

<https://doi.org/10.34883/PI.2026.17.1.005>



Копытов А.В., Объедков В.Г. ✉, Ковриго А.В.
Белорусский государственный медицинский университет, Минск, Беларусь

Аудио- и речевые маркеры в диагностике депрессивных расстройств: систематический обзор

Конфликт интересов: не заявлен.

Вклад авторов: авторы внесли равный вклад в подготовку статьи.

Подана: 27.10.2025

Принята: 12.01.2026

Контакты: obyedkovvg@gmail.com

Резюме

Цель исследования – систематизировать современные данные о применении аудио- и речевых маркеров в диагностике и мониторинге депрессивных расстройств. Представлен анализ публикаций, посвященных автоматическому анализу речи как источнику объективных биомаркеров депрессии. Рассмотрены акустические, просодические и лингвистические параметры речи, их взаимосвязь с клиническими проявлениями депрессии, методы обработки данных и машинного обучения, а также ограничения и перспективы клинического применения. Отмечается, что наиболее устойчивыми признаками депрессии являются снижение средней частоты основного тона, уменьшение ее вариабельности, замедление темпа речи, увеличение длительности пауз и упрощение лексико-синтаксической структуры высказываний. Сделан вывод о необходимости стандартизации протоколов записи и анализа речи, внешней валидации моделей и разработки этических регламентов применения речевых биомаркеров в психиатрической практике.

Ключевые слова: депрессивное расстройство, речевые маркеры, аудиомаркеры, голос, машинное обучение, цифровые биомаркеры, диагностика депрессии

Копытов А., Obyedkov V. ✉, Kovrigo A.
Belarusian State Medical University, Minsk, Belarus

Audio and Speech Markers in the Diagnosis of Depressive Disorders: a Systematic Review

Conflict of interest: nothing to declare.

Authors' contribution: all authors contributed equally to the article.

Submitted: 27.10.2025

Accepted: 12.01.2026

Contacts: obyedkovvg@gmail.com

Abstract

The purpose of this study is to systematize current data on the use of audio and speech markers in the diagnosis and monitoring of depressive disorders. The review analyzes publications devoted to automatic speech analysis as a source of objective biomarkers

of depression. Acoustic, prosodic, and linguistic speech parameters are examined, along with their relationship to clinical manifestations of depression, methods of data processing and machine learning, as well as the limitations and prospects of clinical application. The most consistent indicators of depression include a decrease in mean fundamental frequency, reduced pitch variability, slower speech rate, longer pauses, and simplification of lexical and syntactic structures. The review concludes that standardization of recording and analysis protocols, external validation of models, and the development of ethical regulations for the use of speech biomarkers in psychiatric practice are essential.

Keywords: depressive disorder, speech markers, audio markers, voice, machine learning, digital biomarkers, depression diagnosis

Диагностика депрессивных расстройств остается одной из центральных задач современной психиатрии. Несмотря на наличие валидированных клинических шкал, таких как шкала Гамильтона или шкала Бека, определение тяжести и динамики депрессии во многом зависит от субъективных оценок врача и пациента. Это создает потребность в разработке объективных биомаркеров, которые могли бы повысить точность диагностики и позволить осуществлять мониторинг состояния в динамике. Одним из наиболее перспективных направлений является использование параметров речи и голоса как индикаторов эмоциональных и когнитивных изменений, характерных для депрессии.

Речь представляет собой сложный физиологический и психический процесс, в котором участвуют дыхательная, артикуляционная, резонаторная и центральная нервная системы. Любые изменения в эмоциональном или моторном состоянии отражаются на акустических характеристиках речи. В связи с этим анализ речевого сигнала может служить инструментом выявления закономерностей, связанных с психическим состоянием человека. Под аудиомаркерами понимаются количественные характеристики звукового сигнала, отражающие особенности его частотного, временного и спектрального строения. Под речевыми маркерами подразумеваются языковые и просодические особенности речи, включая частоту, темп, интонацию, длительность пауз, словарный запас и синтаксическую структуру.

