Предписывая дозу прозрачности: качественная оценка объяснений ИИ специалистами в области сердечно-сосудистого здоровья

Источник: Frontiers in AI — Medicine

Оригинал: https://www.frontiersin.org/articles/10.3389/frai.2026.1830201

XAI | доверие к ИИ | кардиология | клиническая практика | объяснимый ИИ | телемедицина

Введение

Хронические заболевания, в частности **сердечная недостаточность**, связаны с высокими показателями смертности и частыми повторными госпитализациями, что создает значительную нагрузку на системы здравоохранения и пациентов. Последние достижения в области телемониторинга с поддержкой **ИИ** (искусственного интеллекта) предлагают способы решения этих проблем посредством раннего обнаружения ухудшения состояния. Тем не менее, природа «черного ящика» ИИ может подорвать доверие врачей, что подчеркивает важность **объяснимого ИИ** (XAI — explainable AI).

Методы

В данном поисковом исследовании изучается восприятие клиницистами XAI в прототипическом сценарии раннего предупреждения о неизбежной декомпенсации сердечной недостаточности на основе табличных лабораторных данных. В ходе интервью с 15 немецкими специалистами в области сердечно-сосудистого здравоохранения с использованием шкалы удовлетворенности объяснениями сравниваются три образцовых метода

объяснения: **SHAP** (SHapley Additive Explanations — аддитивные объяснения Шепли), **CFE** (Counterfactual Explanations — контрфактуальные объяснения) и **Anchors** (Якоря).

Результаты

Метод **SHAP** получил широкое предпочтение благодаря своему интуитивно понятному графическому представлению. **CFE** ценились как метод, ориентированный на действия, хотя их пригодность для выработки терапевтических шагов требует тщательных ограничений и интерпретации. В то же время метод **Anchors** был сочтен слишком поверхностным.

Обсуждение

Наши результаты позволяют предположить, что ни один отдельный метод объяснения не удовлетворяет потребности клиницистов в полной мере. Вместо этого многоуровневые структуры объяснений, сочетающие быстрый обзор с возможностью получения более глубоких данных, могут лучше способствовать доверительному использованию ИИ в сценариях раннего предупреждения на основе табличных лабораторных данных.
