



## OPEN ACCESS

EDITED BY  
Kedra Wallace,  
University of Mississippi Medical  
Center, United States

REVIEWED BY  
Yingli Shen,  
Minzu University of China, China  
Yun Jin,  
Jiangsu Normal University, China  
Durai Raj Vincent P. M.,  
VIT University, India

\*CORRESPONDENCE  
Xin Liu  
✉ lx@xzhmu.edu.cn

†These authors have contributed  
equally to this work and share first  
authorship

SPECIALTY SECTION  
This article was submitted to  
Digital Public Health,  
a section of the journal  
Frontiers in Public Health

RECEIVED 09 July 2022  
ACCEPTED 09 December 2022  
PUBLISHED 23 December 2022

CITATION  
Xu H, Wu X and Liu X (2022) A  
measurement method for mental  
health based on dynamic multimodal  
feature recognition.  
*Front. Public Health* 10:990235.  
doi: 10.3389/fpubh.2022.990235

COPYRIGHT  
© 2022 Xu, Wu and Liu. This is an  
open-access article distributed under  
the terms of the [Creative Commons  
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,  
distribution or reproduction in other  
forums is permitted, provided the  
original author(s) and the copyright  
owner(s) are credited and that the  
original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution  
or reproduction is permitted which  
does not comply with these terms.

# A measurement method for mental health based on dynamic multimodal feature recognition

Haibo Xu<sup>1,2†</sup>, Xiang Wu<sup>3,4†</sup> and Xin Liu<sup>1,2\*</sup>

<sup>1</sup>Center for Mental Health Education and Research, Xuzhou Medical University, Xuzhou, China, <sup>2</sup>School of Management, Xuzhou Medical University, Xuzhou, China, <sup>3</sup>Institute of Medical Information Security, Xuzhou Medical University, Xuzhou, China, <sup>4</sup>School of Medical Information and Engineering, Xuzhou Medical University, Xuzhou, China

**Introduction:** The number of college students with mental problems has increased significantly, particularly during COVID-19. However, the clinical features of early-stage psychological problems are subclinical, so the optimal intervention treatment period can easily be missed. Artificial intelligence technology can efficiently assist in assessing mental health problems by mining the deep correlation of multi-dimensional data of patients, providing ideas for solving the screening of normal psychological problems in large-scale college students. Therefore, we propose a mental health assessment method that integrates traditional scales and multimodal intelligent recognition technology to support the large-scale and normalized screening of mental health problems in colleges and universities.

**Methods:** Firstly, utilize the psychological assessment scales based on human-computer interaction to conduct health questionnaires based on traditional methods. Secondly, integrate machine learning technology to identify the state of college students and assess the severity of psychological problems. Finally, the experiments showed that the proposed multimodal intelligent recognition method has high accuracy and can better proofread normal scale results. This study recruited 1,500 students for this mental health assessment.

**Results:** The results showed that the incidence of moderate or higher stress, anxiety, and depression was 36.3, 48.1, and 23.0%, which is consistent with the results of our multiple targeted tests.

**Conclusion:** Therefore, the interactive multimodality emotion recognition method proposed provides an effective way for large-scale mental health screening, monitoring, and intervening in college students' mental health problems.

## KEYWORDS

mental health assessment, emotion recognition, interactive assessment scale, video feature extraction, deep learning

## 1. Introduction

Most symptoms of early mental health diseases are subclinical, and it is hard to give a definitive diagnosis (1–3). Affected by COVID-19 in recent years, the mental health problems of college students have become more serious. As related research reports that 54% of students had moderate or severe psychological problems, 29% had moderate-to-severe anxiety, and 17% had moderate-to-severe depressive symptoms. So it is an

urgent problem how to make a clear and accurate evaluation of degrees of mental health illnesses of college students and give corresponding or personalized treatment plans in the state of COVID-19 (4, 5).

The traditional mental health evaluation method is the psychological scale, such as Symptom Checklist 90 (SCL-90) (6), Patient Health Questionnaire-9 (PHQ-9) (7), 16-item Quick Inventory of Depressive Symptomatology and Self-Report (QIDS-SR16) (8), Depression Anxiety and Stress Scale (DASS-21) and so on (9). DASS-21 is one of the most widely used and preferred scale versions (10). Although the convergent validity of DASS-21 is crucial, there is still a lack of testing for its convergent validity (11–13). Compared with children and adults, college students have colonial existence frequently, which means that mental health problems are inevitably caused by contradictions between people (14–16). Moreover, some college students failed to provide accurate and timely responses to inquiries about mental health assessments (17–19), which leads to the results of psychological assessments being inaccurate (20, 21). Using objective, intelligent and efficient methods except psychological scales to assess the college student's mental health degree is an urgent problem.

With the rapid development of information technology, Artificial Intelligence (AI) can mine the deep correlation of multiple layers of psychological data, which has attracted widespread attention (22–24). Related studies have shown that individuals with mental health problems have abnormal physical characteristics (mood, voice, body shape, etc.) (25, 26). For example, Simcock et al. examined the association between psychological problems in early adolescence and facial emotion recognition, showing that anxiety and depression were significantly associated with angry expressions (27). Arevian et al. (28) evaluate the feasibility of collecting speech samples from people with serious mental illness and explores the potential utility of tracking changes in clinical state over time. Santore (29) explores the relationship between repetitive behaviors and obsessive-compulsive disorder and autism by studying adolescents with autism spectrum disorders. Therefore, combining multi-modal data and intelligent information technology can effectively assist in evaluating individuals' mental health problems, which is more effective for large-scale college students' mental health screening.

To provide a more objective and effective evaluation method for college students' normalized mental health, we propose a method for assessing psychological problems that combine multimodal data with traditional scales, intelligently correcting the evaluation results by capturing the unconscious characteristic data of individuals and giving the evaluation results quickly and accurately. The main contributions are as follows: propose the psychological assessment scales survey of college students utilizing a human-computer interaction measure model; collect multiple characteristic data for multimodal emotion recognition to assist mental health

representation; the effectiveness of the proposed method is evaluated through multiple dimensions, and a comparative experimental analysis is given. The results show that the proposed mental health assessment method performs better; a complete mental health assessment system for college students has been constructed, providing an essential method for normalized large-scale psychological problem screening.

