Applied Psychophysiology and Biofeedback (2024) 49:71–83
https://doi.org/10.1007/s10484-023-09612-3

Effectiveness of a Biofeedback Intervention Targeting Mental and Physical Health Among College Students Through Speech and Physiology as Biomarkers Using Machine Learning: A Randomized Controlled Trial

Lifei Wang[1,2] · Rongxun Liu[1,2,3] · Yang Wang[1,2,4] · Xiao Xu[5] · Ran Zhang[1,2] · Yange Wei[1,2] · Rongxin Zhu[1,2] · Xizhe Zhang[5] · Fei Wang[1,2,6]

Accepted: 24 November 2023 / Published online: 2 January 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract
Biofeedback therapy is mainly based on the analysis of physiological features to improve an individual's affective state. There are insufficient objective indicators to assess symptom improvement after biofeedback. In addition to psychological and physiological features, speech features can precisely convey information about emotions. The use of speech features can improve the objectivity of psychiatric assessments. Therefore, biofeedback based on subjective symptom scales, objective speech, and physiological features to evaluate efficacy provides a new approach for early screening and treatment of emotional problems in college students. A 4-week, randomized, controlled, parallel biofeedback therapy study was conducted with college students with symptoms of anxiety or depression. Speech samples, physiological samples, and clinical symptoms were collected at baseline and at the end of treatment, and the extracted speech features and physiological features were used for between-group comparisons and correlation analyses between the biofeedback and wait-list groups. Based on the speech features with differences between the biofeedback intervention and wait-list groups, an artificial neural network was used to predict the therapeutic effect and response after biofeedback therapy. Through biofeedback therapy, improvements in depression ($p = 0.001$), anxiety ($p = 0.001$), insomnia ($p = 0.013$), and stress ($p = 0.004$) severity were observed in college-going students (n = 52). The speech and physiological features in the biofeedback group also changed significantly compared to the waitlist group (n = 52) and were related to the change in symptoms. The energy parameters and Mel-Frequency Cepstral Coefficients (MFCC) of speech features can predict whether biofeedback intervention effectively improves anxiety and insomnia symptoms and treatment response. The accuracy of the classification model built using the artificial neural network (ANN) for treatment response and non-response was approximately 60%. The results of this study provide valuable information about biofeedback in improving the mental health of college-going students. The study identified speech features, such as the energy parameters, and MFCC as more accurate and objective indicators for tracking biofeedback therapy response and predicting efficacy. Trial Registration ClinicalTrials.gov ChiCTR2100045542.

Keywords College students · Biofeedback · Speech acoustic features · Formant · Machine learning

Introduction

The period of college, which represents an important stage of development straddling the adolescent and young adulthood life stages, is the peak for the onset of many common psychiatric illnesses, particularly depression, anxiety (Auerbach et al., 2018; Freeman et al., 2017; de Girolamo et al., 2012; Kessler et al., 2007; Sheaves et al., 2016)and insomnia (Freeman et al., 2017; Sheaves et al., 2016). However, while timely and effective treatment is important, a number of students in need of treatment for these disorders receive no pharmaceutical or traditional psychological treatment from college mental health care services, resulting in a substantial unmet need for the treatment of mental disorders among college students (Auerbach et al., 2016, 2018; Beiter et al., 2015). The majority of current interventions

Lifei Wang, Rongxun Liu and Yang Wang have contributed equally to this work.

Extended author information available on the last page of the article

for psychiatric illnesses mainly focus on psychotherapy and medication, and therapy that emphasizes the mind-body connection is seemingly undervalued (Weerdmeester et al., 2020). The use of biofeedback has facilitated access to many therapeutic solutions for mental health, aiming to improve the affective state of the individual (Alneyadi et al., 2021; Maynart et al., 2021; Schoenberg & David, 2014), which could be useful for several psychiatric illnesses such as stress, insomnia, anxiety and depression.

Biofeedback technology enables people to change their physiological activity through measurements of respiratory rate, heart rate (Gaggioli et al., 2014), galvanic skin response (also known as electrodermal response) (Kotwas et al., 2017; Krusemark & Wen, 2012) and others, and provides feedback. Projected information can cause changes in the emotions, thoughts, and behaviors of people (Maynart et al., 2021). In 2016, the Association for Applied Psychophysiology and Biofeedback classified the use of biofeedback techniques as an evidence level of four, effective for depression (Maynart et al., 2021).

Historically, clinical assessment and treatment in psychiatry have relied on patient self-reports, clinician judgment, and subjective clinician rating scales, which are vulnerable to social desirability and other subjective biases, lacking objective and clinically relevant outcome measures (Little et al., 2021; Salekin et al., 2018; Sverdlov et al., 2021). Therefore, introducing innovative and easily accessible therapies that can increase the efficacy of treatment, solving the limitations of conventional methods, and introducing precise and objective indicators for accurate assessment have become a top priority in solving the problem (Maynart et al., 2021).

Of particular concern is that in patients with major depression, movements are slowed in addition to impairments in physiological indicators that impact speech (Bernard & Mittal, 2015). Vocal acoustic features can discriminate between mood changes (Lee et al., 2021). Therefore, the use of speech information, which is noninvasive and easily accessible, for initial judgment provides a new method for the early screening of depression, to reduce the cost of depression detection to some extent (Wang et al., 2019). Previous studies have shown that patients with depression are characterized by slow speaking speed, low intonation, weak voice intensity (Kraepelin, 1921), reduced changes in speech features (Cannizzaro et al., 2004), and more pauses (Mundt et al., 2012). At the same time, changes in voice bandwidth, amplitude, energy, and other changes in patients with depression are reduced (Kuny & Stassen, 1993; Mundt et al., 2012), and the spectral characteristics of patients are also related to the degree of depression (Tolkmitt et al., 1982). In patients with anxiety, many studies found a significant increase in mean fundamental frequency (f0) (König et al., 2021; Low et al., 2020). Jitter and shimmer are also significantly higher in patients with anxiety (König et al., 2021; Oezseven et al., 2018; Silber-Varod et al., 2016). Furthermore, pitch, tone, rate of speech, and valence of language are quantifiable based on natural language models, index motor function, mood, and cognitive functioning (Cannizzaro et al., 2004; Schultebraucks et al., 2020). Speech could, therefore, be a key component in developing an accurate biomarker for depression, and there has been recent interest in automatically analyzing depressed speech (Little et al., 2021).

With the development of computer, the use of deep learning algorithms has proven promising in the field of psychiatry, and deep learning algorithms are considered one of the most promising machine learning techniques. Deep learning can efficiently exploit neuroimaging data and integrate non-imaging biomarkers (Squarcina et al., 2021), thereby bridging the gap between empirical findings and explanatory theories of psychology and cognitive neuroscience (Hasson et al., 2020; Schultebraucks et al., 2020). This allows the integration of multiple empirical associations, including subtle ones, into a computational framework. Several popular deep learning model architectures include deep feedforward neural networks (DFNN), recurrent neural networks (RNN) (Bengio et al., 1994), convolutional neural networks (CNN) (Yann et al. n.d.), and autoencoders (Vincent et al., 2010). Various machine learning methods using acoustic features have been used in the diagnosis of MDD (Di et al., 2021), and the aim of applying deep learning is not to accurately model such theories but, more modestly, to find stable probabilistic patterns in a data-driven manner (Valiant, 1984; Schultebraucks et al., 2020). Visualization methods have been proposed that may help in the interpretability of results by highlighting regions driving the deep learning decision using sensitivity and saliency analysis (Zeiler & Fergus, 2013; Selvaraju et al., 2017; Simonyan et al., 2013; Taha et al., 2020). Recent advances in technology, such as deep learning-based speech detection, allow the accurate detection and analysis of speech in a way that protects the privacy of all participants (Cummins et al., 2018; Little et al., 2021).

