A Depression Detection Auxiliary Decision System Based on Multi-Modal Feature-Level Fusion of EEG and Speech

Zhaolong Ning, *Senior Member, IEEE*, Hao Hu, Ling Yi, Zihan Qie, Amr Tolba, *Senior Member, IEEE*, and Xiaojie Wang, *Senior Member, IEEE*

***Abstract*—By improving the accuracy of depression recognition and designing a consumer-oriented depression detection system, consumers are expected to receive convenient and fast e-health services. Recently, depression recognition methods based on the analysis of physiological and behavioral data have attracted attention. In particular, the research on Electroencephalography (EEG) and speech signals becomes hotspots. However, EEG is susceptible to individual differences, while speech signal is susceptible to environmental factors. In this study, we propose an auxiliary decision-making system for depression detection that considers both physiological and behavioral factors by fusing EEG and speech signals. Compared to existing studies, our proposed multi-modal fusion strategy exploits more linear and nonlinear features to support the recognition of task classifications. In addition, we analyze the functional connectivity of brain regions to facilitate EEG feature extraction. Considering the non-stationary feature of EEG and speech signals, we perform filtering, artifact processing, and time-frequency domain processing. Furthermore, we integrate the EEG and speech signals at the feature level and train their classification. Performance evaluation results show that our proposed multi-modal feature fusion strategy achieves 86.11% accuracy on the dataset of major depressive disorders, and 87.44% recognition accuracy on the healthy controls.**

***Index Terms*—Depression detection, multi-modal fusion, feature level fusion, healthcare consumer electronics.**

I. Introduction

WITH the rapid development of consumer electronic field, information and digital technology has become an integral part of medical research and clinical practice [1]. Traditional healthcare has gradually transformed into e-health [2], mobile healthcare and smart healthcare [3]. The Internet of Medical Things can provide data resources including text, audio, and visual information [4]. Its rapid development popularizes smart healthcare and mobile healthcare, providing pervasive health monitoring services that enable remote patient monitoring and reduce the expenses associated with patient care [5]. However, despite the benefits of this convenient technology in the healthcare consumer electronic field, it has not reduced the number of people with depression.

According to the World Health Organization 2023, an estimated 3.8% of the population suffers from depression, an increasing number of 18.4% since 2012, and more than 75% of the population has no access to treatment in countries with low to moderate income [6]. Depression is a prevalent mental disorder characterized by prolonged negative moods that seriously affect people's physical and mental health [7]. Generally, physicians combine their experience with the depression self-rating scale to give a clinical diagnosis of depression. Thus, this diagnostic method is highly subjective and is limited by the experience of the physician and the cooperation of patients, making the diagnosis of depression rather challenging. In addition, this type of consultation is inconvenient for or even resisted by patients. Therefore, we intend to enable a cloud-based platform to aid in depression detection for the healthcare consumer electronic field, as shown in Fig. 1, where patients can collect their physiological and behavioral data through their health electronic devices and consult remotely through the depression detection system.

In the detection of depression, smart medical devices can provide comprehensive and portable diagnostic information by monitoring physiological indicators, voice analysis, Electroencephalogram (EEG) and so on [8]. Specifically, EEG signals contain much of the patient's physiological information that effectively responds to the implied abnormal EEG information of the cerebral cortex and can be used to diagnose and predict diseases. Simultaneously, clinical studies have shown that depressed patients usually exhibit decreased articulation rate, and monotonous voice [9]. The reason is

Manuscript received 9 February 2024; accepted 23 February 2024. Date of publication 26 February 2024; date of current version 26 April 2024. This work was supported in part by the Natural Science Foundation of China under Grant 62221005 and Grant 62272075; in part by the National Natural Science Foundation of Chongqing under Grant CSTB2022BSXM-JCX0109 and Grant CSTB2022BSXMJCX0110; in part by the Science and Technology Research Program for Chongqing Municipal Education Commission under Grant KJZDM202200601; in part by the Support Program for Overseas Students to Return to China for Entrepreneurship and Innovation under Grant cx2021003 and Grant cx2021053; and in part by the Researchers Supporting Project, King Saud University, Riyadh, Saudi Arabia, under Project RSPD2024R681. *(Corresponding authors: Ling Yi; Xiaojie Wang.)*

Zhaolong Ning, Hao Hu, Ling Yi, and Xiaojie Wang are with the School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: z.ning@ieee.org; s210101045@stu.cqupt.edu.cn; yiling@cqupt.edu.cn; xiaojie.kara.wang@ieee.org).

Zihan Qie is with the Industrial And Commercial Bank of China Baoding Municipal Branch, Hebei, China (e-mail: 343740368@qq.com).

Amr Tolba is with the Department of Computer Science, Community College, King Saud University, Riyadh 11437, Saudi Arabia (e-mail: atolba@ksu.edu.sa).

Digital Object Identifier 10.1109/TCE.2024.3370310

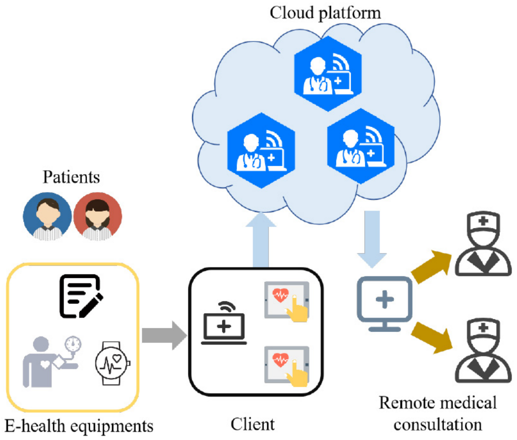


Fig. 1. An illustrative depression recognition platform.

that depression alters the body and the autonomic nervous system, which in turn potentially alters the mechanisms involved in speech production, vocalization, and articulatory musculature. Therefore, the complex nature of depression often leads to patients displaying diverse physiological or behavioral symptoms, necessitating a comprehensive analysis. Moreover, clinical symptoms or psychological analysis may not fully capture the patient's physiological changes, and more physiological and behavioral data are needed to understand the patient's behavioral attributes and psychological state comprehensively.

Given the causes of depression are intricate and complex, high-quality feature extraction and effective multi-modal data fusion strategies are key to achieving an intelligent diagnosis of depression. The challenges faced in this study are as follows:

- Highly characterized signal process and extraction: EEG signal acquisition is vulnerable to interference from noise and motion artifacts. Speech signal acquisition may also be affected by environmental noise and individual differences. Therefore, effective signal preprocessings, such as denoising, filtering, and artifact removal, are needed to improve data quality and accuracy. Meanwhile, EEG and speech signals are complex nonlinear data, and extracting useful features from them is rather challenging.
- The gathering of EEG and speech data may also lack consistency over time. Although the combination of different types of information helps understand the relationship between modalities better each modality contains specific information that goes beyond just complementing the others.

To tackle the above problems, this paper designs an auxiliary decision-making system for depression detection based on multi-modal fusion and validates it in real-world medical scenarios. Specifically, our contributions are as follows:

- In cooperation with Gansu Provincial Key Laboratory of Wearable Equipment, we collect EEG and speech data from a total of 52 subjects (23 depressed patients and 29 normal control subjects), and perform processing operations on them to obtain a high-quality dataset.
- After pre-processing the raw EEG and speech signals with noise reduction and filtering, we extract the features of the EEG and speech signals and we integrate both linear features, i.e., spectral features and Mel-Frequency Cepstral Coefficient (MFCC), and nonlinear features, i.e., Renyi entropy and C_0 complexity.
- We analyze the functional connectivity of brain regions to facilitate the feature extraction of EEG. Then, we perform a feature-level fusion of EEG and speech features, and construct multi-modal feature spaces to train the classifier. We develop a depression detection auxiliary decision system and evaluate it in real medical scenarios to verify its effectiveness. Performance evaluation results show that our proposed multi-modal method performs 11.65% better than the existing method.

