

Использование речевых биомаркеров для выявления болезни Альцгеймера с помощью языковых моделей на основе BERT: исследование машинного обучения

Источник: Frontiers in AI — Medicine

Оригинал: <https://www.frontiersin.org/articles/10.3389/frai.2026.1875702>

NLP

биомаркеры

диагностика

нейродегенеративные заболевания

Актуальность

Ранняя диагностика имеет решающее значение для эффективного ведения болезни Альцгеймера (БА). Хотя предыдущие исследования показали, что речевые характеристики могут указывать на БА, большинство существующих работ объединяют несколько биомаркеров, что затрудняет выделение вклада отдельных признаков.

Цель

В данном исследовании систематически изолируются отдельные речевые биомаркеры для количественной оценки их различительного вклада в эффективность классификации БА с помощью языковых моделей (ЯМ), а также определяется, улучшает ли целенаправленный отбор биомаркеров результаты по сравнению с объединёнными наборами признаков.

Методы

Мы обработали стенограммы речи из DementiaBank для выявления дискриминационных речевых биомаркеров — вербальных пауз, речевых сбоев и неразборчивых слов. Затем мы дообучили и оценили лёгкие

языковые модели (BERT, ALBERT и DistilBERT) на этих стенограммах, модифицированных с учётом биомаркеров, для автоматической классификации БА.

Результаты

Паузы оказались наиболее дискриминационным речевым биомаркером ($F1 = 0,8326$), значительно превосходя базовый уровень без биомаркеров и другие биомаркеры. Объединение всех биомаркеров в среднем снижало эффективность по сравнению с одними паузами, хотя эффект зависел от модели: условие «все биомаркеры» у BERT превзошло его собственный базовый уровень без биомаркеров, что позволяет предположить, что комбинирование признаков приносит пользу моделям с более высокой ёмкостью. BERT показал наилучшие результаты ($F1 = 0,8154$) во всех условиях.

Выводы

Избирательное использование таких речевых биомаркеров, как паузы, может значимо улучшить выявление БА с помощью лёгких языковых моделей, что позволяет предположить, что целенаправленный отбор биомаркеров может предложить более интерпретируемый и клинически применимый путь, чем широкое объединение признаков.