## 2. Materials and methods

### 2.1. Study procedures and participants

The Ethics Committee approved all procedures of Xuzhou Medical University, and all materials are available upon request. Furthermore, all participants were informed about this study's purpose and what was being collected, informed of their rights, and provided informed consent before data collection. The study participants were all college students at Xuzhou Medical University. The specific process is as follows: The Mental Health Center of Xuzhou Medical University publishes the recruitment information through the intranet of Xuzhou Medical University, and students who voluntarily participate in this study sign the written informed consent and enter the test through the evaluation link: <https://www.bowenzz.top:8080/>. The system supports participants' use of mobile phones or computers for testing. The evaluation content includes online questionnaires, facial emotion feature collection, and audio feature collection. The detailed data collection and modeling process are in Section Measurement instruments. Participants need to follow the prompts to complete the corresponding question and answer. The interactive questionnaire system can quickly give the measured results according to the collected multimodal information.

The experimental sample comprised 1,500 college students with the random sampling method, including undergraduates and students aged 17–34. The mean age was 19.20 years, and the standard deviation was 2.31. The proportion of female participants is 60%, and the proportion of male participants is 40%. Some students may have varying severity of mental health problems affected by managing the epidemic blockade for a long time. This study has been supported by the Key Science and Technology Program of Xuzhou: Research on key technologies of psychosocial services for normalized epidemic prevention and control.

### 2.2. Measurement instruments

#### 2.2.1. Design of data acquisition system

The questionnaire used the DASS-21 is a 21-item self-report questionnaire that assesses recent experiences of stress (e.g., "I found it hard to wind down"), anxiety (e.g., "I felt close

to panic”), and depression (e.g., “I felt that I had nothing to look forward to”). Each 7-item subscale is rated on a 4-point Likert scale ranging from 0 (Did not apply to me at all) to 3 (Applied to me very much). Higher scores represent more significant symptomatology.

An interactive questionnaire system has been designed to collect experimental sample data. While conducting human-computer interaction questionnaires, the system can collect interactive videos of participants, extract corresponding facial and audio signals and perform multimodal emotion recognition. In addition, considering personal privacy, the system anonymizes and encrypts the personal information and the collected videos.

In the modeling stage of the multimodal emotion recognition model, we mainly use neural network architectures to train corresponding prediction models, applying transfer learning and ensemble learning methods to handle different types of data, including video signals, emotion detection for audio signals, and emotion recognition for text information. Finally, the highly reliable multimodal emotion recognition detection model was constructed and completed.

### 2.2.2. Interactive scales

As shown in [Figure 1](#), the Interactive Assessment Scales (IAS) is a video and voice interactive question answering system framework that provides composite prompt function, including verbal and non-verbal cues. IAS adopts a modular architecture design consisting of three main modules: a voice interaction module, a video acquisition module, and a statistical analysis module. In the voice interaction module, the synthetic assessment scales are designed to realize the answering function in interactive scales using text-to-speech TTS technology. The video acquisition module integrates face detection, facial feature extraction, and audio emotion analysis functions. In the statistical analysis module, we design non-voice prompting functions, including natural language recognition, text emotion detection, and corresponding functions of scale information statistics. The evaluation process is as follows: First, the system uses voice interaction to describe the multiple-choice questions of the assessment questionnaire to the user. The user answers by saying one of the possible results, and each question can only be answered by choosing from a fixed set of answers. Then, the system checks whether it matches the alternative answer through natural speech recognition and will prompt the user to repeat the answer if the answer cannot be recognized. Otherwise, the answer is recognized, and the user is allowed to answer the following question after confirmation. In this process, by collecting the video and audio signals of the user’s answer, we can analyze the possibility of true and false answers according to the emotional state reflected by the user to adjust the confidence of the evaluation score to match the user’s actual psychological state as much as possible.

### 2.2.3. Multimodal emotion models

Since single-modal information influenced by various noises can not fully reflect emotional states, we proposed a multimodal hybrid deep model, which utilizes ensemble learning to train facial video and audio data of emotion recognition models. The facial video model is based on the Xception method (30), a deep convolutional neural network architecture involving depth-wise separable convolution. The audio emotion detection model is a temporally distributed convolutional neural network; its main idea is to apply a rolling window along the log-mel-spectrogram. These windows are the entry points of a convolutional neural network, and the output of each convolutional network will be fed to a 2-unit LSTM (Long Short Term Memory) to learn long-term contextual dependencies (31). The multimodal data fuses the potentially shared information of each modal and compares the emotion recognition results using an enumeration weighted decision fusion strategy based on statistical rules and probability theory.

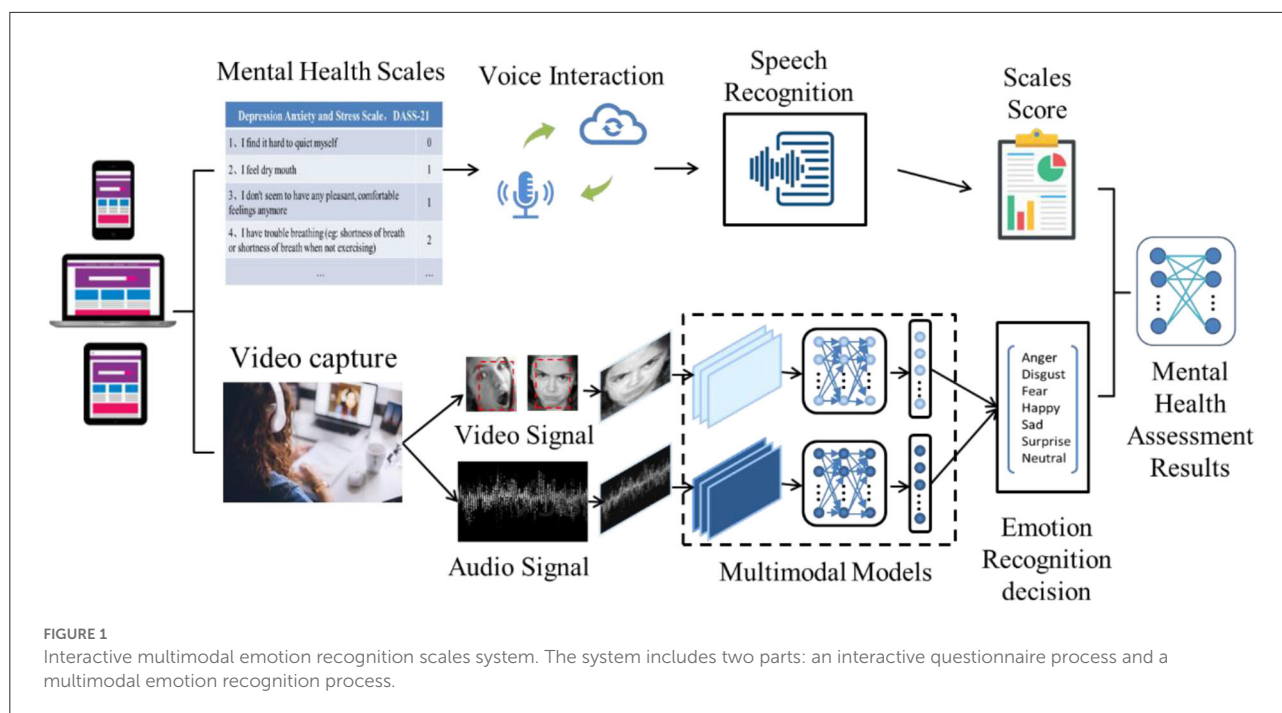
The loss function can measure the contributions of video and audio modalities, and we choose to derive a loss function that updates parameters according to the contributions of different modalities during back propagation. The derived loss function integrated with a deep neural network is defined in Equation (1), where  $l$  is a measure of the difference between the prediction and the ground truth,  $N$  is the number of modalities,  $x_i$  is the output feature of the  $n$ -th modality, and  $\omega_i$  is the fully connected network regarding the weight of the  $n$ th modality, while  $y$  is the ground truth.