The rest of the paper is organized as follows: We review some related work in Section II. Section III presents the system requirements. In Section IV, we construct a detailed system module design, and develop nonlinear feature extraction and feature-level multi-modal fusion methods. Section V details the implementation of main system modules and interfaces. Section VI tests the system systematically and functionally. Finally, we conclude our work in Section VII.

II. Related Work

Detection of mental disorders based solely on clinical symptoms of depression and pharmacological treatments often become inadequate for most patients. Establishing analytic markers from clinically distinct pathologies has proven to be challenging. The detection of psychiatric disorders and relapse through the discovery of new biomarkers requires artificial intelligence [10] and sensor-based approaches [11]. In recent years, research on the potential relationship between physiologic data and mental disorders has received much attention and has resulted in some novel diagnostic applications for mental disorders [12], [13], [14].

EEG feature extraction for independent channels or brain regions based on the time-frequency domain has been widely studied. Authors in [15] extract EEG features by exploring the spatially distributed components of different tasks. Recently, researchers have employed machine learning algorithms to identify brain disorders, and conduct feature extraction from discrete wavelet-transformed EEG using neural networks. Authors in [16] first use a continuous wavelet transform to convert EEG into rhythmic-level images. These images are then inputted into a pre-trained convolutional neural network model for detecting emotions. Authors in [17] propose a method based on Riemannian geometry and transfer learning to recognize depressed patients by extracting features of EEG signals. Authors in [18] introduced a deep learning convolutional neural network for classifying EEG data from depressed and normal subjects and achieved excellent performance.

In recent years, the development of complex network theory has inspired numerous studies, demonstrating the brain exhibits significant topological properties [19]. As a result, research exploring the use of brain networks for detecting depression has surged in popularity. A framework is proposed

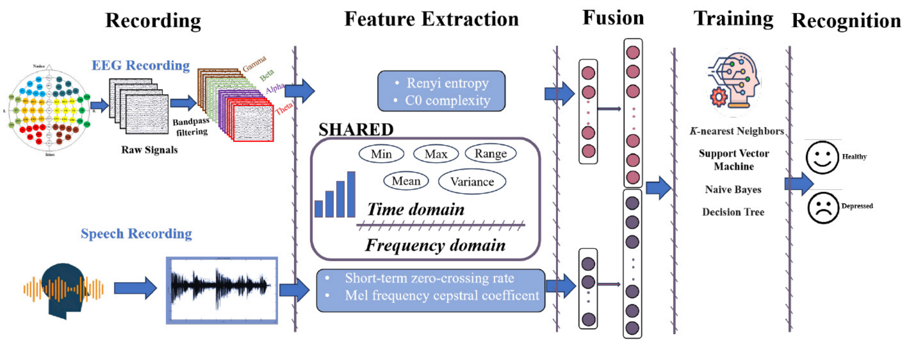


Fig. 2. The diagram of the depression detection system.

in [20] to construct brain functional connectivity matrices and brain functional networks based on resting state and Phase Lag Index (PLI). Meanwhile, speech signals can also reflect the emotional and mental state of people [21]. Authors in [22] extract MFCC features from speech to train model. Specifically, they propose a self-organizing mapping algorithm to cluster MFCC features to improve recognition accuracy.

Although a large number of studies have investigated how to diagnose depression using either, EEG or speech signals, depression diagnosis methods based on multi-modal feature fusion have not been fully developed. Although some work has been done in [23], compared with the conference version, the following work has been further studied: 1) We have added more technique details. For example, details, we have supplemented a detailed data preprocessing process, including useless electrode removal, re-referencing, and filtering of EEG signals, and pre-emphasis of speech signals. 2) We have provided more detailed feature analysis. In the feature extraction of EEG and speech signals, we consider new feature descriptions such as C0 complexity and short-time zero crossing rate. 3) In the section of functional connectivity and feature fusion, we have provided a deeper theoretical analysis, including specific studies and fusion methods. 4) We have analyzed the performance of our method on different classifiers and different frequency bands, and verified the effectiveness of our proposed method on depression detection from distinct perspectives.

III. System Requirements

The consumer-oriented healthcare system is mainly designed for depression recognition based on EEG and speech providing clinicians and professionals with an assisted decision-making tool.

The system mainly consists of data acquisition, data processing, feature extraction and selection, data classification, and result visualization. Fig. 2 shows the whole workflow of the system for depression detection. First of all, a data acquisition function is equipped to collect the EEG data and speech data of individuals before data preprocessing. After denoising, filtering and normalizing operations, the extraction of features related to depression is performed. These features include time-frequency domain features of EEG and speech, as well as complex nonlinear features that take into account of non-stationary stochastic properties. After extracting features, the system selects the relevant ones to reduce data dimensionality and improve classification abilities. The patient's depression detection results are generated by the machine learning model algorithm, and then analyzed.

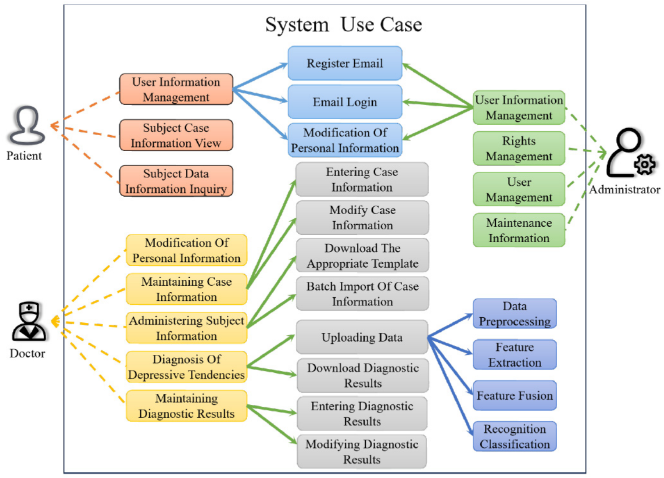

System Use Case
Patient
Doctor
Administrator
User Information Management
Subject Case Information View
Subject Data Information Inquiry
Register Email
Email Login
Modification Of Personal Information
User Information Management
Rights Management
User Management
Maintenance Information
Modification Of Personal Information
Maintaining Case Information
Administering Subject Information
Diagnosis Of Depressive Tendencies
Maintaining Diagnostic Results
Entering Case Information
Modify Case Information
Download The Appropriate Template
Batch Import Of Case Information
Uploading Data
Download Diagnostic Results
Entering Diagnostic Results
Modifying Diagnostic Results
Data Preprocessing
Feature Extraction
Feature Fusion
Recognition Classification

Fig. 3. The use case diagram.

A. Functional Requirement

There are three user roles involved in the system, including administrator, doctor and patient. Fig. 3 shows the diagram of the system use case.

1) Patient: The patient can submit his/her personal information. After the EEG and speech signals are collected, the multi-modal feature recognition can assist the doctor in diagnosis, and the patient can access personal data as well as diagnostic results.

2) Doctor: The doctor can review the depression detection results within the system and determine whether the patient exhibits depressive tendencies based on their scores. Once the doctor has comprehensively analyzed the relevant data, the final diagnosis will be submitted to the system.

3) Administrator: The administrator's privileges include both the privileges of patients and doctors. In addition, the administrator can maintain the information of the system, classify user assign privileges and so on.

B. Non-Functional Requirement

1) Security Analysis: The depression detection auxiliary decision system should protect the confidentiality, integrity and availability of user and medical data from unauthorized access and tampering [24], [25]. We introduce data encryption and backup mechanisms to improve system security. In future research, we will incorporate blockchain into the system to enhance privacy and security and utilize it to develop systems for multi-platform data sharing.

2) Usability Analysis: The system should have a high-quality user interface for ease of understanding and the ability to assist users. We design an easy-to-use pair of user interfaces that allows the system to be well-suited to user's needs for functionality and content finding.