Methods and Materials

Participants

Participants aged 16 to 35 years ($M = 20.37$; $SD = 2.55$) provided informed consent and general information, and filled out screening questionnaires consisting of the PHQ-9 (Patient Health Questionnaire 9-item), GAD-7 (Generalized Anxiety Disorder Scale 7-item), ISI (Insomnia Severity Index), and PSS (Perceived Stress Scale) (Du et al., 2017; Meng et al., 2020; Tong et al., 2016; Yu, 2010). After they completed mental health screening, a randomized, cross-sectional clinical trial was performed to assess the short-term effectiveness of biofeedback treatments. To be included, participants had to meet the following criteria: (a) a total of > 5 on the PHQ-9 or a total of > 5 on the GAD-7, (b) no acute suicidal thoughts, (c) willingness to schedule for a 4-week treatment, (d) having access to a computer with internet connection, and (e) no severe or potentially confounding psychiatric disorders (for example, psychosis and substance misuse). A total of 101 individuals completed the trial and were randomly assigned in a 1:1 ratio to the biofeedback intervention group (n = 51) and the waiting list control (WLC) group (n = 50).

Ethical Approval for this study was obtained from the Ethics Committee of Henan Medical University (HYLL2020005). All the participants volunteered to participate in the study and received individual psychometric results at the end of the measurement. In addition, the participants in the longitudinal study also received free biofeedback (ClinicalTrials.gov ChiCTR2100045542).

Research Tools

Biofeedback Intervention

This experimental instrument was selected by Nanjing VISHEE Medical Technology Co., Ltd., for independent research and development of a biofeedback psychological training product-group wireless biofeedback instrument (Free Mind-S /Free Mind-G). The system consists of self-balanced training software loaded on a computer and a blood volume pulse sensor (also called a PPG sensor). Based on traditional biofeedback technology, a group wireless biofeedback instrument has been further developed.

Physiological Indicators

Collection We collected the feedback heart rate variability characters including time domain features (such as the standard deviation of normal to normal (SDNN), the root mean square of successive differences (RMSSD), and the percentage of successive R-R intervals that differ by more than 50 ms (PNN50)), and frequency domain features (such as time domain includes low frequency (LF), high frequency (HF), very low frequency (VLF), rate of LF to HF (LF/HF) and total power (TP)) calculated by using VISHEE Mental Quality Training Assessment Software V2.0. The details of HRV preprocessing were presented in Supplementary methods. All data in this study were processed and analyzed on a computer using SPSS software to explore whether there were differences in the changes in the participants after the end of the four phases of intervention conditioning in group biofeedback training. Using the principle of biofeedback technology, the dynamic changes in individual physiological and psychological states were presented by real-time acquisition of human blood volume pulse (BVP) and frontal surface electromyography (sEMG) signals during the experiment. The group wireless biofeedback instrument collects neuromuscular bioelectrical signals on the frontal surface through EMG sensors and records, enlarges, transmits, and feedbacks them. This process does not cause any trauma to the participants and helps to objectively realize the quantitative evaluation of body relaxation. Participants can see the changes in their coordination state through the computer and obtain the evaluation report and heart rate variability trend at the end of each training session. They can intuitively understand the changes in the autonomic nervous system after efforts.

Speech Data Acquisition

This study was based on the neutral reading proposed in previous studies. Voice reading text was a neutral reading like 'Life like a summer flower' in acoustic sampling. The audio used was an m4a format audio file with an 8 kHz sampling frequency obtained through the mobile phone, and the duration of the audio was approximately 60 s. The recordings were pretested to ensure that they had a uniform format and parameters. The multimedia video processing tool FFmpeg was used to convert the original audio file into a WAV format file with a 16 kHz sampling frequency to facilitate the subsequent speech feature extraction.

The overview of biofeedback intervention content is shown in Supplementary Methods and Table S1.

Assessment

Resting Assessment

Resting assessments can be divided into pre-resting assessment (pre) and post-resting assessments (post). Pre-resting assessment was a static test conducted before the participants officially started the group biofeedback training. After a period of group biofeedback training, a post-rest evaluation was conducted again. Compared to the first evaluation before the experiment, the test can more effectively evaluate the effect achieved by the test members after training.

Biofeedback Intervention

After randomization, the biofeedback group were granted 4 weeks of access to the self-help biofeedback intervention, and the waiting-list group received delayed treatment after 4 weeks. For all randomized participants, the online assessment was conducted via the WeChat Subscription platform at the following intervals: during the initial pre-assessment phase (baseline) and at week 4 after the biofeedback group finished the last biofeedback session. Sociodemographic and clinical characteristics were recorded during the assessment. Changes in the scores on the four psychiatric questionnaires were used as primary outcome measures: PHQ-9, GAD-7, ISI, and PSS.

Audio Data Preprocessing

The audio used in this study was recorded by the participants in their current environment through a mobile phone; therefore, there were problems such as many blank parts at the beginning and the end or the very short duration of the recording time. Therefore, the first step was to use a voice activity detection algorithm (VAD) to delete the silent parts of the beginning and end of the audio. The VAD algorithm uses a Python-based audio processing toolkit *librosa* to extract each frame of each voice data. The short-term energy and zero-crossing rate of the speech, and then the two thresholds in the VAD are determined to obtain the middle speech part. Secondly, the Audio Segment toolkit is used to read the audio from the processed audio by VAD to obtain the audio time, and the audio data with time is less than 10 s is removed.

Feature Selection and Extraction

In this study, we extracted acoustic features from the data using a speech frame, and the speech frame length was 32 ms. For each frame of speech, eight frame-level features were extracted from the perspective of sound prosodic features, such as fundamental frequency and energy, and 29 low-level descriptors (LLDs) were extracted for each frame of speech. LLDs refer to the calculation of acoustic features in a frame of speech and represent the speech features of a frame. Specific feature information is provided in Supplementary Table S2. The acoustic feature extraction of data used *librosa*, a toolkit for Python (McFee et al., 2020). The details of feature extraction processing are presented in Supplementary Methods.

Statistical Analysis

Data analysis was conducted using SPSS 23.0 (IBM, Chicago). Descriptive analyses were used to summarize sociodemographic data, such as age, sex, and study variables. Differences between the study groups in sociodemographic and baseline variables were assessed using the paired samples *t* test, independent samples *t* test, and Mann–Whitney *U* and c2 tests.

Prediction Model

ANN is a mathematical model that emulates the activity of biological neural networks in the human brain, and we used ANN to construct the classification and regression models used in this study. The basic structure of the ANN model consists of five layers of full connections. The first four layers are used for deep feature extraction, and the last layer is used to map the high-dimensional features learned by the first four layers to label the space, calculate the label weight, and then normalize it to obtain the corresponding probability of the outcome. In terms of model parameters, the grid search method is used to obtain the best parameters. The structure and parameters of this neural network model and the parameters selection processes were presented in Supplementary Methods, Fig. S1 and Table S3. The classification and regression models constructed in this study are the same, except for the output layer. The output of the classification model is the binary label 0, 1, where 1 denotes 'response' and 0 denotes non-response. The label was set as a criterion for assessing efficacy: ≥50% reduction in the GAD-7 and ISS total scores 4 weeks after intervention. The output of the regression model was the difference in the GAD-7 and ISS scores before and after the biofeedback treatment.

Results

Participant Characteristics for the Survey and Intervention

A total of 600 eligible participants fulfilled the inclusion criteria, after confirming their willingness to intervene through telephone communication, 107 participants willing to receive intervention were randomly assigned to the biofeedback or WLC group using a web-based randomization procedure (https://www.random.org/lists/) implemented by an independent research assistant. Symptom scale and vocal acoustic measurements were performed at baseline and after the intervention for the biofeedback (n = 54) and WLC groups (n = 53) separately. The biofeedback group performed 12 intervention sessions over a 4-week period, with the control group waiting for 4 weeks. After the voice quality control, the biofeedback intervention group consisting of 52 biofeedback participants (n = 18, female, 34.62%; mean age, 19.02; SD, 0.94) and the waiting group (n = 52) consisting of 52 WLC participants (n = 18, female, 34.62%; mean age, 18.81; SD, 0.79) finally completed the voice data collection before and after treatment.

