Auxiliary Diagnosis of Depression Based on Multi-modal Data: A Survey

I. Introduction

Depression is one of the most common mental disorders [1]–[3], affecting more than 300 million people worldwide [4]. Persistent grief, losing interest in things that are usually enjoyed, and the inability to participate in daily activities are characteristics of depression [5]. Additionally, it will make patients feel poorly and lose their living ability [6]. Worse still, patients with severe symptoms have thoughts of self-harm or suicide [7]–[9]. It is estimated that close to 70% of depressed patients have had suicidal thoughts, and over 50% have self-harmed [10]. In short, the dangers of depression are staggering.

To alleviate the damage caused by depression, a timely and effective diagnosis is essential. However, there are three main challenges in the clinical diagnosis of depression: (1) Lack of valid and accurate objective indicators. Unlike other physiological diseases, there are no objective and precise markers to assess depression, such as biochemical test indicators, physiological data, or medical imaging. Clinicians draw conclusions based on the clinical diagnostic criteria, their professional experience, and the patient's clinical presentation [11]. However, many patients purposely hide their actual illness while refusing to participate in treatment, which leads to skewed diagnostic outcomes. (2) The severe shortage of medical resources. There is currently a severe shortage of medical professionals and medical resources for diagnosing and treating depression. Up to 75% of patients, especially in low- and middle-income countries, do not obtain early diagnosis and treatment. Additionally, ineffective and time-consuming face-to-face consultations with psychiatrists add to the lack of resources for psychiatric care [12], [13]. (3) Deficiencies in social awareness. Because the general public has low awareness of depression, there is an underlying resistance to or avoidance of mental illness. Given this, early signs of depression may be dismissed or disregarded, losing the ideal window for treatment, which will make the condition worse and make therapy more challenging and complex [14]–[17].

To address the difficulty of clinical diagnosis of depression, in recent years, researchers have proposed massive machine learning approaches. With the help of machine learning, researchers extract practical features from depression data automatically and tap the intrinsic connection between depression clinical symptoms and physiological signals, which has become an essential direction for auxiliary depression diagnosis [18]. The following advantages exist for machine learning-based auxiliary depression diagnosis: (1) High efficiency: by using computer technology to analyze the physiological signals of depressed patients and create objective diagnostic indicators, clinical diagnosis can be made much more quickly and effectively. (2) Low interference: the machine learning method is used for automatically auxiliary depression diagnosis, which eliminates the need for professionals to observe consciousness and lessens the difficulties of diagnosing and treating depression owing to patient resistance.

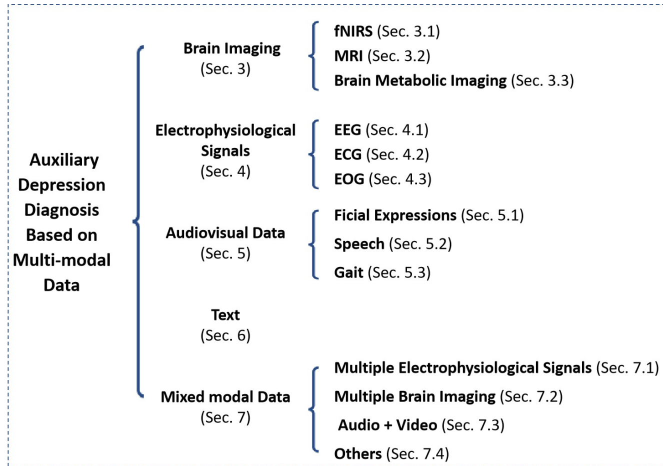


Fig. 1. The taxonomy of ancillary diagnosis of depression from the perspective of various categories of data.

Despite the rapid evolution in this field, there is yet no systematic study to review and discuss existing progress. To fill this gap, a thorough analysis of recent ancillary depression diagnosis studies about the use of machine learning on different categories of data is conducted. Next, the research methodology and procedures of machine learning in ancillary depression diagnosis are outlined, and finally, the research directions and challenges for the future are presented. As shown in Fig 1, we focus on the progress of the methods and potential research directions in five data contexts: *brain imaging* [19]–[23], *electrophysiological signals* [24]–[28], *Audiovisual data* [29]–[31], *text* [32]–[34], and *mixed-modal data* [35]. These commonly used representation forms of physiological signals are illustrated in Fig 2, and the following are brief descriptions of them:

(1) *Brain imaging:* Brain imaging refers to the usually non-invasive or minimally invasive techniques that enable imaging the structure or function of the brain [36]. It is achieved by scanning the subject's brain with various precision instruments. Functional Near-Infrared Spectroscopy (fNIRS), nuclear Magnetic Resonance Imaging (MRI), etc., are its primary data representations. Studies have shown that, to some extent, brain imaging can identify different types of depression depending on the part of the brain affected [37].