$$Loss = \sum_{i=1}^N l(x_i, w_i, y) \quad (1)$$

In the classification task of multimodal emotion recognition, video and audio embedding patterns are helpful to predict real emotions. The cross-entropy loss defined in Equation (2) is chosen as the standard function.

$$l(x) = -\frac{1}{n} \sum_x [y \log x + (1-x)(1-y)] \quad (2)$$

A multimodal emotion recognition model was used to identify participants’ emotions during the assessment process. The participants were summarized in a time series according to the identified expression categories: neutral, happy, angry, surprised, sad, disgust, and fear. The questionnaire covers three mental health characteristics: stress, anxiety, and depression. These characteristics correlated with the multimodal emotions that participants exhibited during the assessment. There was also a correlation with negative emotions that participants displayed during the assessment. And the correlation between the degree of modeling analysis, as close as possible to their actual mental health status.



### 3. Results

#### 3.1. Experimental data collection and processing

The investigation began after more than a month of quarantine and lockdown at the university, and the relevant authorities approved the approval of the experiment. Participants were informed about the purpose of the experiment and what was being collected, informed of their rights, and provided informed consent before data collection. Participants provided demographic information (e.g., age and gender) and emotional state information, and data was collected and analyzed in a secure environment and under strict confidentiality. The mean age of the sample was 19.2 years, and the standard deviation was 2.31. After dis-aggregation, the proportions for each age group were 77.14% before age 20 and 22.86% over 20 years old. Among them, female participants accounted for 60%. Participants then completed an interactive DASS-21 questionnaire and corresponding video signal acquisition. In a word, this study used more complete data types, including structured and unstructured data, to ensure the accuracy and reliability of the analysis results. The DASS-21 questionnaire analyzes the data, and then the multi-modal information is input for emotional model recognition. Finally, the results of the two algorithms are combined to obtain the final evaluation result.

After students completed the online questionnaire, we performed missing value screening and deletion on the collected

data. And calculate statistics based on criteria for demographic variables. Cronbach's alphas were calculated as an indicator of internal consistency. We calculated means, standard deviations, skewness, kurtosis, and corrected item-total correlations for individual items. A three-factor model was used to test structure validity and measurement invariance of the DASS-21. Furthermore, the three one-factor models for each subscale were tested.

In the emotion recognition process, video and audio data embedding contribute to the final prediction result, and the loss function can measure the contributions of these two modalities separately. We choose to derive a loss function that updates parameters based on the contributions of different modalities during backpropagation. In classifying patients with emotion recognition, cross-entropy with logits is used, which is a commonly used method in multi-classification tasks. The cross-entropy loss is chosen as the standard function in the mental health assessment task. The multi-trait-multi-method matrix involved six traits: neutral, happy, angry, surprise, sad, disgust, and fear; two methods were involved, including the DASS-21 subscale and the multimodal emotion recognition model. Calculate the Pearson correlation coefficient between the predicted value and the DASS-21 score to obtain the correlation between different methods measuring the same trait and the correlation between different traits measured by the same method.

We pre-train the multimodal emotion recognition model with the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) and the FER2013 Kaggle Challenge

dataset (32, 33). The RAVDESS database contains 7,356 files of pure audio, audio-video and pure video, of which 1,440 are audio files. The audio included calm, happy, sad, anger, fear, surprise, and disgust, produced at two levels of emotional intensity (regular and intense), with an additional neutral tone. The FER2013 Kaggle Challenge dataset consists of 48 × 48 pixel grayscale images of human faces and contains 28,709 examples. Including seven emotions: anger, disgust, fear, happiness, sadness, surprise, and neutral.

### 3.2. Exploratory factor analysis

The results of the interactive mental health assessment are shown in Table 1. In the depression subscale, the proportion of college students who showed no symptoms, mild symptoms, moderate symptoms, severe symptoms, and extreme symptoms was 67.07, 9.96, 10.56, 2.86, and 9.56%. In the anxiety subscale, the proportions were 34.86, 17.00, 11.35, 8.90, and 27.89%. In the stress subscale, the proportions were 54.91, 8.76, 16.33, 13.55, and 6.44%.

The Cronbach alpha test results in a DASS-21 value of 0.927, a DASS-21-S value of 0.840, a DASS-21-A value of 0.803, and a DASS-21-D value of 0.835. Cronbach alpha values were >0.7 in all cases, indicating that all collected data achieved internal consistency within a large group and subgroups. Because of the non-normal distribution of anxiety, depression, and stress scores in this sample, the median was taken as the most reliable indicator of the midpoint of the sample. The median scores for stress, anxiety, and depression were 6, 4, and 2, respectively. After categorizing outcomes by the midpoint, 45.11% of participants were higher than the median stress score, 48.17% were higher than the median anxiety, and 46.18% were

higher than the median depression variable. When divided into three score scales, the mean and standard deviation of stress, anxiety, and depression factors are as follows: 7.29 (SD = 6.44), 6.52 (SD = 5.75), and 4.54 (SD = 5.47), respectively.

### 3.3. Multimodal emotion recognition

To verify the performance of the multimodal emotion recognition model, we compare the proposed model with other innovative models in video and audio emotion recognition. As shown in Figure 2, the figure shows the recognition accuracy and loss of the multimodal emotion recognition model. The model's accuracy in video and audio reaches 82.3 and 90.2%, which can fully meet the recognition requirements of real scenes.

Table 2 shows the performance of multimodal emotion recognition on RAVDESS compared to Attention-BLSTM-FCN (31), HSF-CRNN (34), and CNN TF Att.pooling (35). Furthermore, compared to KeyFrame (36), ESR-9 (37), and PREP (38) on the FER2013 dataset. In audio emotion recognition, our model outperforms the state-of-the-art models; in video emotion recognition, our model is slightly worse than ESR-9 and outperforms other models. Overall, our model has outstanding performance in video and audio emotion recognition.