At baseline, the biofeedback and waiting-list group did not differ in age (T = 1.241, P = 0.218), sex (X^2 = 0.000, P = 1.000), and depression (T = 1.014, P = 0.313), anxiety (T = 0.141, P = 0.888), insomnia symptoms (T = 0.641, P = 0.523) or perceived stress (T = 0.676, P = 0.501).

Biofeedback Intervention Effect of Clinical Symptoms, Acoustic Speech Features, and Physiological Features

The Mann–Whitney *U* test showed that the biofeedback group at post-treatment had improvements in the PHQ-9, GAD-7, ISI, and PSS scores compared with the WLC group. There were significant differences between the biofeedback and WLC groups in ISI changes after 4 weeks (Table 1).

At baseline and week 4, there were no significant differences in acoustic speech features between the biofeedback intervention and WLC groups. However, the changes in speech before and after 4 weeks in the intervention group were significantly different from those in the WLC group, mainly reflected in the changes in the energy parameters and Mel-Frequency Cepstral Coefficients (MFCC), the details of which are shown in Table 2.

The changes in physiology before and after 4 weeks in the biofeedback intervention group were significantly different from those in the WLC group, mainly reflected in the changes in the body relaxation index and surface electromyography (EMG). The main changes in the biofeedback intervention group after 4 weeks compared to baseline were changes in body relaxation index, surface EMG, heart rate, standard deviation of beat-to-beat intervals, standard deviation of the difference between adjacent RR intervals, and total energy, the details of which are shown in Table 3.

Association Between Clinical Symptoms and Acoustic Speech Features

The results of Spearman correlation analysis showed that at the baseline level, the total PHQ-9, GAD-7, and ISS scores of the intervention group were correlated with the energy parameters and MFCC of the speech features. PHQ-9 was correlated with energy parameters (energy_max: $r = 0.349$, $p = 0.012$) and MFCC (mfcc_para1_max: $r = 0.349$, $p = 0.012$). The GAD-7 was correlated with energy parameters (ste_max: $r = 0.286$, $p = 0.042$; ste_ptp: $r = 0.286$, $p = 0.042$; ste_std: $r = 0.281$, $p = 0.046$; ste_de_std: $r = 0.288$, $p = 0.040$; ste_de2_std: $r = 0.288$, $p = 0.041$). ISS was correlated with energy parameters (ste_de2_max: $r = 0.160$, $p = 0.262$; ste_de2_ptp: $r = 0.122$, $p = 0.392$; ste_de2_std: $r = 0.107$, $p = 0.455$) and MFCC (mfcc_para7_skew: $r = 0.119$, $p = 0.406$).

The changes in the energy parameters (energy_max: $r = 0.280$, $p = 0.046$) and MFCC (mfcc_para1_max: $r = 0.280$, $p = 0.046$) were correlated with the improvement in depression before and after the intervention. Changes in the energy parameters (ste_max: $r = 0.279$, $p = 0.047$; ste_ptp: $r = 0.279$, $p = 0.047$; ste_std: $r = 0.324$, $p = 0.020$; ste_de_std: $r = 0.294$, $p = 0.036$; ste_de2_std: $r = 0.363$, $p = 0.009$) were correlated with improvement in anxiety before and after the intervention. The changes in the energy parameters (ste_de2_max: $r = 0.305$, $p = 0.029$; ste_de2_ptp: $r = 0.329$, $p = 0.018$; ste_de2_std: $r = 0.288$, $p = 0.040$) and MFCC (mfcc_para7_skew: $r = 0.286$, $p = 0.042$) were correlated with the improvement in insomnia before and after the intervention.

Association Between Clinical Symptoms and Physiological Features

The results of the Spearman correlation analysis showed that at the baseline level, the total PHQ-9 score of the intervention group was correlated with the physical relaxation index

Table 1 Primary and secondary clinical measures of symptom severity at baseline and 4 weeks in the Biofeedback and WLC groups

Measure	Group	Baseline	Week 4	Change	Mann–Whitney *U* test (Change)[a]		Wilcoxon signed rank test[b]	
		Median (IQR: p25–p75)			Z	*P*	Z	*P*
PHQ-9	Biofeedback	6 (4–8)	4.5 (2–6)	− 2 (− 4–0)	− 1.659	0.097	− 3.317	0.001
	WLC	5 (3–8)	4.5 (2–7)	− 1 (− 2–1)				
GAD–7	Biofeedback	4 (1.25–5.75)	1.5(0–4)	-1.5 (-3-0.75)	− 1.647	0.100	− 3.385	0.001
	WLC	3 (1–6)	2(0-5.75)	0 (− 1–0)				
ISI	Biofeedback	4 (2–6)	2 (0-5.75)	− 1 (− 3–0)	− 2.743	0.006	− 2.491	0.013
	WLC	3 (1–6)	3 (1-7.75)	0 (− 1.75–2)				
PSS	Biofeedback	25 (19.25–27.75)	21 (15.5–27)	− 2 (− 7.75–1)	− 1.000	0.317	− 2.909	0.004
	WLC	22.5 (18-27.75)	22 (18-27.75)	− 1 (− 6–1.75)				

WLC waitlist control, *PHQ-9* 9-item Patient Health Questionnaire, *GAD-7* 7-item Generalized Anxiety Disorder Scale, *ISI* Insomnia Severity Index, *PSS* Perceived Stress Scale, *SD* Standard Deviation

[a]Mann–Whitney *U* test for the difference between Biofeedback and WLC groups

[b]Wilcoxon signed rank test within Biofeedback treatment group (Baseline vs. Week 4)

Table 2 Speech acoustic measure at baseline and 4 weeks in the biofeedback and WLC groups

Speech acoustic measure	Mann–Whitney *U* test (Baseline)[a]		Mann–Whitney *U* test (Week 4)[b]		Mann–Whitney *U* test (Change)[c]	
	Z	*P*	Z	*P*	Z	*P*
energy_de_skew	− 0.559	0.576	− 3.410	0.001	− 2.694	0.007
spl_de_skew	− 0.599	0.549	− 3.658	< 0.001	− 3.089	0.002
mfcc_para7_skew	− 1.081	0.280	− 2.085	0.037	− 2.547	0.011
mfcc_de_para1_skew	− 0.559	0.576	− 3.410	0.001	− 2.694	0.007
mfcc_de_para9_min	− 1.590	0.112	− 2.078	0.038	− 2.259	0.024
mfcc_de_para12_std	− 0.780	0.436	− 2.413	0.016	− 2.580	0.01
f2_de_kur	− 0.519	0.604	− 2.634	0.008	− 2.045	0.041
qingyin_max	− 1.825	0.068	− 1.199	0.231	− 2.522	0.012
qingyin_ptp	− 1.825	0.068	− 1.199	0.231	− 2.522	0.012
energy_max	− 1.683	0.092	− 1.235	0.217	− 2.306	0.021
ste_max	− 1.429	0.153	− 1.362	0.173	− 2.841	0.004
ste_ptp	− 1.416	0.157	− 1.362	0.173	− 2.828	0.005
ste_mean	− 1.496	0.135	− 1.482	0.138	− 2.868	0.004
ste_std	− 1.503	0.133	− 1.523	0.128	− 2.921	0.003
ste_de_min	− 1.402	0.161	− 1.201	0.230	− 2.647	0.008
ste_de_max	− 1.402	0.161	− 1.201	0.230	− 2.694	0.007
ste_de_ptp	− 1.416	0.157	− 1.268	0.205	− 2.647	0.008
ste_de_mean	− 0.980	0.327	− 0.131	0.896	− 2.132	0.033
ste_de_std	− 1.402	0.161	− 1.462	0.144	− 2.875	0.004
ste_de2_min	− 1.650	0.099	− 1.369	0.171	− 2.841	0.004
ste_de2_max	− 1.590	0.112	− 1.215	0.224	− 2.727	0.006
ste_de2_ptp	− 1.636	0.102	− 1.362	0.173	− 2.794	0.005
ste_de2_std	− 1.529	0.126	− 1.476	0.140	− 3.022	0.003
spl_max	− 1.429	0.153	− 1.362	0.173	− 2.219	0.027
mfcc_para1_max	− 1.683	0.092	− 1.235	0.217	− 2.306	0.021
mfcc_para13_mean	− 0.954	0.340	− 1.516	0.130	− 2.326	0.02
mfcc_de2_para4_min	− 1.904	0.057	− 0.566	0.572	− 2.419	0.016
mfcc_de2_para4_ptp	− 1.476	0.140	− 0.519	0.604	− 2.165	0.03
mfcc_de2_para9_kur	− 0.023	0.981	− 1.917	0.055	− 1.998	0.046
b2_max	− 0.144	0.886	− 1.296	0.195	− 1.978	0.048
shimmer_abs	− 1.690	0.091	− 1.021	0.307	− 2.567	0.01