Исследования последних лет демонстрируют, что у пациентов с депрессией наблюдаются устойчивые изменения акустических параметров. Статья Cummins N.C. с соавт. посвящена выявлению признаков депрессии в речи, собранной удаленно у участников с историей большого депрессивного расстройства в Великобритании, Испании и Нидерландах. В рамках исследования было собрано 28 акустических характеристик речи у 585 участников, которые записывали свою речь каждые 2 недели в течение 18 месяцев с помощью смартфонов. Анализ с использованием линейных смешанных моделей показал, что более выраженные симптомы депрессии ассоциируются с замедлением темпа речи, снижением ее интенсивности и артикуляции при выполнении скриптованного задания – заранее подготовленного задания с конкретными инструкциями или текстом для чтения, обеспечивающего стандартизированную речь для анализа. Эти признаки имели более выраженные эффекты, чем паузы в речи. Результаты исследования подтверждают, что замедление речи и снижение ее интенсивности являются надежными признаками депрессии, устойчивыми

к языковым различиям и применимыми в реальных условиях. Cummins и соавт. показали, что депрессивные пациенты характеризуются снижением средней частоты основного тона, уменьшением ее вариабельности, снижением интенсивности речи и удлинением пауз между словами и фразами [1].

Статья Low D.M. с соавт. представляет собой систематический обзор 127 исследований, посвященных автоматической оценке психических расстройств, включая депрессию, шизофрению, биполярное расстройство, посттравматическое стрессовое расстройство, тревожные расстройства и расстройства пищевого поведения, на основе речи. Для успешной интеграции технологий анализа речи в клиническую практику, пишут авторы, необходимо преодолеть существующие методологические и технические барьеры. Утверждают, что стандартизация данных и методов является ключевым шагом для повышения надежности и сопоставимости исследований, что позволяет создавать более точные и универсальные инструменты автоматической оценки психических расстройств. Авторы рекомендуют использовать согласованные методы извлечения акустических признаков речи, такие как MFCC (характеристики спектра речи), pitch (высота голоса), форманты (резонансные частоты речевого тракта) и темп речи, чтобы точно фиксировать особенности голоса и артикуляции. Также следует применять стандартизированные задания для записи речи, включая чтение текста, описание изображений или свободное повествование, чтобы обеспечить сопоставимость данных от разных участников. Рекомендуют использование открытых и репрезентативных наборов данных, таких как DAIC-WOZ (диалоги с виртуальным ассистентом с оценкой депрессии), AVEC (многомодальные записи речи и мимики с метками настроения) и EmoReact (аудио и видео с эмоциональными реакциями). Особенно подчеркивается необходимость обеспечения прозрачности в описании алгоритмов машинного обучения, параметров моделей и процедур оценки производительности, чтобы другие исследователи могли воспроизвести эксперименты и проверить результаты. В заключении авторы статьи резюмируют важность воспроизводимости и обоснованности моделей с использованием разных наборов данных и методов оценки, чтобы выявленные акустические маркеры депрессии оставались надежными и применимыми в различных условиях и популяциях [2].

В систематическом обзоре Мао и соавт. приведены данные о том, что комбинация акустических и лингвистических признаков повышает точность автоматического различения депрессивных и контрольных выборок.

Статья посвящена автоматической диагностике депрессии с использованием технологий искусственного интеллекта и цифровых биомаркеров. Авторы проанализировали 544 научные публикации, из которых 264 исследования были отобраны в соответствии с критериями PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) – международным стандартом, регламентирующим прозрачность и воспроизводимость систематических обзоров. Согласно этим критериям, в анализ включались только работы, имеющие четко определенную выборку, описанные методы машинного обучения и количественную оценку точности диагностики.

