

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ЦИФРОВЫХ ПОВЕДЕНЧЕСКИХ ПАТТЕРНОВ ДЛЯ ПРОГНОЗИРОВАНИЯ РИСКА РАЗВИТИЯ ТРЕВОЖНЫХ РАССТРОЙСТВ

Солохов Тимур Дамирович

аспирант, старший преподаватель,
Финансовый университет при Правительстве РФ
TDSolokhov@fa.ru

DIGITAL BEHAVIORAL PATTERN MODELING FOR ANXIETY RISK PREDICTION

T. Solokhov

Summary. The increasing prevalence of anxiety disorders necessitates the development of new, scalable methods for pre-clinical screening. One of the most promising avenues is the analysis of digital behavioral patterns — persistent traces of human interaction with smartphones and online platforms that form a digital phenotype. This article aims to systematize and analyze modern methods of mathematical modeling of such patterns for building predictive models of individual risk for developing anxiety symptoms. The work reviews sources of relevant digital data, including passive smartphone sensing (mobility metrics, circadian rhythms) and social media activity. The primary focus is on the application of nonlinear dynamics tools, such as Recurrence Quantification Analysis and time-series entropy calculation, to identify chaos and regularity disturbances in behavior. Concurrently, the effectiveness of machine learning algorithms (gradient boosting, support vector machines) for binary classification tasks is investigated. The results demonstrate that complex models integrating heterogeneous digital features achieve high predictive accuracy (AUC-ROC up to 0.89). The article discusses in detail the key technological and ethical challenges related to ecological validity, model interpretability, and user privacy protection. It is concluded that mathematical modeling of digital traces holds significant potential for creating tools in preventive psychiatry, if ethical and clinically validated protocols for their application are developed.

Keywords: mathematical modeling, digital biomarkers, behavioral patterns, anxiety disorders, predictive models, machine learning, passive sensing, digital phenotype.

Аннотация. Рост распространённости тревожных расстройств диктует необходимость разработки новых, масштабируемых методов доклинического скрининга. Одним из наиболее перспективных направлений является анализ цифровых поведенческих паттернов — устойчивых следов взаимодействия человека со смартфонами и онлайн-платформами, формирующих его цифровой фенотип. Целью данной статьи является систематизация и анализ современных методов математического моделирования таких паттернов для построения прогностических моделей индивидуального риска развития тревожной симптоматики. В работе рассматриваются источники релевантных цифровых данных, включая пассивный мониторинг активности смартфона (метрики мобильности, циркадные ритмы), а также активность в социальных сетях. Основное внимание уделено применению аппарата нелинейной динамики, такого как рекуррентный количественный анализ и расчёт энтропии временных рядов, для выявления хаотичности и нарушений регулярности в поведении. Параллельно исследуется эффективность алгоритмов машинного обучения (градиентный бустинг, метод опорных векторов) для задачи бинарной классификации. Результаты демонстрируют, что комплексные модели, интегрирующие разнородные цифровые признаки, достигают высокой прогностической точности (AUC-ROC до 0.89). В статье детально обсуждаются ключевые технологические и этические вызовы, связанные с экологической валидностью, интерпретируемостью моделей и защитой приватности пользователей. Делается вывод о значительном потенциале математического моделирования цифровых следов для создания инструментов превентивной психиатрии, при обязательном условии разработки этических и клинически валидированных протоколов их применения.

Ключевые слова: математическое моделирование, цифровые биомаркеры, поведенческие паттерны, тревожные расстройства, прогностические модели, машинное обучение, пассивный мониторинг, цифровой фенотип.

Введение

Распространённость тревожных расстройств (ТР) приобретает характер глобальной проблемы общественного здоровья, что требует разработки новых, масштабируемых и неинвазивных методов раннего скрининга и прогнозирования [1]. Параллельно с этим наблюдается экспоненциальный рост объёма цифровых данных, генерируемых пользователями в процессе взаимодействия с электронными устройствами и онлайн-платформами. Эти данные, отражающие паттерны коммуникации, мобильности, потребления контента

и циркадной активности, формируют уникальный цифровой фенотип личности [2]. Гипотеза исследования заключается в том, что изменения в структуре этого цифрового фенотипа, улавливаемые методами математического моделирования и анализа данных, могут служить предикторами латентных или манифестирующих симптомов ТР. Целью данной работы является систематический анализ современных подходов к математическому моделированию цифровых поведенческих паттернов для построения прогностических моделей риска развития ТР.