WLC waitlist control, *max* maximum, *ptp* range

[a]Mann–Whitney *U* test on baseline differences between the biofeedback and WLC groups

[b]Mann–Whitney *U* test for the difference in the fourth week between the biofeedback and WLC groups

[c]Mann–Whitney *U* test for differences in pre- and post- changes in the biofeedback and WLC groups

($p = 0.001$), mental relaxation index ($p = 0.015$), surface EMG ($p = 0.001$), and SDNN ($p = 0.041$). The total GAD-7 score of the intervention group correlated with surface EMG ($p = 0.016$), LF ($p = 0.021$), HF ($p = 0.019$), and LF/HF ($p = 0.020$). And changes in physiological data by group before and after treatment were shown in Supplementary Figure S2–S8.

Association Between Acoustic Speech Features and Physiological Features

The results of the Spearman correlation analysis showed the following. The changes in the energy parameters (spl_max) and MFCC (mfcc_de2_para12_ptp) were correlated with the improvement in EMG before and after the intervention.

Table 3 Physiological features measurement at baseline and 4 weeks in the biofeedback and WLC groups

Physiological features measure	Mann–Whitney U test (Baseline)[a]		Mann–Whitney U test (Week 4)[b]		Mann–Whitney U test (Change)[c]		Wilcoxon signed rank test[d]	
	Z	*P*	Z	*P*	Z	*P*	Z	*P*
body_data	– 3.624	< 0.001	– 0.181	0.856	– 3.097	0.002**	– 3.414	0.001
EMG	– 3.624	< 0.001	– 0.181	0.856	– 3.097	0.002**	– 3.414	0.001
h	– 0.060	0.952	– 1.450	0.147	– 1.430	0.153	– 3.136	0.002
LN(sdnn)	– 1.801	0.072	– 0.522	0.602	– 1.148	0.251	– 2.327	0.02
LN(rmssd)	– 2.326	0.02	– 0.167	0.867	– 0.870	0.384	– 2.226	0.026
LN(pnn50)	– 2.219	0.027	– 0.689	0.491	– 2.406	0.016*	– 3.890	< 0.001
LN(tp)	– 0.532	0.595	– 0.954	0.340	-1.001	0.317	– 1.978	0.048

*$p < 0.05$; **$p < 0.01$

[a]Mann–Whitney *U* test on baseline differences between the biofeedback and WLC groups

[b]Mann–Whitney *U* test for the difference between the biofeedback and WLC groups after 4 weeks

[c]Mann–Whitney *U* test for differences in pre- and post- changes in the biofeedback and WLC groups

[d]Wilcoxon signed rank test for the difference after intervention in the biofeedback group

The changes in the energy parameters (energy_de_skew) and MFCC (mfcc_para7_skew) were correlated with the improvement in heart rate before and after the intervention.

The changes in the energy parameters (energy_de_skew) correlated with the improvement in SDNN before and after the intervention. The changes in the energy parameters (energy_de_skew and ste_max) were correlated with the improvement in RMSSD before and after the intervention. The changes in the energy parameters (ste_ptp, ste_std, and ste_de_std) were correlated with the improvement in RMSSD before and after the intervention. The changes in the energy parameters (energy_max) and MFCC (mfcc_para1_max) correlated with the improvement in LF and HF before and after the intervention. The changes in the energy parameters (energy_max, ste_max, and ste_ptp) and MFCC (mfcc_para1_max) were correlated with the improvement in LF/HF before and after the intervention (Table 4).

Table 4 Correlation between acoustic speech features and physiological features in the biofeedback group

Physiological features	Speech acoustic	Spearman correlation (Change)	
		R value	P value
emg	spl_max	0.276	0.05
	mfcc_de2_para12_ptp	0.298	0.034
hr	energy_de_skew	– 0.358	0.010
	mfcc_para7_skew	0.298	0.034
LN(sdnn)	energy_de_skew	0.425	0.002
LN(rmssd)	energy_de_skew	0.347	0.013
LN(pnn50)	ste_max	0.345	0.013
	ste_ptp	0.345	0.013
	ste_std	0.286	0.042
	ste_de_std	0.286	0.042
LN(lf)	energy_max	– 0.323	0.021
	mfcc_para1_max	– 0.323	0.021
LN(hf)	energy_max	0.299	0.033
	mfcc_para1_max	– 0.323	0.021
LF/HF	energy_max	– 0.285	0.043
	mfcc_para1_max	– 0.323	0.021
	ste_max	– 0.0277	0.049
	ste_ptp	– 0.0277	0.049

Artificial Neural Network Prediction Model by Speech Acoustic Features

Because the LR model only considers linear classification, an ANN-based classification model and a regression model were constructed for verification.

This model, based on ANN binary classification to detect anxiety and insomnia, uses speech feature differences between the responder and non-responder groups as input features. Figures 1 and 2 show the prediction results of this model for 51 participants, in which 0 means "Non-responder," 1 means "Responder," the horizontal axis represents the predicted result, and the vertical axis represents the original value of the label. For anxiety, the correct diagnosis rate of this classification model for the responding group was 0.62, indicating a low misdiagnosis rate; the overall accuracy rate was 62% (Fig. 1). For insomnia, the correct diagnosis rate of this classification model for the responding group was 0.58, indicating a low misdiagnosis rate, and the overall accuracy rate was 58% (Fig. 2). This indicates that the classification model can predict the population that will respond well to treatment with biofeedback.

At the same time, Figs. 3 and 4 show the performance of the classifier per fold when the model was subjected to 5-fold cross-validation. From the ROC curve, the area of the ROC curve of the model in each fold was greater than 0.5,