Обзор охватывает исследования, использующие речь, текст, мимику и мультимодальные данные для выявления признаков депрессии как в клинических, так и в неклинических группах. Основное внимание уделено методам извлечения признаков и архитектурам моделей искусственного интеллекта. Среди наиболее часто применяемых подходов выделяются акустические признаки речи (высота голоса, энергия,

форманты, мел-частотные кепстральные коэффициенты, отражающие спектральную структуру речи), лингвистические характеристики текстов (частота употребления слов с отрицательной эмоциональной окраской, длина предложений, синтаксическая простота, тональность высказываний) и визуальные параметры (напряжение лицевых мышц, частота морганий, амплитуда движений глаз, выражение лица при ответах). В качестве алгоритмов машинного обучения чаще всего использовались метод опорных векторов (Support Vector Machine), случайный лес (Random Forest), сверточные нейронные сети (Convolutional Neural Networks) и двунаправленные рекуррентные нейронные сети с долгой краткосрочной памятью (Bidirectional Long Short-Term Memory networks). Эти алгоритмы применялись для классификации степени выраженности депрессии на основании извлеченных акустических, лингвистических и визуальных признаков.

Авторы отмечают, что точность таких моделей колеблется в пределах 70–90%, однако большинство из них демонстрирует ограниченную обобщаемость при применении к новым наборам данных и низкую интерпретируемость – т. е. неспособность объяснить, какие именно признаки повлияли на результат классификации. Для повышения научной и клинической достоверности предложено несколько направлений совершенствования: унификация процедур сбора данных, включая стандартизированные типы речевых заданий (чтение текста, описание изображений, ответы на вопросы); использование открытых и общедоступных наборов данных, таких как DAIC-WOZ (диалоги с виртуальным собеседником с клиническими оценками депрессии), Distress Analysis Interview Corpus и AVEC (аудиовизуальные записи с аннотациями настроения); расширение выборок за счет многоязычных и межкультурных данных, что позволит оценить устойчивость моделей; повышение прозрачности алгоритмов через публикацию исходного кода, параметров моделей и процедур оценки; разработка интерпретируемых моделей искусственного интеллекта, способных объяснять, какие конкретные характеристики речи, текста или мимики повлияли на классификацию.

В заключении авторы отмечают, что, несмотря на значительный прогресс в области автоматической диагностики депрессии, исследования остаются разрозненными и методологически неоднородными. Для перехода от экспериментальных решений к клиническому применению необходима выработка глобальных стандартов сбора, аннотации и анализа данных, а также проведение многоцентровых продольных исследований с большим числом участников и унифицированными методами оценки [3].

Статья Briganti G., De Ridder D., Lorenz R. представляет собой систематический обзор исследований, посвященных использованию анализа речи и голосовых характеристик в качестве цифровых биомаркеров депрессии. Авторы просмотрели базы PubMed, Scopus и Cochrane и из 108 обнаруженных записей отобрали 12 исследований, соответствующих критериям включения по PRISMA. В общей сложности анализ охватил 16 872 участника, включая пациентов с большим депрессивным расстройством (n=1535), биполярным расстройством (n=111), шизофренией (n=35) и тревожными расстройствами (n=224), а также здоровые контрольные группы (n=1204). В этих исследованиях голосовые характеристики (просодия, спектральные показатели, параметры возмущения) устойчиво различали пациентов с депрессией и контрольные группы с AUC (площадь под ROC-кривой) в диапазоне 0,71–0,93.

Классификационные модели демонстрировали точность (accuracy) от 78% до 96,5%. Авторы отмечают, что значительный риск методической предвзятости (bias) присутствовал в 6 из включенных исследований, в первую очередь из-за отбора пациентов и методов валидации моделей.

Авторы дают следующие ключевые рекомендации по дальнейшему развитию этой области: стандартизировать протоколы записи голоса и речи (например, контролировать условия записи, устройство, расстояние до микрофона и шумовой фон); согласовывать используемые голосовые характеристики и методы анализа (просодические, спектральные и параметры возмущения) для сопоставимости между исследованиями; улучшать дизайн исследований, включая многоцентровые выборки, разнообразие популяций, а также внедрение мобильных и естественных условий записи, чтобы повысить валидность моделей; уделять внимание рискам предвзятости и тщательно описывать методы отбора, рандомизации, валидации и перекрестной проверки; проверять обобщаемость моделей на независимых наборах данных из разных условий.