3) Requirement Analysis: The depression detection auxiliary decision system should support large-scale data processing and high concurrent access. Additionally, the system should be flexible and scalable to allow for future upgrades and expansions.

IV. Overall System Design

The overall architecture of the depression detection auxiliary decision system adopts the model-view-controller layered design pattern, including the user layer, service layer and storage layer. The user layer is responsible for interacting with the front-end page users and visualizing the statistical information in the main interface. The service layer manages information and diagnoses diseases for module's business logic, while the storage layer stores cases and related data for tested information. The system includes a data preprocessing module, a multi-modal feature extraction module, a functional connectivity analysis module, a depression recognition classification module and a data visualization module.

A. Data Preprocessing Module

This module processes multi-modal data collected from patients with Major Depressive Disorders (MDD) and Healthy Controls (HC). The collected EEG and speech data are expected to have environmental noise and other artifacts of ophthalmic electricity and Electromyography (EMG) due to environmental and self-factors. Therefore, preprocessing work, such as filtering and denoising, is necessary prior to uploading the processed data. The timing diagram of the data preprocessing module is shown in Fig. 4.

1) Data Acquisition: We recruit patients with MDD at the Second Hospital of Lanzhou University in Gansu, China, on the recommendation of at least one clinical psychiatrist, and HC via posters. Written informed consent is obtained from all subjects prior to the start of the experiment. 52 subjects consisted of 23 outpatients (16 males and 7 females, age ranging 16-56 years old) and 29 HC (20 males and 9 females, age ranging 18-55 years old), with inclusion criteria of elementary school education or higher.

The experiment is conducted in a quiet room free of strong electromagnetic interference. All participants are asked to complete: a resting state and a point-probe task. We employ a 128-channel geodesic sensor network (250 Hz sampling frequency, impedance ≤ 50 kΩ) for EEG acquisition. All raw electrode signals are referenced to Cz. The ambient noise during voice acquisition must be less than 60 dB. The subjects are required to avoid touching the microphone and keep the distance between the mouth and the microphone about 20 cm. There are 25 recordings per subject (18 interviews, 1 short text reading and 6 word readings). The recording equipment is a Newman TLM102 microphone and an RME FIREFACE UCX audio card (44.1 kHz sampling rate, 24-bit sampling depth).

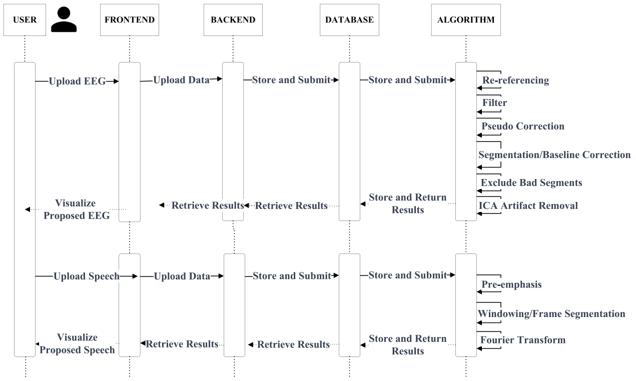


Fig. 4. The data preprocessing module timing diagram.

2) EEG Preprocessing: Predictive processing operations such as filtering, removal and reduction are performed on the acquired EEG data to eliminate noise and extract effective features. The EEG preprocessing process is as follows:

a) Rejection of bad leads or useless electrodes: Some electrodes may be improperly positioned on the scalp, which can affect subsequent analyses. In this study, bad lead data are restored through interpolation with the use of normal lead data as a reference. Any consistently marked bad leads are removed directly from the dataset.

b) Re-referencing: In this study, re-referencing is used to process EEG. The artifacts and noise caused by the reference electrode can be reduced by re-referencing the EEG signal to an average reference or other appropriate reference electrode. The Cz electrode is the selected reference electrode.

c) Filtering: Noise usually includes industrial frequency noise caused by the environment and equipment, EMG interference, and ocular electrical interference. We utilize a 50 Hz notch filter for eliminating AC interference. Muscle contraction generates high-frequency EMG signals above 100 Hz. Since this study examines EEG frequencies from 0.5 Hz to 50 Hz, a Blackman time window-based finite impulse response filter is employed to eliminate EMG-induced high-frequency band noise. In addition, a high-pass filter is used to filter out low-frequency noise.

d) Segmentation and baseline correction: A two-second time window is set to segment the EEG data to eliminate the EEG noise caused by spontaneous brain waves. Subtracting an average baseline segmentation at each point of the segmented data not only makes the signals of each segment flat, but also effectively reduces the noise.

e) Independent component analysis (ICA): Each electrode in EEG acquisition is a linear mixing of multiple independent unknown source signals. ICA performs source signal estimation on the mixed signal separation without the need for a priori information of the source signals with the mixing process.

3) Speech Preprocessing: Predictive processing operations, such as noise reduction, removal of silent parts, and partitioning, are carried out on the collected speech data. Noiseless speech helps in determining various acoustic parameters and improves the subsequent system performance. The speech preprocessing process is as follows:

a) Pre-emphasis: When a speech signal travels through an individual's vocal folds or mouth, it experiences substantial loss of high-frequency signal. To counteract this, the pre-emphasis operation typically entails passing the speech signal through a first-order high-pass filter, which is denoted as:

$$H(Z) = 1 - \frac{\alpha}{Z}, \tag{1}$$

where α represents the pre-emphasis coefficient of the speech signal, and we choose the pre-emphasis coefficient as 0.95.

b) Framing and windowing: We select 25 ms for each frame, and 10 ms for frameshifting. The Hamming window function is used to add the window and divide the frame:

$$\mathbb{W}[n] = \begin{cases} 0.54 - 0.46\cos\left[\frac{2\pi n}{N-1}\right], & 0 \le \mathrm{n} \le N-1; \\ 0, & else. \end{cases} \tag{2}$$

B. Multi-Modal Feature Extraction Module

After preprocessing the data, we extract the corresponding features of EEG and speech, and the traditional feature analysis is mainly linear, such as frequency and power spectrum. Studies have shown that both EEG and speech are non-stationary random signals [26], and limiting to simple linear analysis cannot extract the rich and relevant dynamic information of depressed patients. Therefore, this system extracts the preprocessed signals of the subjects from various aspects.

1) EEG Feature Extraction: Relevant features are extracted from the preprocessed EEG, including frequency domain features and time domain features. Considering the non-linear characteristics of EEG signals, we incorporate $C0$ complexity as an indicator, providing a comprehensive extraction of relevant features.

a) Power spectral density: A substantial body of literature indicates significant differences in the power of EEG signals between normal populations and individuals with depression. In various frequency bands of EEG, the intracranial power in the left and right hemispheres exhibits asymmetry in different populations. We use symbol N to denote the sampling point of the EEG and consider the EEG signal as an energy signal, discrete Fourier transform (DFT) can be formulated as:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-\frac{j2k\pi n}{N}}. \tag{3}$$

The power spectral density $p(k)$ is defined as:

$$p(k) = \frac{1}{N}|X[k]|^2. \tag{4}$$

b) Renyi entropy: This is a feature indicator that reflects the complexity of brain activity. Studies have shown that depression patients have lower complexity in their brain activity signals compared to HC. Unlike other entropy algorithms, it does not require the EEG signal to satisfy a Gaussian or non-smooth distribution. Expression $X = \{x_1, x_2, \ldots, x_i, \ldots, x_I\}$ denotes the set of EEG signal. Based on the EEG amplitude, it can be categorized into I subintervals. According to the corresponding amplitude range, X is assigned to the corresponding interval. Thus, the Renyi entropy is calculated as:

$$\mathbb{R} = \frac{1}{1-\alpha}\log\left(\sum_{i=1}^{I} p_i^{\alpha}\right), \alpha > 0, \alpha \neq 1, \tag{5}$$

where p_i reflects the probability of EEG in each interval, and it is evident that $\sum_{i=1}^{I} p_i = 1$.