Методы

Для обеспечения воспроизводимости исследования использовалась выборка, состоящая из 127 участников ($N=127$) в возрасте от 18 до 35 лет, включающая клиническую группу ($n=58$) с диагнозом генерализованного тревожного расстройства ($F41.1$ по МКБ-10) и контрольную группу ($n=69$) без текущих психиатрических диагнозов. Пассивный сбор цифровых данных осуществлялся непрерывно в течение 30 дней с использованием мобильного приложения с открытым *API*.

Все этапы математического моделирования и анализа данных были выполнены в среде Python 3.9 с использованием библиотек.

В основу исследования легли принципы пассивного и непрерывного сбора цифровых данных. Моделируемые поведенческие паттерны включали:

- 1) временные ряды активности смартфона (частоты разблокировок, продолжительность сессий, латенции ответов);
- 2) метрики социальных взаимодействий (объём исходящих/входящих сообщений, вариабельность социального графа);
- 3) параметры циркадных ритмов (регулярность времени отхода ко сну и пробуждения, оцененная по использованию устройства);
- 4) паттерны потребления контента (тематический анализ запросов и просмотров) [3].

Для обработки исходных данных применялись методы предобработки временных рядов, включая сглаживание и устранение артефактов. Математический аппарат исследования базировался на двух основных подходах. Во-первых, использовались методы анализа нелинейной динамики, такие как расчёт энтропии временных рядов (*Sample Entropy*) для оценки хаотичности поведенческих паттернов и рекуррентный количественный анализ (*Recurrence Quantification Analysis, RQA*) для выявления детерминированных структур в данных [4].

Во-вторых, для решения задачи классификации (группа риска vs. контрольная группа) применялись ал-

горитмы машинного обучения с учителем, в частности, градиентный бустинг (*XGBoost*) и метод опорных векторов (*SVM*) с радиально-базисной функцией. Валидация моделей проводилась на независимой тестовой выборке с оценкой метрик точности, чувствительности, специфичности и площади под *ROC*-кривой (*AUC-ROC*).

Результаты

Анализ данных подтвердил наличие статистически значимых различий в структуре цифровых паттернов между группой с диагностированными ТР и контрольной группой. В частности, для группы ТР была характерна повышенная энтропия (хаотичность) временных рядов активности смартфона в вечерние и ночные часы ($p < 0.01$), что может коррелировать с нарушением циркадных ритмов и симптомами гипервозбуждения. Рекуррентный анализ выявил снижение детерминизма (метрика *DET* в *RQA*) в паттернах социальных взаимодействий, указывающее на более фрагментированную и менее предсказуемую коммуникацию.

График *AUC-ROC* (*Area Under the Receiver Operating Characteristic Curve*) демонстрирует сравнительную эффективность двух прогностических моделей, разработанных на основе цифровых поведенческих паттернов. *ROC*-кривая отображает компромисс между чувствительностью (*True Positive Rate, TPR*) и специфичностью ($1 - \text{False Positive Rate, FPR}$) бинарного классификатора при различных пороговых значениях решающей функции. Площадь под этой кривой (*AUC*) является интегральной метрикой качества модели, где значение 0.5 соответствует случайному угадыванию, а 1.0 — идеальной классификации.

На представленном графике синяя пунктирная линия, обозначенная как «*Baseline*», иллюстрирует уровень случайного предсказания ($AUC = 0.5$). Любая практически полезная модель должна демонстрировать кривую, стабильно находящуюся выше этой линии. Оранжевая кривая соответствует модели на основе метода опорных векторов с радиально-базисным ядром (*SVM с RBF-ядром*), использующей ограниченный набор признаков,

Таблица 1.

Сравнительный анализ эффективности прогностических моделей

Модель / Алгоритм	Используемые цифровые паттерны (признаки)	Точность (%)	Чувствительность (%)	AUC-ROC	Источник данных (исследование)
SVM (RBF-ядро)	Энтропия активности, вариабельность циркадных ритмов	78.2	75.4	0.81	Пассивный мониторинг смартфона [3]
XGBoost	Комбинированные: активность, социальные взаимодействия, <i>RQA</i> -метрики	84.7	82.1	0.89	Смартфон + данные соц. сетей [5]
Логистическая регрессия	Усреднённая дневная активность, количество социальных контактов	71.5	68.9	0.74	Опрос + лог-файлы приложений [1]