Analyzing the participants' video and emotional audio signals during the answering process, it conducted a summary and preliminary analysis, as shown in Figure 3. Figure 3A shows the time series statistics of the participants' emotional changes in the questionnaire process. It can be seen from the figure that in the process of answering, most people showed a stable emotional state, while others showed relatively unstable. This

TABLE 1 Interactive DASS-21 descriptive statistics.

	Total	Age	Normal	Mild	Moderate	Server	Extreme	Confidence
<b>DASS-21</b>	1,500	19.20(2.31)						0.927
Male	600	18.99(2.10)						0.935
Female	900	19.35(2.42)						0.919
<b>DASS-21-S</b>			54.91%	8.76%	16.33%	13.55%	6.44%	0.840
Male	600	18.99(2.10)	55.56%	7.63%	15.42%	14.26%	7.13%	0.858
Female	900	19.35(2.42)	54.39%	9.57%	16.91%	13.13%	6.01%	0.825
<b>DASS-21-A</b>			34.86%	17.00%	11.35%	8.90%	27.89%	0.803
Male	600	18.99(2.10)	33.50%	16.42%	12.60%	8.96%	28.52%	0.809
Female	900	19.35(2.42)	35.71%	17.35%	10.57%	8.90%	27.47%	0.799
<b>DASS-21-D</b>			67.07%	9.96%	10.56%	2.86%	9.56%	0.835
Male	600	18.99(2.10)	62.69%	10.12%	12.11%	3.98%	11.11%	0.848
Female	900	19.35(2.42)	69.97%	9.90%	9.45%	2.11%	8.57%	0.820

DASS-21-S, means the stress of DASS-21; DASS-21-A, means the anxiety of DASS-21; DASS-21-D, means the depression of DASS-21.

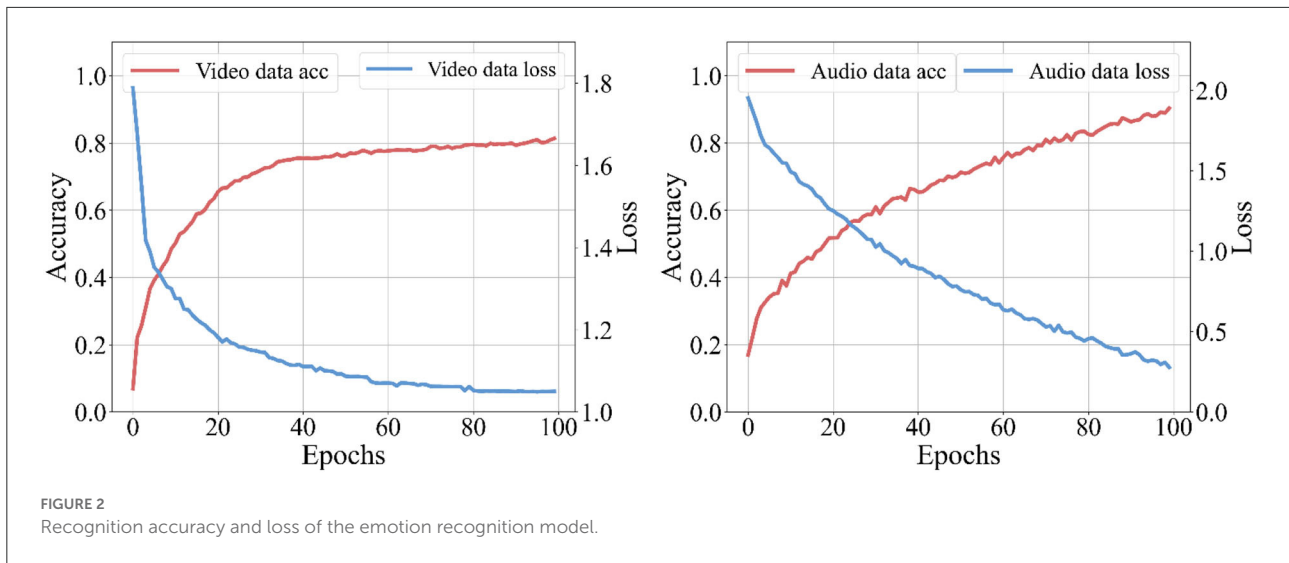


TABLE 2 Performance comparison with other models.

Features	Models	Accuracy (%)
Audio	Attention-BLSTM-FCN	72.5
	HSF-CRNN	81.3
	CNN TF Att.pooling	80.5
	Our model	90.2
Video	KeyFrame	79.1
	ESR-9	85.1
	PREP	75.7
	Our model	82.3

proportion is relatively close to that of students with severe or above various types.

