

Оценка психического состояния говорящего по параметрам устной речи***И.А. Филиппенков¹, В.А. Багрянов¹, В.Я. Чучупал²***¹Московский физико-технический институт (национальный исследовательский университет)²Федеральный исследовательский центр «Информатика и управление»

Раннее выявление депрессивных расстройств представляет значительный научный и практический интерес для современной медицины. В последние годы особое внимание исследователей привлекают методы автоматического анализа речевых характеристик как объективных маркеров аффективных нарушений. Существующие работы, такие как [1,2], демонстрируют высокую эффективность бинарной классификации депрессии по речевым признакам, достигая точности 83-89% при анализе спонтанной речи.

В данной работе сравнивается два метода классификации депрессии по аудиозаписям: 1) использование свёрточно-рекуррентной модели на спектрограммах речи и 2) дообучения предобученной модели wav2vec 2.0.

При первом методе аудиосигнал преобразуется в мел-спектрограмму, которая подаётся на модель, состоящую из сверточных слоев (CNN) и рекуррентных блоков (GRU). Преимуществом этого метода является учет как локальные акустические признаки, так и долговременные зависимости в речи. Обучение проводится с нуля на собранном и размеченном специально для этой задачи датасете с сбалансированными классами.

Второй метод основан на архитектуре wav2vec 2.0, изначально тренированной на обширном корпусе неразмеченных речевых данных. Для решения задачи реализован процесс трансферного обучения: к предобученному базовому модулю добавляется специализированный классификационный слой, который затем адаптируется на размеченных аудиозаписях пациентов. Ключевое преимущество данного подхода заключается в автоматизированном извлечении наиболее релевантных акустических и лингвистических паттернов непосредственно из сырого аудиосигнала.

Данные, предоставленные профессиональными психологами, разделены, как и в других исследованиях [3], на речь с чтением текста и “спонтанную” речь – описание картинки. Был проведен сравнительный анализ работы моделей на 2 типах речи с использованием метрик precision, recall и accuracy. Характерной проблемой для исследователей в данной области является нехватка данных. Для повышения эффективности работы модели была произведена предобработка данных и их аугментация, которая также помогла выявить значимые речевые характеристики.

Литература

1. Lu, X., Shi, D., Liu, Y., & Yuan, J. (2021). Speech depression recognition based on attentional residual network. *Frontiers in Bioscience (Landmark Edition)*, 26(12), 1746–1759.
2. Chen, X., & Pan, Z. (2021). A convenient and low-cost model of depression screening and early warning based on voice data using for public mental health. *International Journal of Environmental Research and Public Health*, 18(12), 6441.
3. Zhang, L., Duvvuri, R., Chandra, K. K. L., Nguyen, T., & Ghomi, R. H. (2020). Automated voice biomarkers for depression symptoms using an online cross-sectional data collection initiative. *Depression and Anxiety*, 37(7), 657–669.