

Прогнозирование риска инфекций репродуктивного тракта среди молодых женщин в Дели с использованием объяснимого искусственного интеллекта

Источник: Frontiers in AI — Medicine

Оригинал: <https://www.frontiersin.org/articles/10.3389/frai.2026.1803913>

диагностика

машинное обучение

объяснимый ИИ

прогнозирование рисков

репродуктивное здоровье

Введение

Инфекции репродуктивных путей (RTI) и инфекции, передающиеся половым путем (STI) представляют собой значительное экономическое бремя и проблему общественного здравоохранения в развивающихся странах, таких как Индия, где неадекватные стратегии раннего выявления и профилактики часто приводят к росту заболеваемости, смертности, стигматизации, развитию рака и неблагоприятным исходам для репродуктивного здоровья как у мужчин, так и у женщин.

Методы

В данном одномоментном (cross-sectional) исследовании использовались модели машинного обучения для прогнозирования риска **RTI/STI** среди молодых женщин в Дели, Индия, путем анализа самозаявленных симптомов наряду с демографическими факторами, показателями репродуктивного здоровья, образом жизни и факторами, связанными с гигиеной. Данные анализировались с использованием нескольких алгоритмов машинного обучения, включая деревья решений (decision trees), случайный лес (random

forest), метод k-ближайших соседей (**KNN**), стохастический градиентный спуск (**SGD**) и **XGBoost**. Перед проведением анализа данные прошли предварительную обработку, включающую несколько этапов, в том числе импутацию пропущенных значений, удаление выбросов с использованием метода межквартильного размаха (**IQR**) и метод синтетического пересэмплирования меньшинства (**SMOTE**) для устранения дисбаланса классов.

Результаты

Сравнительный анализ эффективности моделей показал, что алгоритм **XGBoost** достиг наивысшей точности (89,1%) и наибольшей площади под ROC-кривой (площадь под кривой рабочих характеристик приемника, **ROC-AUC** = 0,935), при этом модель «случайный лес» показала высокие результаты, став второй по эффективности (точность = 87,6%, **ROC-AUC** = 0,921), что продемонстрировало её способность эффективно улавливать нелинейные связи между факторами риска **RTI**. Анализ **SHAP** (SHapley Additive exPlanations — аддитивные объяснения Шепли) определил семейный медицинский анамнез, наличие **RTI** в личном анамнезе и характеристики менструального цикла как наиболее значимые предикторы.

Обсуждение

Полученные результаты подчеркивают потенциал прогностических моделей на основе машинного обучения для стратификации риска **RTI**, поддержки скрининга и выдвижения гипотез.