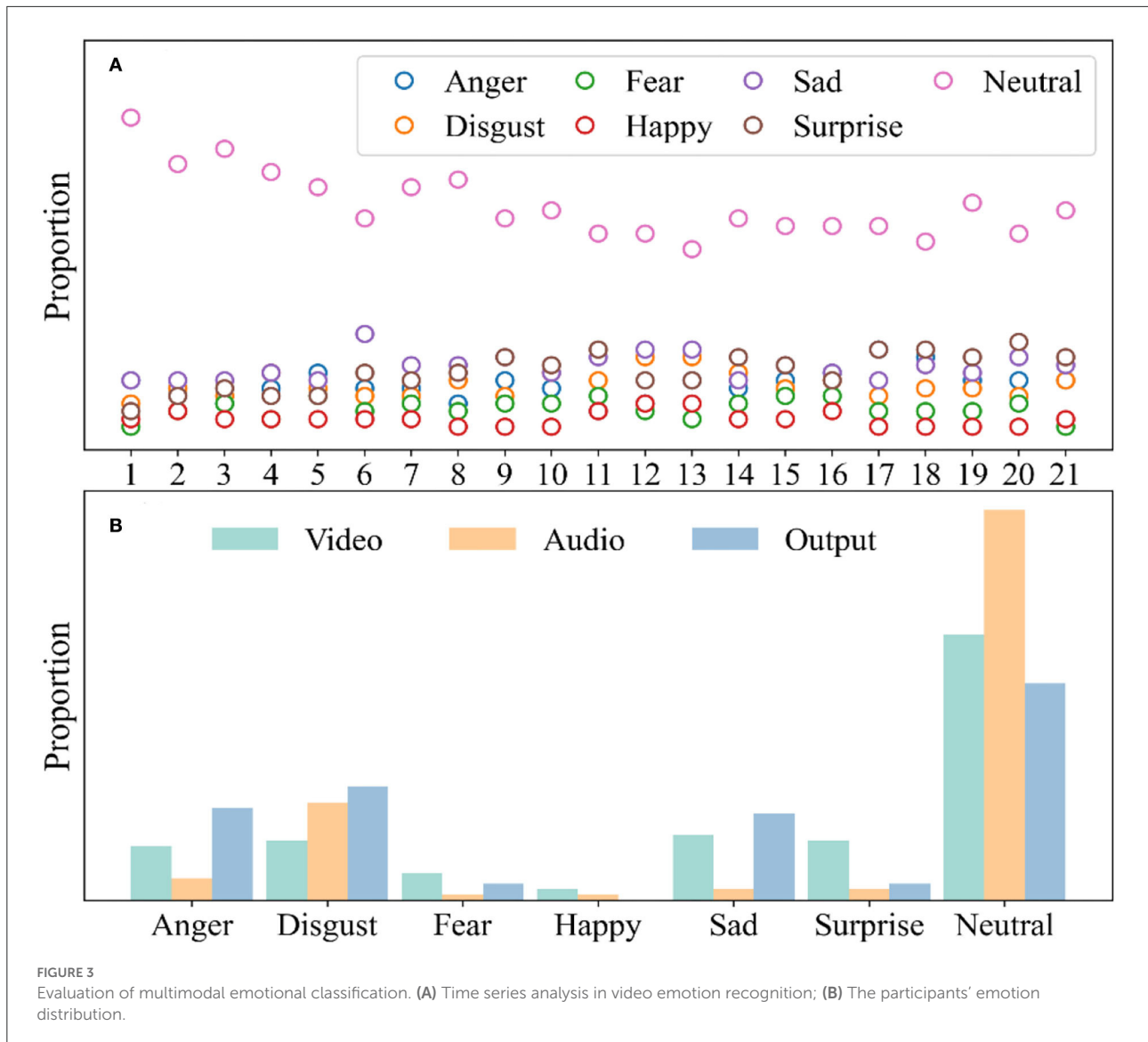
As shown in Figure 3B, the multimodal emotion recognition model identified all emotion distributions in the participant assessment process. Studies have shown that people with poor mental health are more likely to produce and express negative emotions under neutral stimuli (39). Compared with the previous scale questionnaire, most participants' emotions were captured as neutral, and only a few were captured as negative. However, if all negative emotions are counted, the proportion is relatively large. Therefore, it can be considered that there is a real connection between the emotions identified by multimodality and mental health status.

### 3.4. Correlation analysis

The Pearson correlation coefficient between the predicted value and the corresponding subscale's actual score was

calculated, and the standard validity analysis was carried out. We use a multi-feature-multi-method matrix to explore the structural validity of the multimodal sentiment prediction model. The matrix involves three traits feature attributes and seven emotion attributes of 2 methods, including the DASS-21 subscale and the emotion recognition model. The Pearson correlation coefficient between the predicted value of emotion recognition and the DASS-21 score was calculated. Figure 4 presents the zero-order correlation matrix between the variables. It is divided symmetrically on the diagonal line. Numbers indicate the correlation between emotional traits and different psychological symptoms, ranging between  $[-1, 1]$ . A positive and larger value indicates a positive correlation between attributes, and a negative and smaller value indicates a negative correlation between the attributes. At the same time, the color features in the graph are also used to represent the correlation features between the attributes. The results show that there is some correlation between the degree of mental health symptoms and multimodal emotions, especially some emotion with negative attributes. These results demonstrate that the proposed method can effectively establish the association with mental health status. However, the correlation degree of related attributes is not particularly significant, so there is still a deep correlation between various psychological symptoms and emotional features to be explored.

Based on the above analysis, the mental health of participants who exhibited negative emotions in the evaluation was corrected by establishing the relationship between the questionnaire and multimodal emotions. Figure 5 shows the evaluation results after correction based on emotion recognition and the comparison with before correction. Figure 5A shows the proportion of symptoms in the three subscales of DASS-21, including the proportion of the original scores and the proportion of the revised scores. Figure 5B shows the deviation



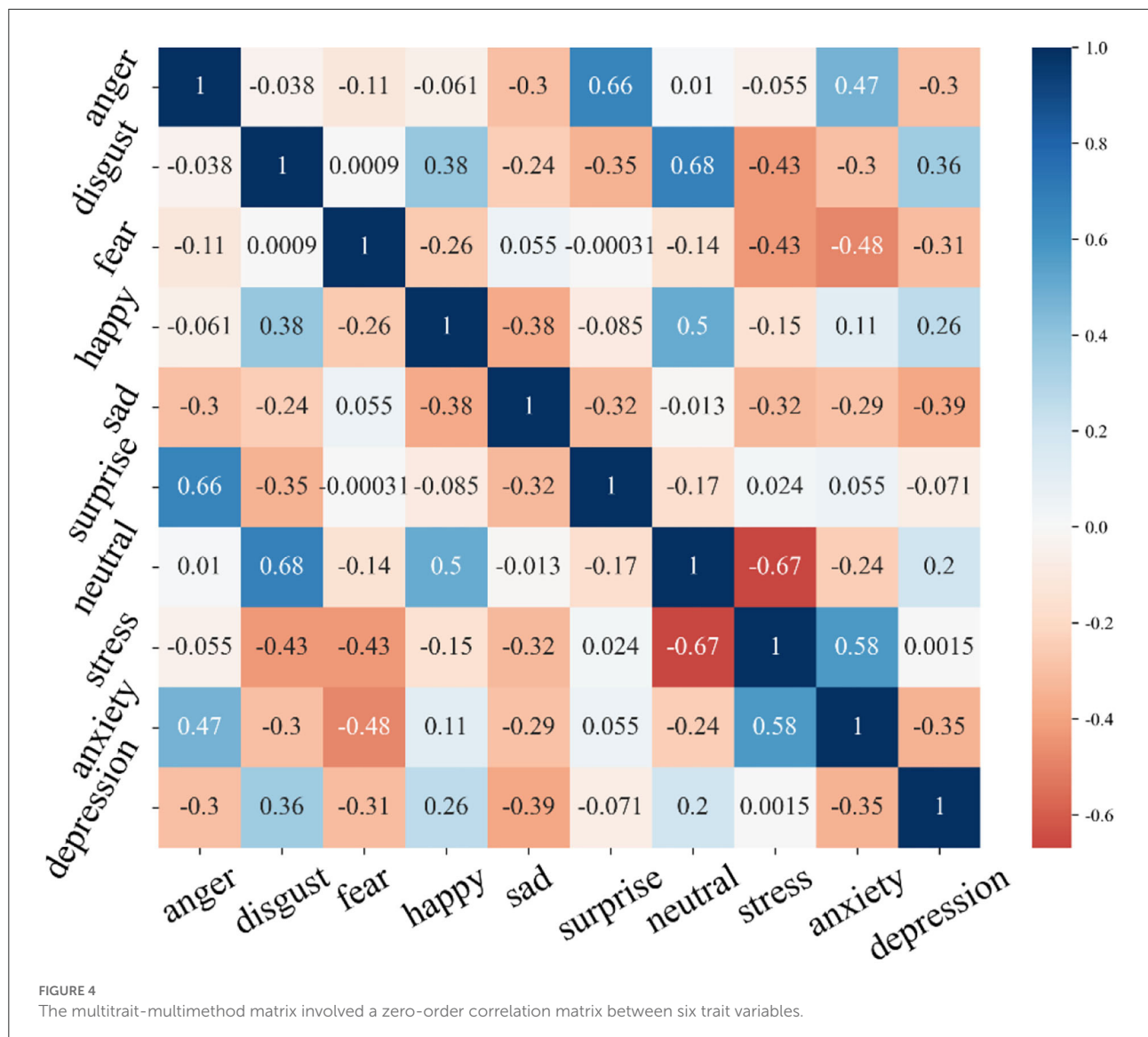
of the results of the two proportions of three symptom severity before and after correction. It can be seen that the emotion recognition correction has the most significant impact on the results of the extreme of each symptom, which is related to the fact that the extreme participants are more likely to show obvious emotional characteristics.