c) C_0 Complexity: This feature represents the proportion of irregular components in the original signal and is used to characterize the complexity of the EEG. Based on the characteristics of EEG signals, the corresponding C_0 complexity is extracted by initially subjecting the signal sequence to a Fast Fourier Transform (FFT) to obtain its frequency spectrum sequence $X[k]$. Then the mean the power spectrum amplitude of $X(k)$ can be calculated as:

$$U = \frac{1}{N}\sum_{k=0}^{N-1} |X[k]|^2. \tag{6}$$

We then replace values in $X[k]$ that are less than or equal to U with 0 to obtain new sequence $Y[k]$:

$$Y[k] = \begin{cases} X[k], & |X[k]|^2 > U; \\ 0, & |X[k]|^2 \le U. \end{cases} \tag{7}$$

By performing an inverse FFT on the obtained new sequence $Y[k]$ to obtain a new time sequence $y[n]$, C_0 complexity can be calculated by:

$$C_0 = \frac{\sum_{n=0}^{N-1} |x[n] - y[n]|^2}{\sum_{n=0}^{N-1} |x[n]|^2}. \tag{8}$$

2) Speech Feature Extraction: Depression patients exhibit slow speech, monotonous intonation, and repetitive content at the speech level. Considering these characteristics, our proposed depression-assist decision system extracts relevant features from preprocessed speech data, including time-domain and frequency-domain features. The former includes energy and short-time zero crossing rates, and the latter includes spectral features and MCFFs.

a) Energy: Energy is the most important feature in speech signals, since it defines the boundary between voiced and unvoiced sounds. Short-term energy of speech signals displays a wide range of diversity. It reflects the loudness of sounds perceived by the human ear for content at various distances. Suppose that the speech is divided into N samples, where $x[n]$ represents the n-th frame of the speech. The energy of one sample is given by:

$$E_n = \sum_{m=-\infty}^{\infty} (x[n]\mathbb{W}[n-m])^2. \tag{9}$$

b) Short-time zero crossing rate: It reflects the frequency at which a signal crosses zero within a short time. Depression patients may exhibit reduced emotional expression, slower speech rate, and less fluent speech. Calculating the zero-crossing rate from the beginning of the audio signal can determine the speech components in the unvoiced part. In frames with n samples, the zero-crossing rate is obtained by:

$$\mathbb{Z} = \frac{1}{N}\sum_{n=0}^{N-1}\frac{|\operatorname{sgn}[x[n]] - \operatorname{sgn}[x[n-1]]|}{2}. \tag{10}$$

c) Spectral features: Spectral features demonstrate the frequency components and energy distribution of the speech signal by mapping the speech signal to the frequency domain. Specifically, with spectral features, we can observe the different frequency components contained in the speech signal and see the energy distribution of the speech signal at different frequencies. The spectrum can be obtained by the following equation:

$$\mathbb{X}[k] = \frac{1}{N}\sum_{n=0}^{N-1} x[n]e^{-\frac{2\pi jk}{N}}. \tag{11}$$

d) MFCC: MFCC is frequently utilized as a speech feature due to its alignment with human hearing and low-frequency properties. The technique portrays the nonlinear qualities of human auditory frequency perception, and its correlation with sound frequency can be computed by [27]:

$$\mathbb{M}(f) = 2595lg\left(1 + \frac{f}{700}\right), \tag{12}$$

where $\mathbb{M}$ represents that the frequency in Mel-scale and f represents the original frequency. Consider the Mel-scale-based triangular filter and let symbol q denotes the index of triangle filters. Its frequency response $H_q(k)$ is expressed as:

$$H_q[k] = \begin{cases} 0, & k < f(q-1); \\ \frac{2(k-f(q-1))}{(f(q+1)\ -f(q-1))(f(q)-f(q-1))}, & f(q-1)\ \le k \le f(q); \\ \frac{2f(q+1)\ +k)}{(f(q+1)\ +f(q+1)(q-1)\ -f(q+1))}, & f(q) < k \le f(q+1); \\ 0, & k > f(q+1). \end{cases} \tag{13}$$

The logarithmic energy for the output of each filter bank can be calculated by:

$$s(q) = ln\left(\sum_{k=0}^{N-1} H_q(k)|X(k)|^2\right). \tag{14}$$

By applying the obtained logarithmic energies to the discrete cosine transform, we have:

$$C(l) = \sum_{q=0}^{Q-1} s(q)\cos\frac{\pi l(q-0.5)}{Q}, l = 1, 2, \ldots, L, \tag{15}$$

where L represents the order of MFCCs. The width of the triangular filter can be determined based on the distribution of the center frequency on the Mel-Frequency scale, to determine the number of triangular filters Q to be 26.

C. Functional Connectivity Analysis Module

For high-level cognitive functions in the brain, a single brain region cannot perform all the functions and it requires the collaboration of different brain regions. Therefore, it is crucial to examine the functional connectivity of various brain regions to comprehend the mechanisms that govern cognitive processes [28].

Due to the presence of volume conduction effects, PLI is chosen as the method to measure the synchrony between two channel signals since it is insensitive to these effects. At time t, the phase difference between any pair of EEG signals can be formulated by:

$$\left|\Delta\varphi_{i,j}(t)\right| = \left|n\varphi_i(t) - m\varphi_j(t)\right|, \tag{16}$$

where $\varphi_i(t)$ and $\varphi_j(t)$ represent the instantaneous phases of channel signals $X_i(t)$ and $X_j(t)$, respectively. Take i as an example, it can be calculated by:

$$\varphi_i(t) = \arctan\frac{\tilde{x}_i(t)}{x_i(t)}, \tag{17}$$

where $\tilde{x}_i(t)$ is the Hilbert transform of $x_i(t)$, and its calculation formula is:

$$\tilde{x}_i(t) = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{x_i(\tau)}{t-\tau}d\tau. \tag{18}$$

The PLI between two channels is derived from the following formula:

$$\mathbb{P}_{ij} = \left|\frac{1}{Q}\sum_{n=1}^{N-1} sgn\left(\Delta\varphi_{ij}(n)\right)\right|, \tag{19}$$

where $0 \le \mathbb{P}_{ij} \le 1$. A higher value indicates a stronger degree of phase synchronization between two signals.

The system displays nodes for both HC and MDD patients located in distinct brain regions. The coordinates of each node correspond to the center of its respective brain region, with node size relative to its strength value.

D. Depression Recognition Classification Module

A multi-modal-based classification module is a core part for recognizing depression disorders. In this module, machine learning algorithms are utilized to input fused features into a pre-trained model for recognizing depression, with the results stored in a database for display in the front-end scheduling.

Most current studies, such as [16], [29], [30], only focus on unimodal feature extraction for recognition and classification tasks. However, unimodal is susceptible to interference from individual differences, thus leading to a reduction in classification accuracy [31]. To improve traditional depression detection, we propose a feature fusion of multi-modal information by effectively introducing both EEG and speech signals, creating a multi-modal feature space. Given that MFCCs can effectively identify different features in speech signal recognition and the cepstral feature class plays a critical role in identifying different subjects, this study employs feature splicing to construct a multi-modal feature space using both EEG features and cepstral feature MFCC splicing, as illustrated in Fig. 5. Herein, symbols x_i and y_j denote EEG

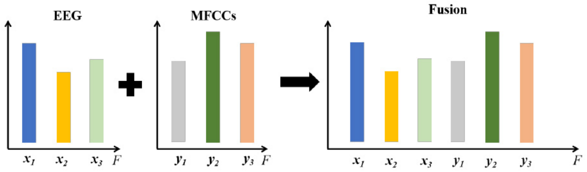


Fig. 5. The multi-modal feature-level fusion strategy.

and MFCC features, respectively, and F denotes the number of features.