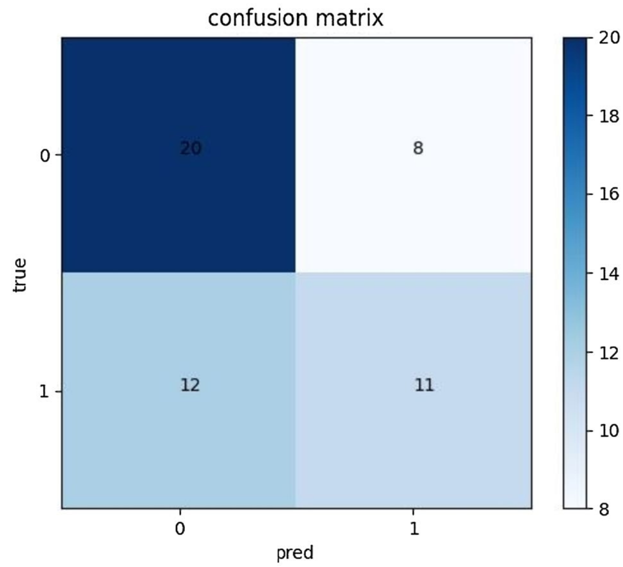

Fig. 1 Prediction results of anxiety

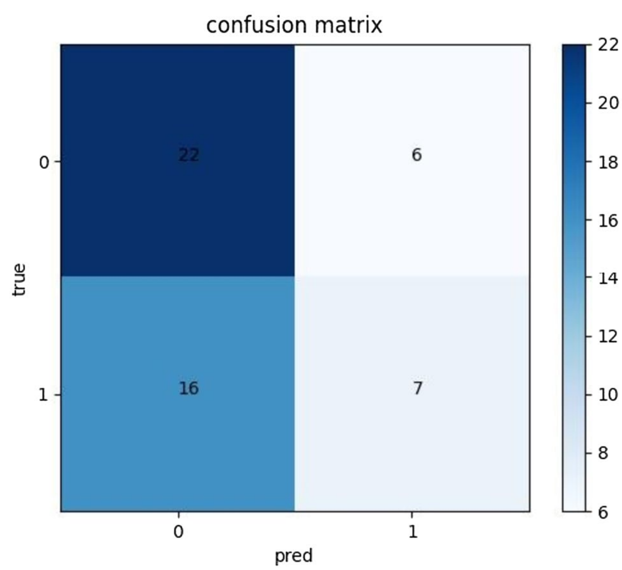

Fig. 2 Prediction results of insomnia

which proves that the training of this classifier in each fold was effective.

Discussion

Principal Results

Our study found that speech features, such as energy parameters and MFCC, can be more accurate and objective indicators for tracking biofeedback therapy response

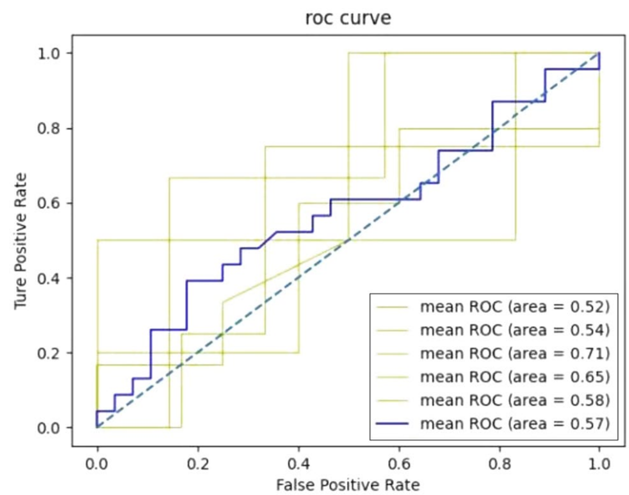

Fig. 3 ROC curve of 5-fold cross-validation model of anxiety

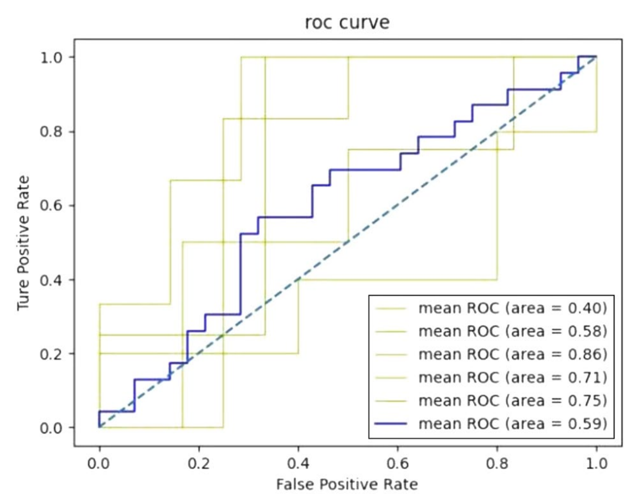

Fig. 4 ROC curve of 5-fold cross-validation model of insomnia

and predicting efficacy. Voice is also associated with mood. For example, depression causes changes in the somatic and automatic nervous systems that reflect muscle tension and respiratory rate (Ellgring & Scherer, 1996). Previous studies have suggested that patients with depression have fewer vocal tract changes, which may be related to tighter vocal tracts due to psychomotor retardation. Increased muscle tension and changes in salivation and mucus secretion affect the vocal tract and limit articulatory movement (Albuquerque et al., 2021; Ellgring & Scherer, 1996).

Information contained in speech can be rather precise and convey extra information about our physical and mental states, as well as emotions (Kacur et al., 2021). Zhu et al. found that five types of features were considered the most useful to reflect a speaker's emotion: pitch, short-term energy, short-term zero-crossing rate format, and MFCC

(Zhu et al., 2017). This conclusion supports our finding that, after treatment, the difference between the intervention and waiting groups was reflected in the energy parameters and MFCC. Short-term energy reflects the amplitude characteristics of speech signals, which can be used to distinguish between voice and noise (Zhu et al., 2017), clear and cloudy sounds, and vowels from rhymes. They can be used as suprasegmental information for speech recognition. Spectral features usually reflect the short-term characteristics of the speech signal, related to changes in muscle tone and control associated with vocalization, and reflect the correlation between changes in vocal tract shape and the occurrence of motion. They are also correlated with individuals' mood (France et al., 2000; Ozdas et al., 2004; Taguchi et al., 2018). The MFCC is one of the most widely used features for speech analysis. This spectral feature describes the envelope characteristics of the short-time power of speech. It is extracted based on the human auditory system, which provides a natural and real reference for speech recognition (Zhu et al., 2017). Taguchi et al. investigated the differences in the MFCCs of individuals with and without depression and found evidence of higher levels of sensitivity and specificity in the second dimension of MFCC, concluding that this dimension could be a discriminating factor between patients with depression and healthy patients and, consequently, a depression biomarker (Taguchi et al., 2018). As anxiety disorders are reflected in people's voices due to the somatic symptoms associated with the respiratory system, acoustic parameters could be used as an objective method to assist in the assessment of anxiety symptoms (Albuquerque et al., 2021; Oezseven et al., 2018). For example, in general, MFCCs decrease with anxiety (Albuquerque et al., 2021). These previous findings directly or indirectly support our view that energy parameters and the MFCC can be used as speech markers for the diagnosis of depression and anxiety. In addition, we found that energy parameters and MFCC can be used as voice markers for the diagnosis of insomnia. Insomnia often appears as a comorbidity of depression and anxiety, but there are almost no previous studies on this topic, and our results fill this literature gap.

Biofeedback therapy transforms normally imperceptible biological signals into directly observable information, allowing patients to more visually and consciously learn to control physiological activities that are not otherwise controlled by the mind, and consequently, better control their mental state (Dormal et al., 2021). The assessment of clinical and physiological outcomes was good for the effect of the biofeedback intervention. Previous studies have confirmed our findings that physical and mental health-related symptoms such as depression, anxiety, insomnia, and stress improved after biofeedback intervention (Alneyadi et al., 2021; Dormal et al., 2021; Maynart et al., 2021). Some studies found that some physiological indicators were also improved after biofeedback intervention, such as HRV, EMG, and measures of lung function (Delk et al., 1994; Dormal et al., 2021; Giggins et al., 2013). However, there are relatively few articles on biofeedback interventions for changes in speech features, and Graf et al. found that after biofeedback, the intensities for all vowels increased (Hoyer & Graf, 2019).

The main principle of deep learning is the use of NN: many hidden layers of increasing levels of abstraction are employed to learn a hierarchical representation of the data. The use of deep learning methods has proven promising in the field of psychiatry, and deep learning algorithms are considered one of the most promising machine learning techniques, even if the results are not easy to interpret (Squarcina et al., 2021; Zhang et al., 2018).