### 4. Discussions

The mental health problems in college students have traditionally been detected based on scores on the mental health questionnaire alone. However, researchers may be unable to accurately provide mental health counseling and treatment due to withholding facts and psychologists making misjudgments. Furthermore, previous studies have shown that the DASS-21

scales are biased in measuring the mental health of college students (40, 41). Therefore, this study combines the interactive mental health questionnaire with an intelligent algorithm-based multimodal emotion recognition model. To optimize the scale results by multimodal emotion recognition of video signals, we propose a very suitable evaluation system that adopts a multi-factor structure to measure the mental health of college students, effectively assessing the mental health degree of college students.

A reported cross-sectional survey sheds light on stress, anxiety, and depression among college students in the COVID-19 pandemic (42). Results of the study showed that anxiety factors had the lowest prevalence among students (19.7%) in long-term closed environments, while the highest prevalence of anxiety among general college students was previously reported (47.3%). In the context of the COVID-19 epidemic,

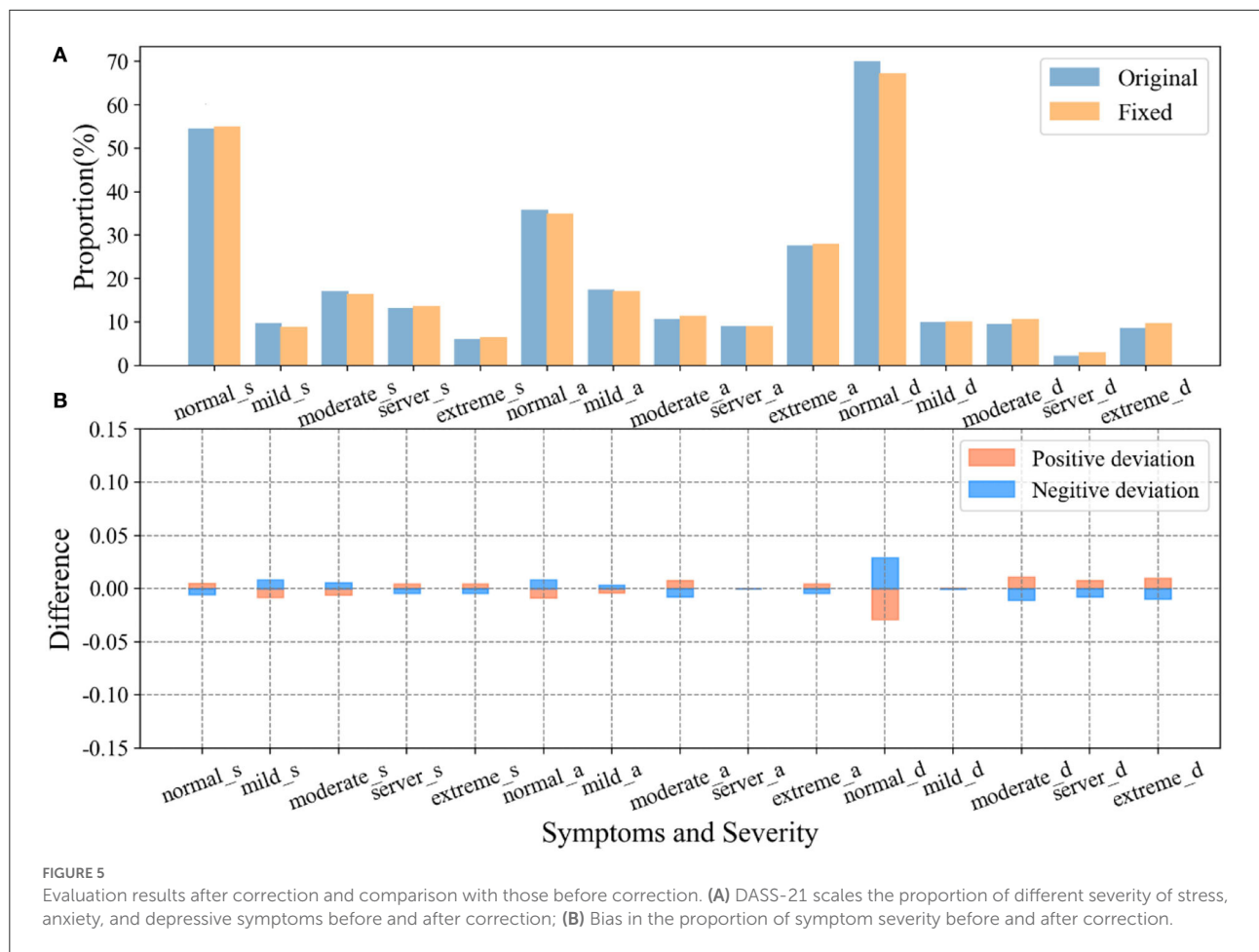


college students' direct psychological responses indicated that students had lower rates of stress but higher rates of anxiety and depression. This study's results also indicate significant differences in anxiety rates among students using the interactive multimodal emotion recognition DASS-21 scale. The student's anxiety rates increased significantly from 47.3% using the DASS-21 to 48.1% using this method, while stress and depression up to 36.3 and 23.0%, respectively. Many students may miss diagnoses because of their mental health levels when using the DASS-21. This means that the validity of the DASS-21 scale needs to be explicitly validated before it is applied to each specific setting.