Dataset S consists of σ samples, where EEG features have α dimensions, speech features have β dimensions, and class labels are denoted as z_i. The formal representation of the dataset is as follows:

$$S = \{x_i, y_j, z_i\}, \tag{20}$$

where $|S| = \sigma, x_i \in R^{\alpha}, z_i \in R^{\beta}, z_i \in \{0, 1\}$.

By concatenating the features to integrate their modal information, new feature data can be obtained as follows:

$$S' = \left\{r_i, z_i'\right\}, \tag{21}$$

where $|S'| = \sigma, r_i \in R^{\alpha+\beta}, z_i' \in \{0, 1\}$, and $r_i = [x_i, y_j]$. The feature space of S' comprises not only the physiological information of the subject's profile such as EEG features, but also their behavioral information, i.e., MFCC features.

E. Data Visualization Module

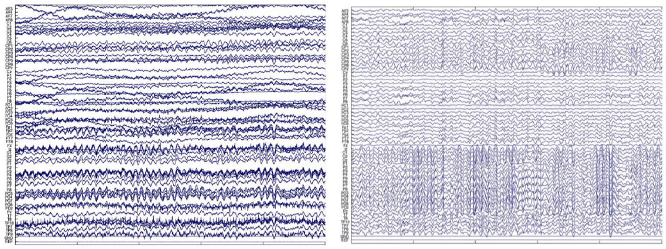
The data visualization module should offer a straightforward interface to enable smooth user interaction, specifically in inputting EEG and speech data for recognition, as well as access results and confidence levels. Comprehensive visual data are presented in the main panel, such as recently evaluated cases, prevalence of depression disorders, and EEG statistics. In the recently tested information list, administrators and doctors can conveniently view patient information by filtering and sorting data. The page provides detailed personal information about the patient, including his/her name, gender, and specific case information. The data are entered into the system after a joint diagnosis based on the results of the back-end classification of depressive disorders.

V. System Implementation

In this section, we present the implementation of the main system modules and interfaces.

A. Data Preprocessing Module

The system preprocesses the multi-modal data collected from individuals with MDD and HC. Environmental and self-factors result in the presence of environmental noise, ocular and electromyographic artifacts in the collected EEG and speech data. It is necessary to carry out preprocessing work, such as filtering and denoising, to remove these artifacts before uploading the processed data. The initial EEG graph is displayed in Fig. 6(a). The EEG data are processed using the data preprocessing module, which calls relevant back-end programs. The processed EEG graph is displayed in Fig. 6(b).



(a) The original EEG (b) The preprocessed EEG

Fig. 6. Pre-processing of EEG signals.

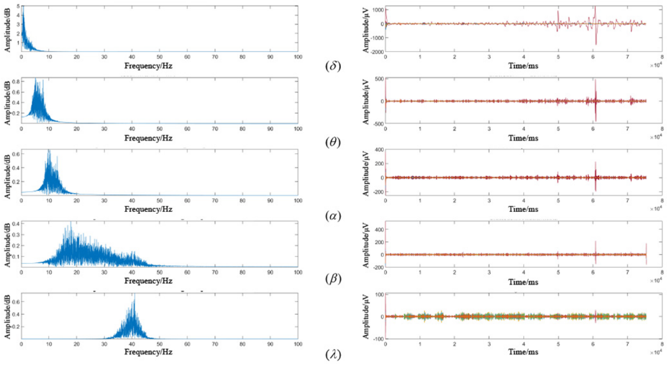


Fig. 7. The EEG spectrogram in different frequency bands after filtering.

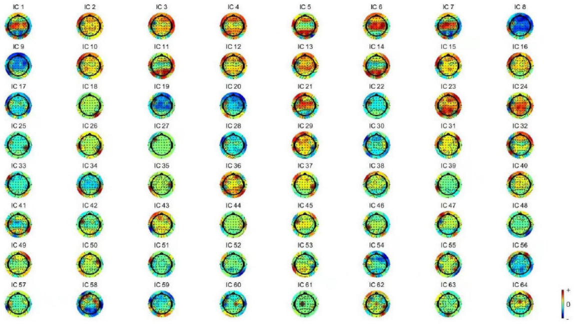

IC 1 IC 2 IC 3 IC 4 IC 5 IC 6 IC 7 IC 8
IC 9 IC 10 IC 11 IC 12 IC 13 IC 14 IC 15 IC 16
IC 17 IC 18 IC 19 IC 20 IC 21 IC 22 IC 23 IC 24
IC 25 IC 26 IC 27 IC 28 IC 29 IC 30 IC 31 IC 32
IC 33 IC 34 IC 35 IC 36 IC 37 IC 38 IC 39 IC 40
IC 41 IC 42 IC 43 IC 44 IC 45 IC 46 IC 47 IC 48
IC 49 IC 50 IC 51 IC 52 IC 53 IC 54 IC 55 IC 56
IC 57 IC 58 IC 59 IC 60 IC 61 IC 62 IC 63 IC 64

Fig. 8. The EEG data obtained with 64 channels after ICA processing.

B. Multi-Modal Feature Extraction Module

This module extracts EEG and speech features from preprocessed data and conducts a comprehensive analysis of multi-modal datasets acquired from subjects. This module depicts spectrograms for each band post-EEG filtering, as shown in Fig. 7, and Fig. 8 displays the 64-channel EEG data information obtained after ICA. The spectrogram of EEG can be used to extract the spectral features of EEG to obtain the energy distribution of EEG at different frequencies. With MFCC processing, we can extract the frequency features in the speech signal and reduce the dimensionality of different features to improve the computation efficiency of the system, as shown in Fig. 9.

C. Functional Connectivity Analysis Module

This module visualizes the functional connectivity of brain regions in different frequency bands in depressed and normal populations. It maps nodes in different left and right brain

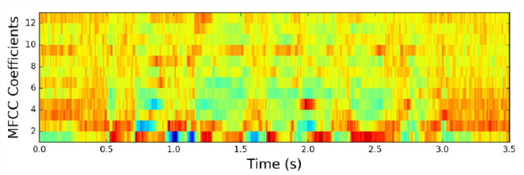
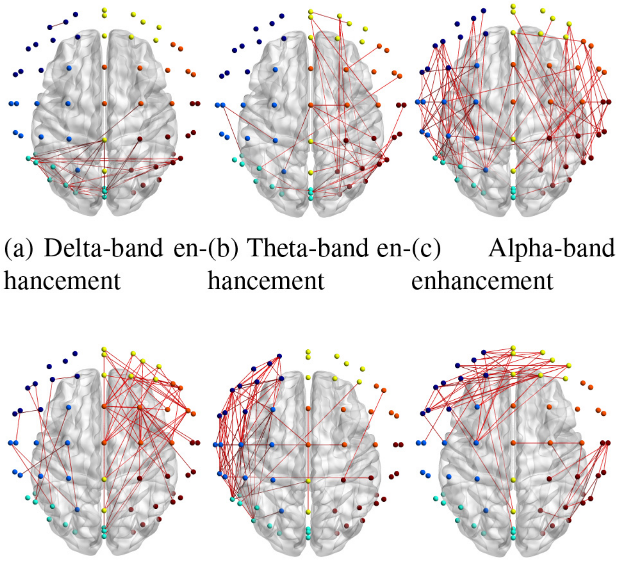


Fig. 9. The MFCCs spectrum.



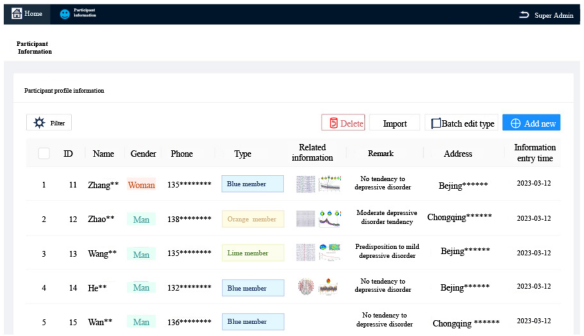
(a) Delta-band enhancement (b) Theta-band enhancement (c) Alpha-band enhancement

(d) Delta-band reduction (e) Theta-band reduction (f) Alpha-band reduction

Fig. 10. Functional connectivity analysis of brain regions.