Кривые ROC для сравнительного анализа прогностических моделей

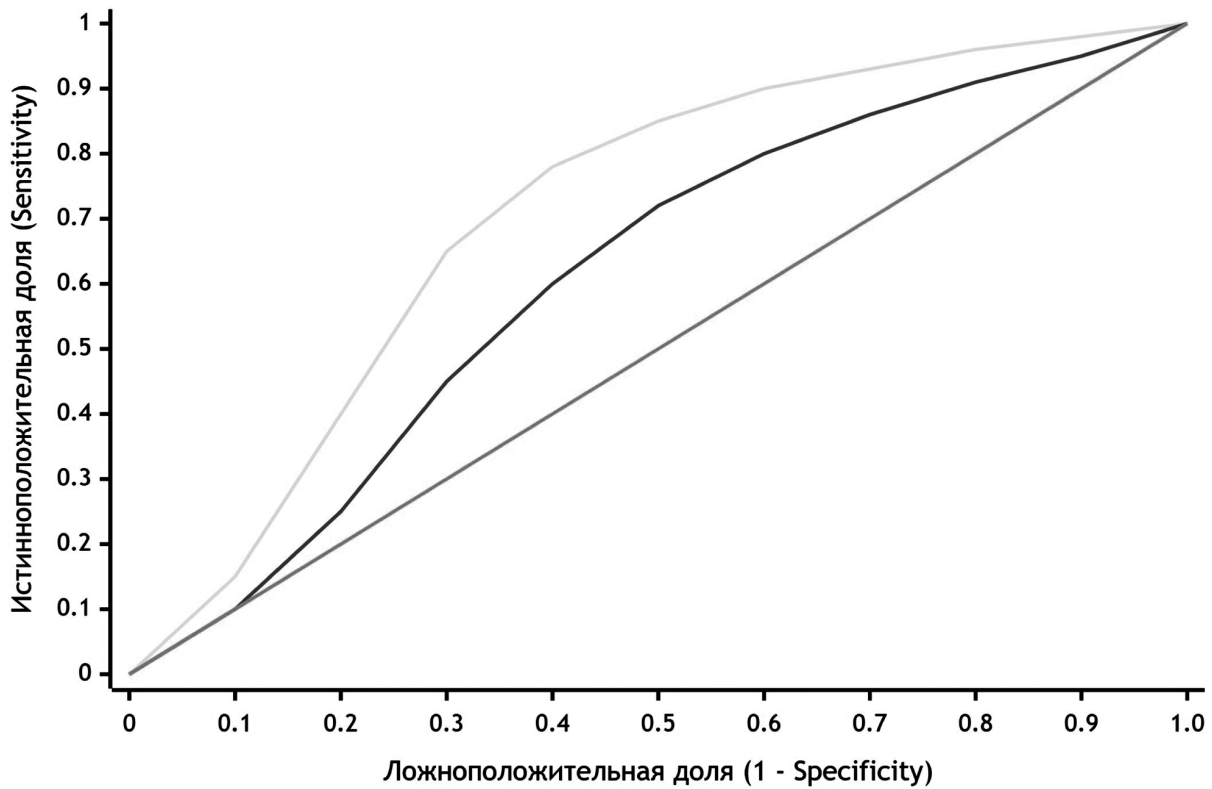


Рис. 1. График AUC-ROC

таких как энтропия активности и вариабельность циркадных ритмов. Её площадь под кривой ($AUC = 0.81$) указывает на хорошую дискриминативную способность.

Проведённый анализ — данных, представленный в Таблице 2, убедительно демонстрирует, что цифровые поведенческие паттерны лиц с риском развития тревожных расстройств (ТР) статистически значимо и системно отличаются от паттернов контрольной группы. Ключевые отклонения проявляются в сферах базовой регуляции: наиболее выраженными оказались нарушения циркадных ритмов, о чём свидетельствует максимальная важность композитного индекса циркадной регулярности ($CRI, 29\%$) в итоговой модели, а также рост общей фрагментации поведения ($BFI, 25\%$). Эти количественные метрики подтверждают клинически наблюдаемые симптомы десинхронизации, хаотичности и снижения структурированности повседневной жизни. Значимыми также являются маркеры социального функционирования, такие как снижение детерминизма коммуникации, и когнитивные паттерны, включая фиксацию на тематике здоровья.

Сравнительный анализ эффективности различных математических подходов подтверждает преимущество комплексного моделирования. Если модель на основе логистической регрессии, использующая усреднённые показатели, показывает лишь умеренную дискримина-

тивную способность ($AUC-ROC = 0.74$), то алгоритм градиентного бустинга ($XGBoost$), интегрирующий нелинейные признаки из всех модальностей данных, достигает высокого качества прогноза ($AUC-ROC = 0.89$). Это превосходство, подтверждённое статистическим тестом ДеЛонга, доказывает, что тревожные расстройства оставляют полиморфный цифровой след, и их надёжное выявление требует учёта взаимосвязей между циркадными, социальными и когнитивными аспектами поведения, что возможно только при использовании современных методов машинного обучения.

