антагонистические траектории спасения, определяющие индивидуализированные топологии выживаемости, невидимые для аддитивных моделей. Архитектура была развернута как CancerRCDPredictor — цифровой молекулярный онкологический консилиум со встроенными возможностями больших языковых моделей (LLM). MVL SuperLearner достиг медианного C-индекса 0,749 (IQR: 0,722–0,836) в 96 моделируемых стратах, при этом 95% бутстреп-доверительные интервалы подтвердили точность (медианная ширина: 0,052), а пермутационная значимость наблюдалась в 93,8% страт ($p < 0,001$). Внешняя валидация CPTAC на десяти типах рака продемонстрировала значимую межкогортную обобщаемость для светлоклеточного почечно-клеточного рака (KIRC; C-индекс 0,675, $p = 0,017$) и умеренную производительность для остальных типов рака с достаточной мощностью (медиана 0,582), что подчеркивает необходимость более крупных мультиинституциональных валидационных когорт.

Заключение

Этот пан-раковый мультиомиксный SuperLearner обходит линейные топологические ограничения, продвигаясь за пределы обобщенной стратификации к созданию детерминированно картированной архитектуры для прогнозирования топологий выживаемости, связанных с РКГ. Через интерфейс CancerRCDPredictor мультиомиксные данные трансформируются в индивидуализированное исследование топологий выживаемости, обеспечивая основу для будущей валидации в точной онкологии.

Перевод выполнен: 04.09.2026 | ai4med.ru

Машинный перевод. Рекомендуем сверять с оригиналом при клиническом использовании.