(2) *Electrophysiological signals:* Electrophysiological sig-

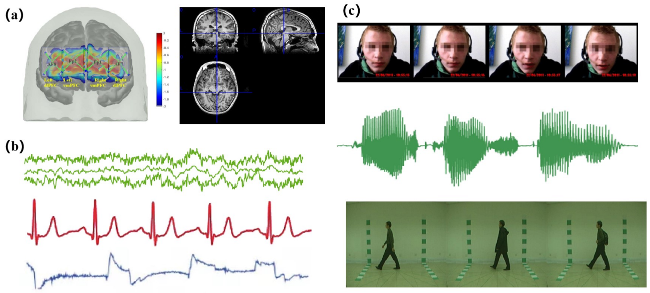


Fig. 2. Common types of physiological signals: subgraphs $a-c$ represent brain imaging,electrophysiological signals, and audiovisual data, respectively. These categories in turn have different representations: fNIRS, MRI, EEG, ECG, EOG, facial expression, speech, and gait.

nals are caused by changes in the membrane potential of individual cells [38], which are usually recorded by metal electrodes placed on the body surface. Electroencephalography (EEG, tracing brain tissue activities), electrocardiography (ECG, recordings of the cardiac movement), and electrooculogram (EOG, examining the behavior of eyes) are common electrophysiological signals relevant to clinical interest. They have been widely utilized to treat auxiliary depression diagnosis due to their non-invasive detection and simplicity of use.

(3) *Audiovisual data:* Audiovisual data, whose forms mainly include facial expressions, speech, and gait, are captured by video and audio recording devices. Audiovisual data-based auxiliary depression diagnosis is the first recognition method proposed by researchers and the most widely used method, which has the advantage of low acquisition cost but can achieve better recognition results.

(4) *Text:* Text is the most direct vehicle for people to express their thoughts and emotions. Especially on social media platforms, people are more inclined to express their true feelings. Therefore, numerous academics have looked at the disparities in textual expression between depressed people and those without depression in the setting of social media platforms from a textual perspective [39].

(5) *Mixed-modal data:* In addition to using single physiological or behavioural data to assist in the diagnosis of depression, many studies have sought to employ mixed-modal data for auxiliary depression diagnosis [40]–[42]. Mixed-modal data have more features than unimodal data, which can give a more complete picture of the symptomatology of depressed patients.

In this paper, we review the current state of research on machine learning in auxiliary depression diagnosis regarding the five types mentioned above of data, with a focus on the progress of the use of machine learning techniques in different data contexts and potential research directions. The main contributions of this study are threefold.

(1) To the best of our knowledge, this is the first comprehensive survey about auxiliary depression diagnosis on different modalities of data, which will provide researchers and clinical psychologists with a better understanding of the use of machine learning In this area.

(2) We provide an in-depth review of advanced auxiliary depression diagnosis studies, summarize and compare the performance of of different methods in each data modality.

(3) We summarize and analyze the advantages and disadvantages of different data types for auxiliary depression diagnosis and give insights on promising research directions in this field.

The rest of this survey will be organized as follows: Section 2 presents a summary of standard approaches to diagnosing clinical depression and abnormal changes in the physiology and behaviour of depressed patients and introduces widely-used questionnaires, datasets, and metrics. Section 3-7 provides an in-depth analysis of machine learning-based approaches to depression diagnosis from the perspective of different data types. Section 8 provides insights into promising research directions. We conclude the survey in Section 9.

II. BACKGROUND

This section presents some basic knowledge in auxiliary depression diagnosis task including scales, datasets and assessment metrics which are important to be known ahead. Firstly, introduces some scales of diagnosing depression which are used as labels. Secondly, shows some public datasets with different modal. Finally, lists some commonly used assessment metrics, which are important in evaluating the performance of methods.

A. Depression Diagnosis Scale

As the most commonly used tools, a series of questionnaires capturing depressive symptoms had been developed and showed great success in psychiatric practice. The common depression test scales are shown in table I.

For each scale, there are several items with 3-4 options which rating to different scores. Subjects match one option they deem appropriate in each item and the different range of overall score is divided to different degree of depression. For example, the BDI-II scale is a 4-point scale which containing 21 items about sleep, appetite, mood, and suicidal thoughts. Each item has 4 options and is rated ranging from 0 to 3. The highest overall score is 63, and higher scores indicate more severe depressive symptoms. Subjects will match one option of each item by the feelings in the past two weeks. After finishing the BDI-II scale, subject will get the degree of depression by overall score they matched.