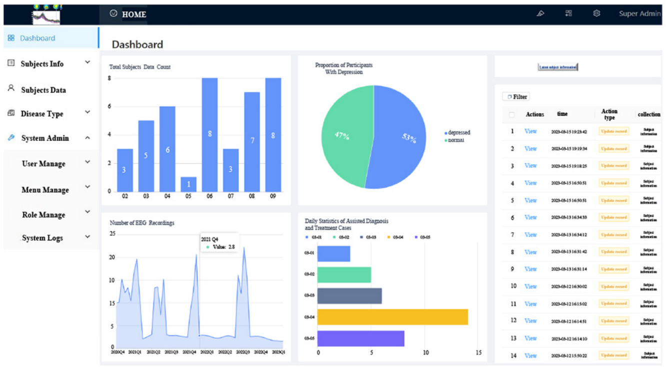
regions in healthy populations and patients with depressive disorders, and investigates the spatial distribution of these functional connectivity features with the categorization ability. Increased and decreased functional connectivity for MDD patients relative to HC is located between the relevant brain regions, allowing for an intuitive analysis of the identification of depressive disorders. The visualization of functional connectivity analysis of brain regions is shown in Fig. 10.

Figs. 10(a) and 10(d) represent the parts of the brain with increased and decreased functional connectivity of MDD relative to HC in the Delta frequency band, respectively. It can be observed that the enhanced parts are mainly concentrated in the parietal and occipital parts, and the decreased parts are mainly in the right frontal lobe. In Figs. 10(b) and 10(e), we can observe that on the Theta band, the increase in brain functional connectivity is mainly in the right occipital lobe, while the decrease in the functional connectivity is mainly in the left frontal lobe. Similarly, Figs. 10(c) and 10(f) show that the regions of increased functional connectivity in the brain on the Alpha band are in the frontal and parietal lobes and are symmetrically distributed. The regions of the reduced functional connectivity are concentrated in the left frontal and right parietal lobes. By observing and analyzing the functional connectivity of the brain as described above, it is known that the functional connectivity between the left frontal lobe and the whole parietal lobe of MDD patients is abnormal.



□	ID	Name	Gender	Phone	Type	Related information	Remark	Address	Information entry time
1	11	Zhang**	Woman	135********	Blue member		No tendency to depressive disorder	Bejing******	2023-03-12
2	12	Zhao**	Man	138********	Orange member		Moderate depressive disorder tendency	Chongqing******	2023-03-12
3	13	Wang**	Man	135********	Lime member		Predisposition to mild depressive disorder	Bejing******	2023-03-12
4	14	He**	Man	132********	Blue member		No tendency to depressive disorder	Bejing******	2023-03-12
5	15	Wan**	Man	136********	Blue member		No tendency to depressive disorder	Chongqing ******	2023-03-12

Fig. 11. The results of depression recognition and classification.



□	Actions	time	Action type	collection
1	View	2023-03-15 19:23:42	Update record	Subject information
2	View	2023-03-15 19:19:34	Update record	Subject information
3	View	2023-03-15 19:18:25	Update record	Subject information
4	View	2023-03-15 16:50:51	Update record	Subject information
5	View	2023-03-15 16:50:51	Update record	Subject information
6	View	2023-03-13 16:34:33	Update record	Subject information
7	View	2023-03-13 16:34:12	Update record	Subject information
8	View	2023-03-13 16:31:42	Update record	Subject information
9	View	2023-03-13 16:31:14	Update record	Subject information
10	View	2023-03-12 16:30:02	Update record	Subject information
11	View	2023-03-12 16:15:02	Update record	Subject information
12	View	2023-03-12 16:14:51	Update record	Subject information
13	View	2023-03-12 16:14:10	Update record	Subject information
14	View	2023-03-12 15:50:22	Update record	Subject information

Fig. 12. The data visualization module.

D. Depression Recognition Classification Module

The main function of this module is to use machine learning algorithms to identify and classify the fused features, and deliver them to the next module for visualization. Specifically, the fusion feature method is first used to fuse the features to form higher-dimensional feature vectors. Then the classifier is trained and tested using the dataset, and Fig. 11 shows some results of depression recognition and classification. By uploading the subject's relevant data, a classifier is used to obtain auxiliary results, and then a professional doctor can make a diagnosis of the subject by combining the classification results.

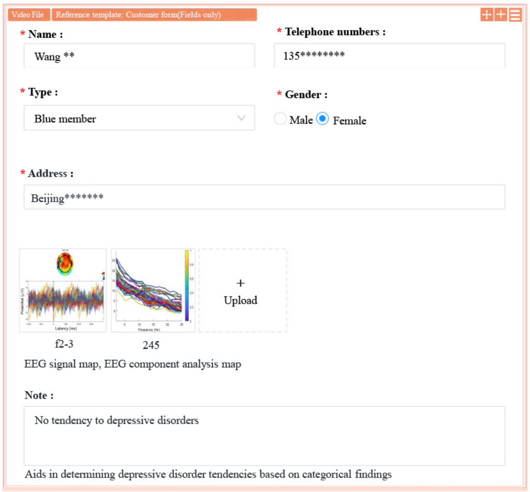
E. Data Visualization Module

The data visualization module is shown in Fig. 12. After administrators and doctors enter the system, they can view the visualized data information, such as the total number of recently tested information statistics, the percentage of people with depression disorder classification, and EEG data statistics in the main panel. On the data visualization page, one can click different buttons to view other visual information.

The subject's diagnosis results are shown in Fig. 13. It provides specific details regarding the patient's personal information, such as their name, gender, home address, and other personal information. Additionally, it contains comprehensive case information featuring depression disorder identification classification results, which are entered into the system by the administrator and the doctor after the joint diagnosis of the patient. The case information section features

TABLE I
FEATURE NUMBERS AND CLASSIFICATION RESULTS OF CLASSIFIERS IN DIFFERENT FREQUENCY BANDS

Classifier	SVM		KNN		DT		NB	
	accuracy	features	accuracy	features	accuracy	features	accuracy	features
full(1-40Hz)	81.50	50	85.18	239	83.60	15	82.66	50
delta(1-4Hz)	82.69	26	83.44	255	81.09	20	79.40	20
theta(4-8Hz)	78.57	55	79.32	267	79.11	20	80.93	225
alpha(8-13Hz)	77.92	15	82.20	600	77.45	600	78.21	25
beta(13-30Hz)	80.09	20	82.13	725	80.91	25	79.70	20



* Name : Wang **

* Telephone numbers : 135********

* Type : Blue member

* Gender : ○ Male ● Female

* Address : Beijing*******

f2-3 245

EEG signal map, EEG component analysis map

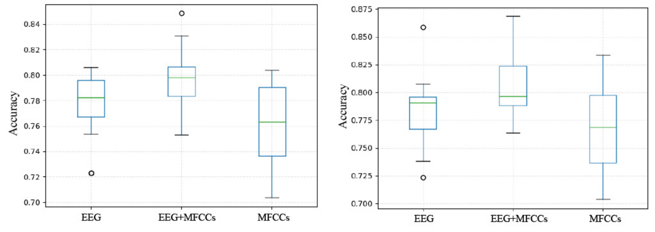
Note : No tendency to depressive disorders

Aids in determining depressive disorder tendencies based on categorical findings

Fig. 13. The subject diagnostic results interface.