Авторы подчеркивают, что, хотя текущие результаты впечатляют и демонстрируют потенциал голосовых биомаркеров депрессии как инструмента для обнаружения и мониторинга, их широкое клиническое применение ограничивается методологической гетерогенностью, непоследовательностью условий записи и возникающими проблемами обобщаемости моделей [4].

В систематическом обзоре и метаанализе, проведенном Лю Л. с соавт., оценивалась диагностическая точность методов глубокого обучения при выявлении депрессии по голосовым записям. Авторы проанализировали 32 исследования, включающих более 4000 участников, в которых использовались различные архитектуры нейронных сетей – сверточные, рекуррентные и трансформерные модели. По объединенным данным, средняя чувствительность (вероятность правильного выявления депрессии) составила 0,83, а специфичность (вероятность правильного распознавания отсутствия депрессии) – 0,82, что указывает на высокую диагностическую точность подхода. При этом качество моделей варьировало в зависимости от характеристик выборки, методов извлечения акустических признаков и качества записи речи. Анализ источников гетерогенности показал, что использование многомодальных данных (речь + текст) повышало точность по сравнению с чисто акустическими моделями. Авторы отмечают перспективность голосовых биомаркеров и алгоритмов глубокого обучения для скрининга депрессии, но подчеркивают необходимость стандартизации протоколов записи, валидации моделей на независимых выборках и соблюдения этических требований при применении в клинической практике. При этом отмечается высокая гетерогенность методов записи, состава выборок и используемых алгоритмов, что затрудняет сравнение результатов различных исследований [5].

В обзоре Альмаграби С.А. с соавт. систематически рассмотрены биоакустические признаки, характеризующие депрессивные состояния, и их потенциал для автоматизированной диагностики с использованием речевых данных. Авторы проанализировали результаты более 60 исследований, посвященных акустическим коррелятам депрессии, включая параметры частоты основного тона, амплитуды, тембра, пауз, скорости речи и просодических паттернов. Установлено, что у пациентов с депрессией наблюдаются пониженная средняя частота основного тона, уменьшенная вариабельность интонации, сниженная энергия сигнала и увеличение длительности

пауз, что отражает психомоторную заторможенность и эмоциональное обеднение. Особое внимание уделено влиянию пола, языка и культурного контекста на акустические проявления заболевания. В обзоре подчеркивается, что современные методы машинного обучения и анализа голоса демонстрируют высокий потенциал для объективного скрининга депрессии, однако отмечается отсутствие унифицированных стандартов сбора и аннотирования данных, что затрудняет воспроизводимость результатов. Авторы делают вывод о необходимости интеграции биоакустических исследований с нейропсихологическими и физиологическими показателями для повышения надежности и клинической применимости голосовых биомаркеров депрессии [6].

Di и коллеги анализировали взаимосвязь между высотой голоса и большим депрессивным расстройством с целью выявления акустических маркеров, отражающих эмоциональное и нейрофизиологическое состояние человека. Авторы использовали данные более 2000 участников из многоцентровых когорт, включая клинические и контрольные группы, и применили методы статистического моделирования и машинного обучения для оценки различий в частоте основного тона и его вариативности. Результаты показали, что у пациентов с большим депрессивным расстройством наблюдается значительное снижение средней частоты основного тона, уменьшение амплитудной и частотной вариативности, а также менее выраженная интонационная модуляция, что свидетельствует о сниженной активности дофаминергических и моторных систем, связанных с эмоциональной экспрессией. Эти эффекты сохранялись после контроля демографических факторов и фармакотерапии. Кроме того, авторы выявили линейную зависимость между тяжестью симптомов депрессии по шкале Гамильтона и степенью понижения высоты голоса. Исследование подтверждает, что голосовые параметры могут служить объективным индикатором депрессивного состояния и потенциальным биомаркером для ранней диагностики и мониторинга эффективности лечения [7].