In addition, patients with the same categorical diagnosis may respond differently to the same treatment (Squarcina et al., 2021). In summary, we utilized a deep learning neural network approach to predict the efficacy of a biofeedback intervention. The energy parameters and MFCC of speech features can predict whether a biofeedback intervention can effectively improve anxiety and insomnia symptoms and treatment response. Advances in deep learning now allow for rapid and accurate measurement of myriad markers that have already demonstrated robust effects in clinical populations (Schultebraucks et al., 2020). Predictive models were constructed using deep learning to explore the acoustic features of anxiety and insomnia, consider possible complex nonlinear correlations between acoustic features and anxiety and insomnia, and assess the validity of the predictive models. Previous studies have found that voice prosody and speech content all contribute uniquely to the classification and prediction of MDD (Little et al., 2021; Muzammel et al., 2021; Schultebraucks et al., 2020). The current results are encouraging, and the use of speech data sources is scalable, as it can be integrated into cellphones and web-based tele-medicine applications. This can significantly increase the assessment of clinical function (Schultebraucks et al., 2022).

Physical indicators are closely related to emotional problems; for example, heart rate and HRV are widely used markers of physiological activation associated with exaggerated emotional responses during situations of acute emotional provocation (Mather, 2021; Predatu et al. 2021). In addition, EMG is related to mental stress, and Schleifert et al. found that when mental stress increased, EMG also increased (Schleifer et al., 2008). This corroborates our observations that biofeedback improves insomnia, anxiety, depression, and stress. Physiological characteristics can be used as objective assessment indicators of the effectiveness of biofeedback interventions.

Limitations

This study aimed to provide more knowledge on the treatment of college-going students with depression and anxiety through biofeedback. However, there are some limitations to this study, and not enough previous studies that can confirm the association between insomnia and related voice characteristics. This finding needs to be verified in future studies.

Comparison with Prior Work

Almost no previous studies have used speech features as an evaluation of biofeedback efficacy, and we used speech features as an objective indicator and obtained positive results. We pioneered the use of deep learning methods to build models using speech indicators as predictors of individual efficacy with biofeedback. The results obtained were better than expected.

Conclusions

Depression, anxiety, insomnia, and stress in the college-going student population improved after biofeedback, with insomnia showing a particularly significant improvement. After 4 weeks of intervention, the differences in speech were reflected in energy parameters and MFCC in the intervention group compared to the waiting group. The differences in the physiological aspects were reflected in the changes in the body relaxation index and surface EMG. Moreover, there was a correlation between clinical symptoms, speech characteristics, and physical characteristics. Finally, we found that logistic regression and deep learning models constructed using the above speech features could effectively classify and predict insomnia and anxiety symptoms. Overall, the results of this study will provide valuable information about the role of biofeedback in improving the mental health of college students, as well as more accurate and objective indicators for assessing emotional problems.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10484-023-09612-3.

Acknowledgements The authors thank all participants who took part in this study. The authors thank all participants who took part in this study. The support to recruitment by personnel in Xinxiang Medical University is acknowledged. The authors thank Xinxiang Medical University and The Affiliated Brain Hospital of Nanjing Medical University for excellent research assistance.

Author Contributions The two authors contributed equally to this work: Lifei Wang , Rongxun LiuCorresponding authors: Fei Wang fei.wang@yale.edu. Xizhe Zhang zhangxizhe@gmail.com.Xiao Xu prepared Figs. 1, 2, 3 and 4. All authors reviewed the manuscript.

Funding This study was funded by National Science Fund for Distinguished Young Scholars (81725005 to Fei Wang), National Natural Science Foundation Regional Innovation and Development Joint Fund (U20A6005 to Fei Wang), Jiangsu Provincial Key Research and Development Program (BE2021617), National Natural Science Foundation of China (62176129 to Xizhe Zhang) and Key Project supported by Medical Science and Technology Development Foundation, Jiangsu Commission of Health (ZD2021026 to Zhu Rongxin).

Declarations

Conflict of interest The authors declare that the authors have no conflict of interest. The authors have full control of all primary data and that agree to allow the journal to review their data if requested

References

Albuquerque, L., Valente, A. R. S., Figueiredo, D., Teixeira, A., Sa-Couto, P., & Oliveira, C. (2021). Association between acoustic speech features and non-severe levels of anxiety and depression symptoms across lifespan. *PloS One, 16*(4), e0248842. https://doi.org/10.1371/journal.pone.0248842

Alneyadi, M., Drissi, N., & Almeqbaali, M. (2021). Biofeedback-based connected mental health interventions for anxiety: Systematic literature review. *JMIR mHealth and uHealth, 9*(4), e26038. https://doi.org/10.2196/26038

Auerbach, R. P., Alonso, J., Axinn, W. G., Cuijpers, P., Ebert, D. D., Green, J. G., Hwang, I., Kessler, R. C., Liu, H., & Mortier, P. (2016). Mental disorders among college students in the world health organization world mental health surveys. *Psychological Medicine, 1*(14), 2955–2970. https://doi.org/10.1017/S0033291716001665

Auerbach, R. P., Mortier, P., Bruffaerts, R., Alonso, J., Benjet, C., Cuijpers, P., Demyttenaere, K., Ebert, D. D., Green, J. G., Hasking, P., Murray, E., Nock, M. K., Pinder-Amaker, S., Sampson, N. A., Stein, D. J., Vilagut, G., Zaslavsky, A. M., Kessler, R. C., WHO, WMH-ICS, & Collaborators. (2018). WHO world mental health surveys international college student project: prevalence and distribution of mental disorders. *Journal of Abnormal Psychology, 127*(7), 623–38. https://doi.org/10.1037/abn0000362

Beiter, R., Nash, R., Mccrady, M., Rhoades, D., Linscomb, M., Clarahan, M., & Sammut, S. (2015). The prevalence and correlates of depression, anxiety, and stress in a sample of college students. *Journal of Affective Disorders, 173*, 90–96. https://doi.org/10.1016/j.jad.2014.10.054

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, *5*(2), 157–166. https://doi.org/10.1109/72.279181.

Bernard, J. A., & Mittal, V. A. (2015). Updating the research domain criteria: The utility of a motor dimension. *Psychological Medicine, 45*(13), 2685–2689. https://doi.org/10.1017/S0033291715000872

Cannizzaro, M., Harel, B., Reilly, N., Chappell, P., & Peter, J. S. (2004). Voice acoustical measurement of the severity of major depression. *Brain and Cognition, 56*(1), 30–35. https://doi.org/10.1016/j.bandc.2004.05.003

Cummins, N., Baird, A., & Björn, W. S. (2018). Speech analysis for health: Current state-of-the-art and the increasing impact of deep learning. *Methods (San Diego Calif), 151*, 41–54. https://doi.org/10.1016/j.ymeth.2018.07.007

da Maynart, W. H., de Albuquerque, M. C. D. S., Santos, R. C. S., de Albuquerque Sarmento, P., da Silva, J. J., Costa, C. D. S. G., da

Silva Santos, J. D., de Oliveira Pontes, C., dos Santos Barros, M. C., Belo, F. M. P., & Costa, C. R. B. (2021). The use of biofeedback intervention in the improvement of depression levels: A randomised trial. *Acta Neuropsychiatrica, 33*(3), 126–133. https://doi.org/10.1017/neu.2020.46

de Girolamo, G., Dagani, J., Purcell, R., Cocchi, A., & McGorry, P. D. (2012). Age of onset of mental disorders and use of mental health services: Needs, opportunities and obstacles. *Epidemiology and Psychiatric Sciences, 21*(1), 47–57. https://doi.org/10.1017/s2045796011000746

Delk, K. K., Gevirtz, R., Hicks, D. A., & Carden, F. (1994). The effects of biofeedback assisted breathing retraining on lung functions in patients with cystic fibrosis. *Chest, 105*(1), 23–28. https://doi.org/10.1378/chest.105.1.23

Di, Y., Wang, J., Li, W., & Zhu, T. (2021). Using I-Vectors from voice features to identify major depressive disorder. *Journal of Affective Disorders, 288*, 161–166. https://doi.org/10.1016/j.jad.2021.04.004