TABLE II
DETECTION ACCURACY OF EACH CLASSIFIER UNDER MULTI-MODAL CONDITIONS

Classifier	KNN		SVM		DT		NB	
	HC	MDD	HC	MDD	HC	MDD	HC	MDD
EEG	77.20	79.30	82.85	81.17	69.96	71.15	71.73	69.00
MFCC	78.05	76.27	78.32	79.95	70.40	68.05	72.56	67.57
EEG+MFCC	79.80	79.82	87.44	86.11	76.02	76.20	76.80	74.93



(a) HC (b) MDD

Fig. 14. The recognition accuracy with KNN classifier

detailed notes, related diagnosis materials, and classification results.

VI. SYSTEM TESTING

We perform evaluations on EEG and speech signals of 23 depressed patients and 29 healthy controls. EEG signals are divided into 5 sets, i.e., full band, delta band, theta band, alpha band and beta band. The classifiers are Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Decision Trees (DT), and Naive Bayes (NB), where the hyperparameter K of KNN is set to 1 to 5, the soft interval parameter under the kernel of SVM is set as $C = [10^{-3}, 10^{-2}, \dots, 10^{3}]$. The validation is executed by the nested cross-validation method. The used dataset is the public MODMA dataset from Lanzhou University (http://uais.lzu.edu.cn/? p=4068).

A. Algorithm Testing

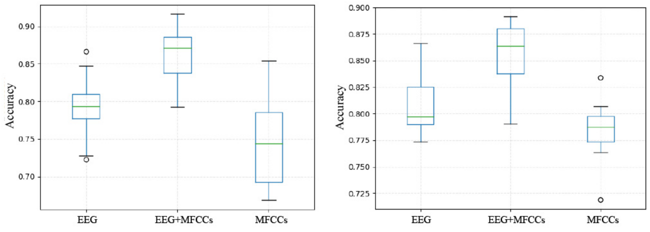
To verify the effectiveness of the depression recognition system, a comparison is conducted regarding the disparity in depression recognition accuracy between single–modality and multi-modality. The use of various classifiers under different frequency bands of the EEG in varying classification accuracies is indicated in Table I.

When selecting 255 features with the largest correlation coefficients from the PLI functional connectivity in the Delta band for classification, the SVM classifier achieves an accuracy of 83.44%. In the Alpha and Beta bands, the KNN classifier yields the highest recognition accuracy, specifically 82.20% and 82.13%, respectively. The highest classification accuracy, at 85.18%, is achieved by implementing the SVM classifier to select 239 features with the largest correlation coefficients from the PLI functional connectivity across the full frequency band. We can observe that the KNN classifier performs well and stably in different frequency bands in terms of classification accuracy, and performs the best in the full band, delta band, alpha band and beta band. For each classifier model, cross-validation should be used in the training set to select hyper-parameters. Once hyper-parameters are set, the outer loop of nested cross-validation tests the classification accuracy and feature layer fusion of the individual classifiers for different modalities. Table II displays the recognition accuracy after feature-level fusion.

As depicted in Fig. 14, combining EEG and MFCCs features under the KNN classifier results in better elevated accuracy than single one feature. As shown in Fig. 15, combining EEG and MFCCs features with the SVM classifier increases accuracy, with higher mean intervals indicating MDD in recognition results. Fig. 16 shows that the method fusing EEG and MFCC features with the DT classifier exhibits the best performance on both HC and MDD datasets. As shown in Fig. 17, compared to the NB classifier, the performance of the DT classifier is more stable.

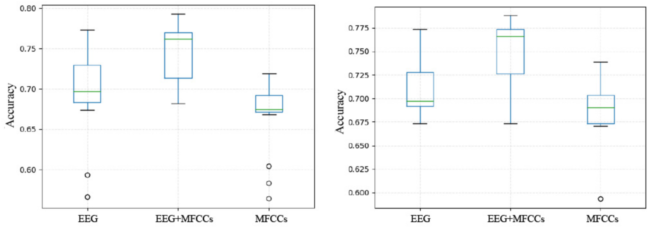
TABLE III
TEST CASES OF DATA AND DEVICE MANAGEMENT MODULE

NO.	Description	Steps	Expected results	Test results
01	Data import	After entering the information to be tested according to the template, upload the file and click the "Start Import" button	Display a successful import message and show the imported data on the tested information page	Pass
02	Result entry	Modify the tested data information and re-upload the EEG analysis graph	Display a success message upon modification and show the newly uploaded image on the tested information page	Pass
03	Result entry	Modify the disease type of the tested data	Display a success message upon modification and show the updated disease type	Pass
04	Result entry	Modify the provided assisted diagnosis results in the comments	Indicates that the modification is successful, and the comment displays the corrected results	Pass
05	Data visualization	Click on "Visualization of Assisted Diagnosis Count"	Display a bar chart visualization of the assisted diagnosis count	Pass
06	Data visualization	Click on "Visualization of the Total Number of Participants"	Display a bar chart visualization of the total number of participants	Pass
07	Data visualization	Click on "Visualization of the Classification Ratio for Depression in Participant Data"	Display a pie chart visualization showing the classification ratio for depression in participant data	Pass
08	Data visualization	Click on "Visualization of EEG Data Statistics and Record Counts"	Display visualization of EEG data statistics and record counts	Pass



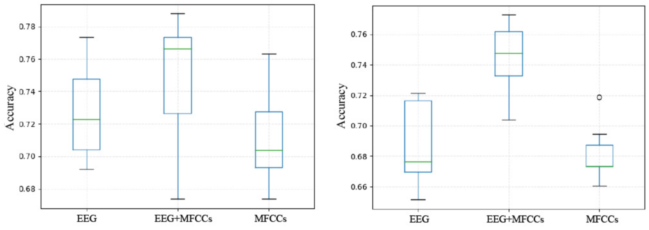
(a) HC (b) MDD

Fig. 15. The recognition accuracy with SVM classifier.



(a) HC (b) MDD

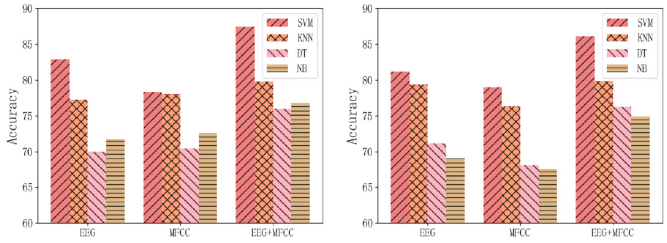
Fig. 16. The recognition accuracy with DT classifier.



(a) HC (b) MDD

Fig. 17. The recognition accuracy with NB classifier.

Fig. 18 shows the performance comparison among our proposed method and the unimodal methods with different classifiers. It can be seen that the classification accuracy of our algorithms under different classifiers is better than the EEG feature and speech feature methods. Also, the performance of each method is the best with the SVM classifier, followed by KNN. It also shows that the methods of fusing features can compensate for their respective shortcomings to obtain better results.



(a) HC (b) MDD

Fig. 18. Comparison of classification accuracy with unimodal and multi-modal feature methods.

The fusion of multi-modalities not only improves model stability but also enhances performance due to the introduction of modal information. In particular, using SVM to learn EEG + MFCC fusion features leads to the highest HC and MDD classification accuracy, with recognition rates of 87.44% and 86.11%, respectively.

B. Functional Testing

Functional test cases are shown in Table III. Its purpose is to verify that the system meets the expected requirements. During the process, developers or users compare the system's operations and results through manual or automated testing, and assess the system's suitability, completeness, and operability according to user needs.

VII. Conclusion

In this paper, we propose a multi-modal depression recognition algorithm based on EEG and speech, and implement the design of a depression recognition system, to provide an effective detection method for depression. First, we eliminated noise and interference from the data using a well-designed preprocessing technique that effectively ensures the accuracy and reliability of the input data. Meanwhile, the introduction of functional connectivity analysis helps to reveal depression-related brain activities, thus enriching our recognition model. Then, we propose a multi-modal feature-level fusion method based on EEG and speech to comprehensively explore the deep features latent in physiological and behavioral signals. Compared to a single type of data, the proposed feature fusion method, facilitates the model to obtain higher recognition accuracy. Finally, we develop a functional depression detection auxiliary decision system to aid in the diagnosis of depression, which demonstrates good efficacy and feasibility in real-world healthcare environments.