This study further clarified the association between mental health and emotions. Consistent with previous research on emotion-evoking conditions, facial emotion and voice emotion can also be used to differentiate between mentally ill and

healthy individuals under neutral conditions, with those with poorer mental health more likely to produce than healthy individuals and expressing negative emotions under neutral stimuli (43). Although each model had significant standard validity, it is worth noting that the standard validity of the depression and anxiety models was lower than that of the other symptom dimensions. This is because individuals with depression and anxiety disorders may have different subtypes, leading to various emotional manifestations, resulting in slightly lower legal validity. Related studies have also pointed to differences in performance between individuals with multiple symptoms and those with only depression or anxiety disorders (44, 45). Therefore, our study shows that different psychological symptoms of mental health may have different emotional information, which can correspond to the interactive DASS





assessment process. These scores are detailed and refined. Future research needs to explore each symptom’s unique expressions of dimensions and underlying emotional mechanisms.

The study also has some limitations. First, the number of samples is insufficient to meet the needs, and some calculations based on statistical theory may be biased. Although age and gender were balanced, the participants’ shared experiences may also have contributed to some sampling bias. Second, the accuracy and applicability of the multimodal emotion recognition model need to be improved. This is mainly the choice of a machine learning algorithm to ensure that it can match the corresponding data set to ensure that the applicability will improve the results. Third, considering the purpose of the study, we used the DASS-21, whose subscales had very high correlations, which resulted in low discriminant validity. Further work should consider comorbidities between symptoms and strive to obtain corresponding unique emotions for each symptom. Furthermore, another potential use of this comprehensive model could be in predicting the risk and timing

of mental health events in graduate students. Finally, by tracking the progress of students’ emotions over time, understanding the mental health changes in different periods is one of our future research directions.

## 5. Conclusions

This study explores a multimodal emotion recognition model using an interactive assessment method based on the DASS-21 questionnaire for detecting the prevalence of anxiety and depressive symptoms in college students during the COVID-19 pandemic. The experimental results show that the interactive multimodal emotion recognition method has high reliability and satisfactory results. This study provides a feasible method for future large-scale college student mental health screening and monitoring, providing policymakers with valuable insights for targeted interventions.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving human participants were reviewed and approved by Xuzhou Medical University. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

## Author contributions

HX and XW performed the majority of the study and conducted experiments. XL reviewed the data, revised the manuscript, and designed the overview framework. All authors approved the final version of the manuscript.

## References

- Pfefferbaum B, North CS. Mental health and the Covid-19 pandemic. *N Engl J Med.* (2020) 383:510–2. doi: 10.1056/NEJMp2008017
- Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. *Lancet.* (2020) 395:912–20. doi: 10.1016/S0140-6736(20)30460-8
- Teo KY, Ching C, Chuah JH, Murphy BP, Lai KW. Early Detection of Readmission Risk for Decision Support Based on Clinical Notes. *J Med Imag Health Inform.* (2021) 11:529–34. doi: 10.1166/jmihi.2021.3304
- Talevi D, Socci V, Carai M, Carnaghi G, Faleri S, Trebbi E, et al. Mental health outcomes of the COVID-19 pandemic. *Riv Psichiatr.* (2020) 55:137–44. doi: 10.32440/ar.2020.137.03.rev03
- Zhang X, Liu X, Mi Y, Wang W, Xu H. Resilience and depressive symptoms mediated pathways from social support to suicidal ideation among undergraduates during the COVID-19 campus lockdown in China. *Psychol Res Behav Manag.* (2022) 15:2291–301. doi: 10.2147/PRBM.S377158
- Weimin Dang Y, Jun J. Study of the SCL-90 scale and changes in the Chinese norms. *Front. Psychiatry.* (2021) 11:524395–524395. doi: 10.3389/fpsy.2020.524395
- Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. *Psychiatr Ann.* (2002) 32:509–21. doi: 10.3928/0048-5713-20020901-06
- Rush MHT, Ibrahim HM, Carmody TJ, Arnow B, Klein DK, Keller MB. The 16-Item quick inventory of depressive symptomatology (QIDS) clinician rating (QIDS-C) and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biological Psychiatry.* (2003) 54:573–83. doi: 10.1016/S0006-3223(02)01866-8
- Antony MM, Bieling PJ, Cox BJ, Enns MW, Swinson RP. Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. *Psychol. Assess.* (1998) 10:176–81. doi: 10.1037/1040-3590.10.2.176
- Henry JD, John R, Crawford. The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *Br J Clin Psychol.* (2005) 4:227–39. doi: 10.1348/014466505X29657

## Funding

This work was supported by Xuzhou Science and Technology Project under Grant No. KC21306.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Lee D. The convergent, discriminant, and nomological validity of the Depression Anxiety Stress Scales-21 (DASS-21). *J Affect Disord.* (2019) 259:136–42. doi: 10.1016/j.jad.2019.06.036
- Oei TP, Sawang S, Goh YW, Mukhtar F. Using the depression anxiety stress scale 21 (DASS-21) across cultures. *Int J Psychol.* (2013) 48:1018–29. doi: 10.1080/00207594.2012.755535
- Wang K, Shi HS, Geng FL, Zou LQ, Tan SP, Wang Y, et al. Cross-cultural validation of the depression anxiety stress scale–21 in China. *Psychol Assess.* (2016) 28:e88–e100. doi: 10.1037/pas0000207
- Andreou E, Alexopoulos EC, Lionis C, Varvogli L, Gnardellis C, Chrousos GP, et al. Perceived stress scale: reliability and validity study in Greece. *Int J Environ Res Public Health.* (2011) 8:3287–98. doi: 10.3390/ijerph8083287
- Bibi A, Lin M, Zhang XC, Margraf J. Psychometric properties and measurement invariance of Depression, Anxiety and Stress Scales (DASS-21) across cultures. *Int J Psychol.* (2020) 55:916–25. doi: 10.1002/ijop.12671
- Scholten S, Velten J, Bieda A, Zhang XC, Margraf J. Testing measurement invariance of the Depression, Anxiety, and Stress Scales (DASS-21) across four countries. *Psychol. Assess.* (2017) 29:1376–90. doi: 10.1037/pas0000440
- Son C, Hegde S, Smith A, Wang X, Sasangohar F. Effects of COVID-19 on college students' mental health in the United States: interview survey study. *J Med Internet Res.* (2020) 22:e21279. doi: 10.2196/21279
- Li Y, Zhao J, Ma Z, McReynolds LS, Lin D, Chen Z, et al. Mental health among college students during the COVID-19 pandemic in China: a 2-wave longitudinal survey. *J Affect Disord.* (2021) 281:597–604. doi: 10.1016/j.jad.2020.11.109
- Lin L, Chen X, Shen Y, Zhang L. Towards automatic depression detection: a bilstm/1d cnn-based model. *Appl Sci.* (2020) 10:8701. doi: 10.3390/app10238701
- Salimi N, Gere B, Talley W, Iriogbe B. College students mental health challenges: concerns and considerations in the COVID-19 pandemic. *J Coll Stud Psychother.* (2021) 1–13. doi: 10.1080/87568225.2021.1890298
- Xu H, Peng L, Wang Z, Zeng P, Liu X. Interpersonal sensitivity on college freshmen's depression: a moderated moderation model of psychological capital and family support. *Front Psychiatry.* (2022) 13:921045. doi: 10.3389/fpsy.2022.921045