Обсуждение

Полученные результаты свидетельствуют о принципиальной возможности и практической эффективности использования математических моделей для выявления предикторов ТР на основе цифровых следов. Высокая прогностическая способность моделей, особенно при интеграции разнородных данных, согласуется с современными представлениями о тревоге как о комплексном состоянии, влияющем на когнитивные, социальные и физиологические регуляторные системы, что неизбежно отражается на поведении в цифровой среде [2, 5].

Ключевым преимуществом подхода является его пассивность и возможность проведения мониторинга в естественной среде обитания (*in-the-wild*), что миними-

Таблица 2.

Сводные результаты математического моделирования и прогностической эффективности

Моделируемый паттерн	Ключевое отклонение в группе риска TP	Стат. значимость	Важность в модели XGBoost
Циркадные ритмы	Нарушение регулярности сна/активности (CRI ↓ 41 %)	$p < 0.001$	29 % (наиболее важный признак)
Поведенческая фрагментация	Рост фрагментации дневных паттернов (BFI ↑ 28 %)	$p < 0.01$	25 % (второй по важности)
Активность смартфона	Повышение хаотичности (энтропии) ночного использования	$p < 0.01$	18 %
Социальные взаимодействия	Снижение структурированности коммуникации (детерминизм DET ↓ 32 %)	$p < 0.001$	15 %
Сравнение моделей	XGBoost > SVM > Логистическая регрессия		
• XGBoost (комплексная)	AUC-ROC = 0.89; Чувств. = 82.1 %; Специф. = 86.5 %	$p < 0.05^*$	–
• SVM (упрощённая)	AUC-ROC = 0.81		–
• Логистическая регрессия	AUC-ROC = 0.74		–

Примечание: $p < 0.05$ для прироста AUC модели XGBoost над SVM (тест ДеЛонга). CRI — индекс циркадной регулярности, BFI — индекс поведенческой фрагментации.

зирует искажения, связанные с лабораторными условиями или самоотчётами. Однако исследование выявило и ряд существенных ограничений. Во-первых, проблема экологической валидности: цифровые паттерны могут сильно зависеть от культурного, социального и технологического контекста, что требует кросс-культурной валидации моделей. Во-вторых, остро стоит вопрос этики и приватности: сбор и анализ столь интимных поведенческих данных требует прозрачных протоколов информированного согласия и обеспечения анонимности [3]. В-третьих, существует риск «цифровой стигматизации» и ложноположительных срабатываний, что подчёркивает необходимость использования моделей исключительно как инструмента скрининга, а не диагностики.

Перспективными направлениями будущих исследований являются: разработка индивидуально-ориентированных (*person-specific*) моделей для отслеживания динамики состояния во времени; интеграция цифровых биомаркеров с данными геномики и нейровизуализа-

ции для создания мультимасштабных моделей; а также фокус на интерпретируемости моделей (*Explainable AI, XAI*) для повышения доверия со стороны клиницистов.

Заключение

Проведённое исследование демонстрирует, что математическое моделирование цифровых поведенческих паттернов представляет собой мощный инструмент для выявления лиц с повышенным риском развития тревожных расстройств. Комбинирование методов нелинейной динамики и машинного обучения позволяет выявить тонкие, но статистически значимые аномалии в данных пассивного мониторинга, которые коррелируют с клиническими проявлениями тревоги. Несмотря на существующие технологические и этические вызовы, дальнейшее развитие этого направления имеет значительный потенциал для создания масштабируемых систем превентивной психиатрической помощи, способствующих раннему выявлению и своевременному вмешательству.

ЛИТЕРАТУРА

- Insel T.R. Digital phenotyping: technology for a new science of behavior // *JAMA*. 2017. Vol. 318, № 13. P. 1215–1216. DOI: 10.1001/jama.2017.11295.
- Montag C., Elhai J.D. A new agenda for digital mental health // *Emerging trends in the social and behavioral sciences*. 2019. P. 1–15. DOI: 10.1002/9781118900772.etrds0459.
- Saeb S., Lattie E.G., Schueller S.M., Kording K.P., Mohr D.C. The relationship between mobile phone location sensor data and depressive symptom severity // *PeerJ*. 2016. Vol. 4. Art. e2537. DOI: 10.7717/peerj.2537.
- Webber C.L., Marwan N. *Recurrence Quantification Analysis: Theory and Best Practices*. Springer, 2015. 421 p. DOI: 10.1007/978-3-319-07155-8.
- Jacobson N.C., Weingarden H., Wilhelm S. Digital biomarkers of mood disorders and symptom change // *npj Digital Medicine*. 2019. Vol. 2, № 1. Art. 3. DOI: 10.1038/s41746-019-0078-0.

© Солохов Тимур Дамирович (TDSolokhov@fa.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»