Dormal, V., Vermeulen, N., & Mejias, S. (2021). Is heart rate variability biofeedback useful in children and adolescents? A systematic review. *Journalof Child Psychology and Psychiatry.* https://doi.org/10.1111/jcpp.13463

Du, N., Yu, K., Ye, Y., & Chen, S. (2017). Validity study of patient health questionnaire-9 items for internet screening in depression among Chinese university students. *Asia–Pacific Psychiatry: Official Journal of the Pacific Rim College of Psychiatrists*. https://doi.org/10.1111/appy.12266

Ellgring, H., & Scherer, K. R. (1996). Vocal indicators of mood change in depression. *Journal of Nonverbal Behavior, 20*(2), 83–110. https://doi.org/10.1007/BF02253071

France, D. J., Shiavi, R. G., Silverman, S., Silverman, M., & Wilkes, D. M. (2000). Acoustical properties of speech as indicators of depression and suicidal risk. *IEEE Transactions on Bio-Medical Engineering, 47*(7), 829–837. https://doi.org/10.1109/10.846676

Freeman, D., Sheaves, B., Goodwin, G. M., Yu, L. M., Nickless, A., Harrison, P. J., Emsley, R., Luik, A. I., Foster, R. G., Wadekar, V., Hinds, C., Gumley, A., Jones, R., Lightman, S., Jones, S., Bentall, R., Kinderman, P., Rowse, G., Brugha, T., … Espie, C. A. (2017). The effects of improving sleep on mental health (OASIS): A randomised controlled trial with mediation analysis. *The Lancet Psychiatry, 4*(10), 749–758. https://doi.org/10.1016/S2215-0366(17)30328-0

Gaggioli, A., Pallavicini, F., Morganti, L., & Serino, S. (2014). Experiential virtual scenarios with real-time monitoring (interreality) for the management of psychological stress: A block randomized controlled trial. *Journal of Medical Internet Research, 16*(7), e167. https://doi.org/10.2196/jmir.3235

Giggins, O. M., Persson, U. M. C., & Caulfield, B. (2013). Biofeedback in rehabilitation. *Journal of Neuroengineering and Rehabilitation, 10*, 60. https://doi.org/10.1186/1743-0003-10-60

Hasson, U., Nastase, S. A., & Goldstein, A. (2020). Direct fit to nature: An evolutionary perspective on biological and artificial neural networks. *Neuron, 105*(3), 416–434. https://doi.org/10.1016/j.neuron.2019.12.002

Hoyer, P., & Graf, S. (2019). Adjustment of the vocal tract shape via biofeedback: A case study. *Journal of Voice: Official Journal of the Voice Foundation, 33*(4), 482–489. https://doi.org/10.1016/j.jvoice.2018.01.018

Kacur, J., Puterka, B., & Pavlovicova, J. (2021). On the speech properties and feature extraction methods in speech emotion recognition. *Sensors (Basel Switzerland), 21*(5), 1888. https://doi.org/10.3390/s21051888

Kessler, R. C., Amminger, G. P., Aguilar-Gaxiola, S., Alonso, J., Lee, S., & Üstün, T. B. (2007). Age of onset of mental disorders: a review of recent literature. *Current Opinion in Psychiatry, 20*(4), 359–364. https://doi.org/10.1097/YCO.0b013e32816ebc8c

König, A., Mallick, E., Tröger, J., Linz, N., Zeghari, R., & Manera, V. (2021). Measuring neuropsychiatric symptoms in patients with early cognitive decline using speech analysis. *European Psychiatry: The Journal of the Association of European Psychiatrists, 64*(1), e64. https://doi.org/10.1192/j.eurpsy.2021.2236

Kotwas, I., Mcgonigal, A., Khalfa, S., Bastien-Toniazzo, M., & Bartolomei, F. (2017). A case-control study of skin conductance biofeedback on seizure frequency and emotion regulation in drug-resistant temporal lobe epilepsy. *International Journal of Psychophysiology*. https://doi.org/10.1016/j.ijpsycho.2017.10.005

Kraepelin, E. (1921). Manic depressive insanity and paranoia. *The Journal of Nervous and Mental Disease, 53*(4), 350. https://doi.org/10.1097/00005053-192104000-00057

Krusemark, E. A., & Wen, L. (2012). Enhanced olfactory sensory perception of threat in anxiety: An event-related FMRI study. *Chemosensory Perception, 5*(1), 37–45. https://doi.org/10.1007/s12078-011-9111-7

Kuny, S., & Stassen, H. H. (1993). Speaking behavior and voice sound characteristics in depressive patients during recovery. *Journal of Psychiatric Research, 27*(3), 289–307. https://doi.org/10.1016/0022-3956(93)90040-9

Lee, S., Suh, S. W., Kim, T., Kim, K., Lee, K. H., Lee, J. R., Han, G., Hong, J. W., Han, J. W., Lee, K., & Kim, K. W. (2021). Screening major depressive disorder using vocal acoustic features in the elderly by sex. *Journal of Affective Disorders, 291*, 15–23. https://doi.org/10.1016/j.jad.2021.04.098

Little, B., Alshabrawy, O., Stow, D., Nicol Ferrier, I., McNaney, R., Jackson, D. G., Ladha, K., Ladha, C., Ploetz, T., Bacardit, J., Olivier, P., Gallagher, P., & O'Brien, J. T. (2021). Deep learning-based automated speech detection as a marker of social functioning in late-life depression. *Psychological Medicine, 51*(9), 1441–1450. https://doi.org/10.1017/S0033291719003994

Low, D. M., Bentley, K. H., & Ghosh, S. S. (2020). Automated assessment of psychiatric disorders using speech: A systematic review. *Laryngoscope Investigative Otolaryngology, 5*(1), 96–116. https://doi.org/10.1002/lio2.354

Mather, M. (2021). Is there a maximum desirable heart rate variability? *Neuroscience & Biobehavioral Reviews, 128*, 87–89. https://doi.org/10.1016/j.neubiorev.2021.06.015

McFee, Brian, Vincent Lostanlen, Alexandros Metsai, Matt McVicar, Stefan Balke, Carl Thomé, Colin Raffel, Frank Zalkow, Ayoub Malek, Dana, Kyungyun Lee, Oriol Nieto, Jack Mason, Dan Ellis, Eric Battenberg, Scott Seyfarth, Ryuichi Yamamoto, Keunwoo Choi, viktorandreevichmorozov, Josh Moore, Rachel Bittner, Shunsuke Hidaka, Ziyao Wei, nullmightybofo, Darío Hereñú, Fabian-Robert Stöter, Pius Friesch, Adam Weiss, Matt Vollrath, and Taewoon Kim. 2020. "Librosa/Librosa: 0.8.0.

Meng, R., Li, J., Wang, Z., Zhang, D., Liu, B., & Luo, Y. (2020). The Chinese version of the perceived stress questionnaire: Development and validation amongst medical students and workers. *Health and Quality of Life Outcomes, 18*(1), 70. https://doi.org/10.1186/s12955-020-01307-1

Mundt, J. C., Adam, P., Vogel, D. E., Feltner, & Lenderking, W. R. (2012). Vocal acoustic biomarkers of depression severity and treatment response. *Biological Psychiatry, 72*(7), 580–587. https://doi.org/10.1016/j.biopsych.2012.03.015

Muzammel, M., Salam, H., & Othmani, A. (2021). End-to-end multimodal clinical depression recognition using deep neural networks: A comparative analysis. *Computer Methods and Programs in Biomedicine, 211*, 106433. https://doi.org/10.1016/j.cmpb.2021.106433

Oezseven, T., Duegenci, M., Doruk, A., & Hilal, I. K. (2018). Voice traces of anxiety: Acoustic parameters affected by anxiety disorder. *Archives of Acoustics*, *43*(4), 625–636. https://doi.org/10.24425/aoa.2018.125156.