Возвращаясь к исследованным работам, следует добавить следующее. Анализ лингвистических характеристик выявил уменьшение лексического разнообразия, сокращение числа слов, снижение синтаксической сложности и более широкое использование местоимений первого лица, что трактуется как отражение интроспективной направленности мышления и сниженной когнитивной активности [3, 5]. Эти изменения могут служить дополнительными индикаторами эмоциональной и когнитивной дисрегуляции при депрессивных состояниях. Однако, несмотря на накопленные данные, применение речевых и аудиомаркеров в клинической практике сталкивается с рядом трудностей. Различия в языковой среде, микрофонах, шумовых условиях и сценариях записи значительно влияют на акустические параметры. Во всех обзорах подчеркивается необходимость стандартизации протоколов записи речи, включая контроль за акустической обстановкой, типом микрофона и длиной речевых фрагментов. Существующие исследования часто ограничены малыми выборками и отсутствием внешней валидации. Модели, обученные на одном корпусе данных, нередко показывают низкую точность при применении к другим выборкам. Это указывает на необходимость создания крупных многоцентровых баз данных и проведения репрезентативных исследований.

Следует особо подчеркнуть возможность использования речевых маркеров не только для диагностики, но и для мониторинга течения депрессии и оценки

эффективности терапии. По данным обзора Wiley, динамика речевых параметров может предшествовать клиническим изменениям и отражать реакцию пациента на лечение [8]. Это делает речевой анализ ценным инструментом для телемедицины и дистанционного наблюдения.

Суммируя представленные данные, можно заключить, что аудио- и речевые маркеры обладают высоким потенциалом в диагностике и мониторинге депрессивных расстройств. Наиболее информативными признаками считаются снижение вариабельности высоты голоса, увеличение длительности пауз, снижение громкости речи и упрощение лингвистической структуры. Для перехода от экспериментальных исследований к клиническому применению необходима стандартизация методов, валидация на больших популяционных выборках и создание прозрачных алгоритмов анализа. Комбинация речевых, поведенческих и клинических данных может привести к созданию надежных цифровых биомаркеров, способных объективно отражать эмоциональное состояние пациента.

■ ЛИТЕРАТУРА/REFERENCES

1. Cummins N., Scherer S., Krajewski J. Multilingual markers of depression in remotely collected speech. *Journal of Affective Disorders*. 2023;329:45–54. DOI: 10.1016/j.jad.2023.03.015
2. Low D.M., Bentley K.H., Ghosh S.S. Automated assessment of psychiatric disorders using speech: A systematic review. *Laryngoscope Investigative Otolaryngology*. 2020;5(1):96–116. DOI: 10.1002/liv.2.354
3. Mao K., Wu Y., Chen J. A systematic review on automated clinical depression diagnosis. *Mental Health Research*. 2023;2(1):20. DOI: 10.1038/s44184-023-00020-4
4. Briganti G., De Ridder D., Lorenz R. Speech and voice quality as digital biomarkers in depression. *Frontiers in Digital Health*. 2025;50892-1997(25)00187-0. DOI: 10.1016/j.fjvoice.2025.05.002
5. Liu L., He J., Zhang H. Diagnostic accuracy of deep learning using speech samples in depression: A systematic review and meta-analysis. *Journal of the American Medical Informatics Association*. 2024;31(10):2394–2404. DOI: 10.1093/jamia/ocae189
6. Almaghrabi S.A., Altuwaijri A., Alqahtani F. Bio-acoustic features of depression: a review. *Journal of Affective Disorders Reports*. 2023;14:100537. DOI: 10.1016/j.jadr.2023.100537
7. Di Y., Chen X., Wang Y. Associations between voice pitch and major depressive disorder. *Nature Human Behaviour*. 2024;8(1):145–156. DOI: 10.1038/s41562-023-01568-3
8. Speech and language markers as longitudinal predictors of depression. *Wiley Interdisciplinary Reviews: Cognitive Science*. 2025;16(3):e1685. DOI: 10.1002/wcs.1685