References

[1] Y. Liu, X. Wang, G. Zheng, X. Wan, and Z. Ning, "An AoI-aware data transmission algorithm in blockchain-based intelligent healthcare systems," *IEEE Trans. Consum. Electron.*, early access, Feb. 13, 2023, doi: 10.1109/TCE.2024.3365198.

[2] Y. Zhang, C. Xu, H. Li, K. Yang, J. Zhou, and X. Lin, "HealthDep: An efficient and secure deduplication scheme for cloud-assisted eHealth systems," *IEEE Trans. Ind. Informat.*, vol. 14, no. 9, pp. 4101–4112, Sep. 2018.

[3] V. P. Yanambaka, S. P. Mohanty, E. Kougianos, and D. Puthal, "PMsec: Physical unclonable function-based robust and lightweight authentication in the Internet of Medical Things," *IEEE Trans. Consum. Electron.*, vol. 65, no. 3, pp. 388–397, Aug. 2019.

[4] T. K. Reddy, V. Gupta, and L. Behera, "Autoencoding convolutional representations for real-time eye-gaze detection," in *Proc. Comput. Intell., Theor., Appl. Future Direct.*, 2019, pp. 229–238.

[5] W. Wang et al., "Realizing the potential of the Internet of Things for smart tourism with 5G and AI," *IEEE Netw.*, vol. 34, no. 6, pp. 295–301, Dec. 2020.

[6] *Depressive Disorder (Depression)*, World Health Org., Geneva, Switzerland, 2023.

[7] J. Li, Y. Hao, W. Zhang, X. Li, and B. Hu, "Effective connectivity based EEG revealing the inhibitory deficits for distracting stimuli in major depression disorders," *IEEE Trans. Affect. Comput.*, vol. 14, no. 1, pp. 694–705, Mar. 2023.

[8] W. Wang et al., "Cross-modality LGE-CMR segmentation using image-to-image translation based data augmentation," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 20, no. 4, pp. 2367–2375, Aug. 2023.

[9] R. Horwitz, T. F. Quatieri, B. S. Helfer, B. Yu, J. R. Williamson, and J. Mundt, "On the relative importance of vocal source, system, and prosody in human depression," in *Proc. IEEE Int. Conf. Body Sensor Netw.*, 2013, pp. 1–6.

[10] V. Arora, L. Behera, T. K. Reddy, and A. P. Yadav, "HJB equation based learning scheme for neural networks," in *Proc. Int. Joint Conf. Neural Netw.*, 2017, pp. 2298–2305.

[11] Y. Zhu, Y. Shang, Z. Shao, and G. Guo, "Automated depression diagnosis based on deep networks to encode facial appearance and dynamics," *IEEE Trans. Affect. Comput.*, vol. 9, no. 4, pp. 578–584, Dec. 2017.

[12] G. A. Prabhakar, B. Basel, A. Dutta, and C. V. Rama Rao, "Multichannel CNN-BLSTM architecture for speech emotion recognition system by fusion of magnitude and phase spectral features using DCCA for consumer applications," *IEEE Trans. Consum. Electron.*, vol. 69, no. 2, pp. 226–235, May 2023.

[13] J.-S. Park, J.-H. Kim, and Y.-H. Oh, "Feature vector classification based speech emotion recognition for service robots," *IEEE Trans. Consum. Electron.*, vol. 55, no. 3, pp. 1590–1596, Aug. 2009.

[14] V. Singh and T. K. Reddy, "EEG-based reaction time prediction with fuzzy common spatial patterns and phase cohesion using deep autoencoder based data fusion," in *Proc. IEEE 4th Annu. Flagship India Council Int. Subsect. Conf.*, 2023, pp. 1–5.

[15] V. Mishuhina and X. Jiang, "Feature weighting and Regularization of common spatial patterns in EEG-based motor imagery BCI," *IEEE Signal Process. Lett.*, vol. 25, no. 6, pp. 783–787, Jun. 2018.

[16] F. Demir, N. Sobahi, S. Siuly, and A. Sengur, "Exploring deep learning features for automatic classification of human emotion using EEG rhythms," *IEEE Sensors J.*, vol. 21, no. 13, pp. 14923–14930, Jul. 2021.

[17] S. D. Reddy, S. Goyal, and T. K. Reddy, "Riemannian approach based depression classification using transfer learning for MEG signals," in *Proc. IEEE 4th Annu. Flagship India Council Int. Subsect. Conf.*, 2023, pp. 1–4.

[18] A. Seal, R. Bajpai, J. Agnihotri, A. Yazidi, E. Herrera-Viedma, and O. Krejcar, "DeprNet: A deep convolution neural network framework for detecting depression using EEG," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, Jan. 2021.

[19] Y. He and A. Evans, "Graph theoretical modeling of brain connectivity," *Current Opinion Neurol.*, vol. 23, no. 4, pp. 341–350, 2010.

[20] B. Zhang, G. Yan, Z. Yang, Y. Su, J. Wang, and T. Lei, "Brain functional networks based on resting-state EEG data for major depressive disorder analysis and classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 215–229, Dec. 2020.

[21] R. Chatterjee, S. Mazumdar, R. S. Sherratt, R. Halder, T. Maitra, and D. Giri, "Real-time speech emotion analysis for smart home assistants," *IEEE Trans. Consum. Electron.*, vol. 67, no. 1, pp. 68–76, Feb. 2021.

[22] Z. Huang, J. Epps, D. Joachim, and V. Sethu, "Natural language processing methods for acoustic and landmark event-based features in speech-based depression detection," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 2, pp. 435–448, Feb. 2020.

[23] X. Wang, X. Wan, Z. Ning, Z. Qie, J. Li, and Y. Xiao, "A multimodal fusion depression recognition assisted decision-making system based on EEG and speech signals," in *Proc. IEEE Int. Conf. Commun., Comput., Cybersecurity, Inform.*, 2023, pp. 1–8.

[24] X. Wang, Z. Ning, L. Guo, S. Guo, X. Gao, and G. Wang, "Mean-field learning for edge computing in mobile blockchain networks," *IEEE Trans. Mobile Comput.*, vol. 22, no. 10, pp. 5978–5994, Oct. 2023.

[25] Z. Ning et al., "Blockchain-enabled intelligent transportation systems: A distributed crowdsensing framework," *IEEE Trans. Mobile Comput.*, vol. 21, no. 12, pp. 4201–4217, Dec. 2022.

[26] T. K. Reddy, Y.-K. Wang, C.-T. Lin, and J. Andreu-Perez, "Joint approximate diagonalization divergence based scheme for EEG drowsiness detection brain computer interfaces," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2021, pp. 1–6.

[27] X. Mu and C.-H. Min, "MFCC as features for speaker classification using machine learning," in *Proc. IEEE World AI IoT Congr.*, 2023, pp. 566–570.

[28] X. Shao et al., "Analysis of functional brain network in MDD based on improved empirical mode decomposition with resting state EEG data," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1546–1556, Jun. 2021.

[29] J. Shen et al., "Exploring the intrinsic features of EEG signals via empirical mode decomposition for depression recognition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 356–365, Nov. 2022.

[30] J. Shen et al., "An optimal channel selection for EEG-based depression detection via kernel-target alignment," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 7, pp. 2545–2556, Jul. 2021.

[31] X. Zhang et al., "Emotion recognition from multimodal physiological signals using a regularized deep fusion of kernel machine," *IEEE Trans. Cybern.*, vol. 51, no. 9, pp. 4386–4399, Sep. 2021.