22. Søgaard Neilsen A, Wilson RL. Combining e-mental health intervention development with human computer interaction (HCI) design to enhance technology-facilitated recovery for people with depression and/or anxiety conditions: an integrative literature review. *Int J Ment Health Nurs.* (2019) 28:22–39. doi: 10.1111/inm.12527
23. Pascual AM, Valverde EC, Kim J-I, Jeong J-W, Jung Y, Kim S-H, et al. Deep learning-based facial emotion recognition for human–computer interaction applications. *Neural Comput Appl.* (2021) 1–18. doi: 10.1007/s00521-021-06012-8
24. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. *J Med Internet Res.* (2019) 21:e12869. doi: 10.2196/12869
25. Haq S, Jackson PJ. Multimodal emotion recognition. Machine audition: principles, algorithms and systems. *IGI Global.* (2011) 398–423. doi: 10.4018/978-1-61520-919-4.ch017
26. Amminger GP, Schäfer MR, Klier CM, Schlögelhofer M, Mossaheb N, Thompson A, et al. Facial and vocal affect perception in people at ultra-high risk of psychosis, first-episode schizophrenia and healthy controls. *Early Interv Psychiatry.* (2012) 6:450–4. doi: 10.1111/j.1751-7893.2012.00362.x
27. Simcock G, McLoughlin LT, De Regt T, Broadhouse KM, Beaudequin D, Lagopoulos J, et al. Associations between facial emotion recognition and mental health in early adolescence. *Int J Environ Res Public Health.* (2020) 17:330. doi: 10.3390/ijerph17010330
28. Arevian AC, Bone D, Malandrakis N, Martinez VR, Wells KB, Miklowitz DJ, et al. Clinical state tracking in serious mental illness through computational analysis of speech. *PLoS ONE.* (2020) 15:e0225695. doi: 10.1371/journal.pone.0225695
29. Santore Lee A, Gerber Alan, Gioia Ayla N, et al. Felt but not seen: Observed restricted repetitive behaviors are associated with self-report—but not parent-report—obsessive-compulsive disorder symptoms in youth with autism spectrum disorder. *Autism.* (2020) 24:983–94. doi: 10.1177/1362361320909177
30. Chollet F. *Xception: Deep Learning with Depthwise Separable Convolutions.* 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017: 1800–1807 (2017).
31. Zhao Z, Zhao Y, Bao Z, Wang H, Zhang Z, Li C. Deep Spectrum Feature Representations for Speech Emotion Recognition. In: *Session: Emotion Recognition from Multimedia ASMMC-MMAC'18.* (2018), 27–33.
32. Livingstone SR, Russo FA. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): a dynamic, multimodal set of facial and vocal expressions in North American English. *PLoS ONE.* (2018) 13:e0196391. doi: 10.1371/journal.pone.0196391
33. *The Video Dataset.* Available online at: <https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data>
34. Luo D, Zou Y, Huang D. Investigation on Joint Representation Learning for Robust Feature Extraction in Speech Emotion Recognition. In: *INTERSPEECH.* (2018) 152–6.
35. Li P, Song Y, McLoughlin IV, Guo W, Dai LR. An Attention Pooling based Representation Learning Method for Speech Emotion Recognition. *International Speech Communication Association.* (2018). Canterbury, England: University of Kent.
36. Wei J, Yang X, Dong Y. User-generated video emotion recognition based on key frames. *Multimed Tools Appl.* (2021) 80:14343–61. doi: 10.1007/s11042-020-10203-1
37. Siqueira H, Magg S, Wermter S. Efficient facial feature learning with wide ensemble-based convolutional neural networks. *Comput Vis Pattern Recogn.* (2020) 34:AAAI-20 Technical Tracks 4. doi: 10.1609/aaai.v34i04.6037
38. Kim BK, Roh J, Dong SY, Lee SY. Hierarchical committee of deep convolutional neural networks for robust facial expression recognition. *J Multimodal User Interfaces.* (2016) 10:173–89. doi: 10.1007/s12193-015-0209-0
39. Bertocci MA, Bebko GM, Mullin BC, Langenecker SA, Ladouceur CD, Almeida JRC, et al. Abnormal anterior cingulate cortical activity during emotional n-back task performance distinguishes bipolar from unipolar depressed females. *Psychol Med.* (2012) 42:1417–28. doi: 10.1017/S003329171100242X
40. Gurrola G, Balcázar P, Bonilla M, Virseda J. Depression, Anxiety and Stress Scales (DASS-21): construct validity problem in Hispanics. *Eur J Investig Health Psychol Educ.* (2020) 10:375–89. doi: 10.3390/ejihpe10010028
41. Lan HT, Long NT, Hanh NV. Validation of Depression, Anxiety and Stress Scales (DASS-21): immediate psychological responses of students in the E-learning environment. *Int J High Educ.* (2020) 9:125–33. doi: 10.5430/ijhe.v9n5p125
42. Meléndez JC, Satorres E, Reyes-Olmedo M, Delhom I, Real E, Lora Y. Emotion recognition changes in a confinement situation due to COVID-19. *J Environ Psychol.* (2020) 72: 101518. doi: 10.1016/j.jenvp.2020.101518
43. Hamaideh SH, Al-Modallal H, Tanash MA, Hamdan-Mansour 3 A. Depression, anxiety and stress among undergraduate students: a cross sectional study. *Nurs Open.* (2022) 9:1423–31. doi: 10.1002/nop.2.918
44. Gilboa-Schechtman E, Foa E, Vaknin Y, Marom S, Hermesh H. Interpersonal sensitivity and response bias in social phobia and depression: labeling emotional expressions. *Cogn Ther Res.* (2008) 32:605–18. doi: 10.1007/s10608-008-9208-8
45. Wölfling K, Beutel ME, Koch A, Dickenhorst U, Müller KW. Comorbid internet addiction in male clients of inpatient addiction rehabilitation centers: psychiatric symptoms and mental comorbidity. *J Nerv Mental Dis.* (2013) 201:934–40. doi: 10.1097/NMD.0000000000000035