Ozdas, A., Shiavi, R. G., Silverman, S. E., Silverman, M. K., & Mitchell Wilkes, D. (2004). Investigation of vocal jitter and glottal flow spectrum as possible cues for depression and near-term suicidal risk. *IEEE Transactions on Bio-Medical Engineering, 51*(9), 1530–1540. https://doi.org/10.1109/TBME.2004.827544

Predatu, R., David, O. A., Ovidiu David, D., & Maffei, A. (2021). Emotion regulation abilities as a predictor of anxiety, positive emotions, heart rate, and alpha asymmetry during an impromptu speech task. *Anxiety, Stress, & Coping, 34*(6), 719–733. https://doi.org/10.1080/10615806.2020.1867717

Salekin, A., Eberle, J. W., Glenn, J. J., Teachman, B. A., & Stankovic, J. A. (2018). A weakly supervised learning framework for detecting social anxiety and depression. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2*(2), 81. https://doi.org/10.1145/3214284

Schleifer, L. M., Spalding, T. W., Kerick, S. E., Cram, J. R., & Ley, R. (2008). Mental stress and trapezius muscle activation under psychomotor challenge: A focus on EMG gaps during computer work. *Psychophysiology, 45*(3), 356–365. https://doi.org/10.1111/j.1469-8986.2008.00645.x

Schoenberg, P., & David, A. S. (2014). Biofeedback for psychiatric disorders: A systematic review. *Applied Psychophysiology and Biofeedback, 39*(2), 109. https://doi.org/10.1007/s10484-014-9246-9

Schultebraucks, K., Yadav, V., Shalev, A. Y., Bonanno, G. A., & Galatzer-Levy, I. R. (2020). Deep learning-based classification of posttraumatic stress disorder and depression following trauma utilizing visual and auditory markers of arousal and mood. *Psychological Medicine, 52*(5), 957–967. https://doi.org/10.1017/S0033291720002718

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via griadient-based localization. in IEEE international conference on computer vision.

Sheaves, B., Porcheret, K., Tsanas, A., Espie, C. A., Foster, R. G., Harrison, D. F. P. J., & Wulff, K. (2016). Insomnia, nightmares, and chronotype as markers of risk for severe mental illness: Results from a student population. *Sleep, 39*(1), 173–181. https://doi.org/10.5665/sleep.5342

Silber-Varod, V., Kreiner, H., Lovett, R., Levi-Belz, Y., & Amir, N. (2016). Do social anxiety individuals hesitate more? The Prosodic profile of hesitation disfluencies in social anxiety disorder individuals. in Speech Prosody, Boston 2016.

Simonyan, K., Vedaldi, A., & Zisserman, A. (2013). Deep inside convolutional networks: Visualising image classification models and saliency maps. *Computer Ence*. https://doi.org/10.48550/arXiv.1312.6034

Squarcina, L., Villa, F. M., Nobile, M., & Grisan, E. (2021). Deep learning for the prediction of treatment response in depression. *Journal of Affective Disorders, 281*, 618–622. https://doi.org/10.1016/j.jad.2020.11.104

Sverdlov, A., Curcic, J., Hannesdottir, K., De Luca, V., Ambrosetti, F., Zhang, B., Praestgaard, J., Vallejo, V., Dolman, A., Gomez-Mancilla, B., & Biliouris, K. (2021). A study of new exploratory tools, digital technologies and biomarkers to characterize depression. *Frontiers in Psychiatry*. https://doi.org/10.3389/fpsyt.2021.640741

Taguchi, T., Tachikawa, H., Nemoto, K., Suzuki, M., Nagano, T., Tachibana, R., & Nishimura, M. (2018). Major depressive disorder discrimination using vocal acoustic features. *Journal of Affective Disorders, 225*, 214–220. https://doi.org/10.1016/j.jad.2017.08.038

Taha, A., Yang, X., Shrivastava, A., & Davis, L. (2020). A generic visualization approach for convolutional neural networks. https://doi.org/10.1007/978-3-030-58520-4_43.

Tolkmitt, F., Helfrich, H., Standke, R., & Scherer, K. R. (1982). Vocal indicators of psychiatric treatment effects in depressives and schizophrenics. *Journal of Communication Disorders, 15*(3), 209–222. https://doi.org/10.1016/0021-9924(82)90034-x

Tong, X., An, D., McGonigal, A., & Park, S. P. (2016). Validation of the generalized anxiety disorder-7 (GAD-7) among Chinese people with epilepsy. *Epilepsy Research, 120*, 31–36. https://doi.org/10.1016/j.eplepsyres.2015.11.019

Valiant, L. G. (1984). A theory of the learnable. *Communications of the ACM*. https://doi.org/10.1145/1968.1972

Vincent, P., Larochelle, H., Lajoie, I., & Bengio, Y., and Pierre Antoine Manzagol (2010). Stacked Denoising autoencoders: Learning useful representations in a Deep Network with a local Denoising Criterion. *Journal of Machine Learning Research*, *11*(12), 3371–3408. https://doi.org/10.1016/j.mechatronics.2010.09.004.

Wang, J., Zhang, L., Liu, T., & Wei Pan, B. H. (2019). Acoustic differences between healthy and depressed people: A cross-situation study. *Bmc Psychiatry, 19*(1), 300. https://doi.org/10.1186/s12888-019-2300-7

Weerdmeester, J., van Rooij, M. M., & Engels, R. C. (2020). An integrative model for the effectiveness of biofeedback interventions for anxiety regulation: Viewpoint. *Journal of Medical Internet Research, 22*(7), e14958. https://doi.org/10.2196/14958

Yann, L., Bottou, L., Bengio, Y., & Haffner, P. n.d. Gradientbased learning applied to document recognition.

Yu, D. S. F. (2010). Insomnia severity index: Psychometric properties with Chinese community-dwelling older people. *Journal of Advanced Nursing, 66*(10), 2350–2359. https://doi.org/10.1111/j.1365-2648.2010.05394.x

Zeiler, M. D., & Fergus, R. (2013). Visualizing and understanding convolutional neural networks. *Springer International Publishing*. https://doi.org/10.1007/978-3-319-10590-1_53.

Zhang, Q. S., Zhu, S. C., & California, U. O. (2018). Visual interpretability for deep learning: A survey. *Frontiers of Information Technology & Electronic Engineering*. https://doi.org/10.1631/FITEE.1700808

Zhu, L., Chen, L., & Zhao, D. (2017). Emotion recognition from chinese speech for smart affective services using a combination of SVM and DBN. *Sensors (Basel Switzerland), 17*(7), e1694. https://doi.org/10.3390/s17071694

Publisher's Note Springer nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Lifei Wang[1,2] · Rongxun Liu[1,2,3] · Yang Wang[1,2,4] · Xiao Xu[5] · Ran Zhang[1,2] · Yange Wei[1,2] · Rongxin Zhu[1,2] · Xizhe Zhang[5] · Fei Wang[1,2,6]

✉ Xizhe Zhang
zhangxizhe@njmu.edu.cn

✉ Fei Wang
fei.wang@yale.edu

Lifei Wang
lifeiwang0428@163.com

Rongxun Liu
liurongxun@xxmu.edu.cn

Yang Wang
yangwang201224@163.com

Xiao Xu
xuxiao@stu.njmu.edu.cn

Ran Zhang
ran_1994@yeah.net

Yange Wei
weiyange@xxmu.edu.cn

Rongxin Zhu
zhurx2000@njmu.edu.cn

1 Early Intervention Unit, Department of Psychiatry, The Affiliated Brain Hospital of Nanjing Medical University, Nanjing, People's Republic of China

2 Functional Brain Imaging Institute of Nanjing Medical University, Nanjing, People's Republic of China

3 Henan Key Laboratory of Immunology and Targeted Drugs, School of Laboratory Medicine, Xinxiang Medical University, Xinxiang, People's Republic of China

4 Psychology Institute, Inner Mongolia Normal University, Hohhot, Inner Mongolia, People's Republic of China

5 School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, Jiangsu, China

6 Department of Mental Health, School of Public Health, Nanjing Medical University, Nanjing, China