The scales are divided into other-rating scales and self-rating scales by the way of testing scales. For subjects, testing other-rating scales needs communication with psychiatrist, psychologist or other trained medical personnel at a professional place like hospital or clinic, which is more reliable but less convenient. Self-rating scale like BDI-II is a kind of self-assessment under the Instruction manual, which can be easy complete in 5-10 minutes but not quite reliable.

The scale is tried to test the emotions of subjects, because negative emotions are the main performance of depressed people. However, people with negative emotions are not the same as with depression. Measuring depression on a scale alone has limitations, so that physiological electrical signals or brain images from subjects for auxiliary depression diagnosis during professional diagnosis. In recent years, some research showed that abnormal behavioral signals like facial expressions, voice and gait can reflect mental disease, which is

TABLE I
COMMONLY USED DEPRESSION DIAGNOSIS SCALES

Name	Items	Scores	Grading	Usage
Beck Depression Inventory (BDI) [43]	21	0-63	0-9 minimal, 10-18 mild, 19-29 moderate, 30-63 severe	self- rating
Beck Depression Inventory-II (BDI-II) [44]	21	0-63	0-13 minimal, 14-19 mild, 20-28 moderate, 29-63 severe	
Self-rating depression scale (SDS) [45]	20	20-80	20-53 minimal, 53-62 mild, 63-72 moderate, 73-80 severe	
Patient Health Questionnaire (PHQ-9) [46]	9	0-27	0-5 minimal, 5-9 mild, 10-14 moderate, 15-19 medium severe, 20-27 severe	
Center for Epidemiological Survey, Depression Scale (CES-D) [47]	20	0-60	0-16 minimal, 16-20 mild, 21-25 moderate, 26-60 severe	
Geriatric Depression Scale (GDS) [48]	30	0-30	0-10 minimal, 11-20 mild, 21-30 moderate	
Hamilton Depression Scale (HAMD) [49]	17 /21 **/24**	0-34 /0-42 **/0-48**	0-8 minimal, 8-20 mild, 20-35 moderate, 35-48 severe	others-rating
Montgomery-Asberg Depression Rating Scale (MADRS) [50]	10	0-60	0-12 minimal, 12-21 mild, 22-29 moderate, 30-34 medium severe, 35-60 severe	
International Classification of Diseases (ICD-10) [51]	-	-	Clinical diagnosis criteria	
Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) [52]	-	-		
Chinese Classification and Diagnostic Criteria of Mental Disorders (CCMD-3) [53]	-	-		

There are three versions of the HAMD scale, 17, 21 and 24-item, and only the 24-item version is described here in detail.
Clinical diagnosis criteria: Different criteria are listed for different numbers of symptoms, and doctors confirm the diagnosis of the condition based on the number of symptom matches and the duration of the patient's symptoms.

more easily to observe compared with physiological electrical signals or brain images. For example, depressed people have dull facial expressions, rarely make eye contact, use phrases with a flat tone when speaking, and walk slowly. By analyzing the abnormal behavioral signals can be well in auxiliary diagnosing depression.

B. Datasets

This section introduces some public datasets as shown in Table II, including the name, information of subject and data mode. Due to privacy issues, only several of the modal data has public dataset.

The Reddit Self-reported Depression Diagnosis (RSDD) dataset consists of Reddit posts for 9,210 users who have claimed to have been diagnosed with depression ("diagnosed users") and 107,274 matched control users, which is a large-scale general forum dataset. The data is in JSON format, and each row is an array representing a user. The label field includes the user's label (control or depression), and the posts field contains timestamp and untokenized pairs.

The eRsik dataset consists of text examples collected from messages of 887 Reddit users. The main idea of the task is to classify users into risk case of depression and non-risk case.

Multi-modal Open Dataset for Mental-disorder Analysis (MODMA) consists data mainly from clinically depressed patients and matching normal controls, which is a multi-modal open dataset for mental-disorder analysis. All patients in MODMA were carefully diagnosed and selected by professional psychiatrists in hospitals. EEG and speech recording data are made publicly available in MODMA. The EEG signals were recorded as both in resting state and under stimulation, and EEG dataset includes not only data collected using traditional 128-electrodes mounted elastic cap, but also a novel wearable 3-electrode EEG collector for pervasive applications. The speech data were recorded as during interviewing, reading and picture description.

Audio-Visual Emotion Challenge (AVEC) is an expression recognition challenge held every year since 2011, which is recognized as the top international competition in the field of emotional computing. AVEC2013 began to introduce the task of depression recognition, which considers the analysis of depression based on auditory vision as a classification problem or regression problem.

Audio-Visual Depression Corpus (AViD-Corpus) contains 340 video clips of subjects performing a Human-Computer Interaction task while being recorded by a webcam and a microphone. The speakers were recorded between one and four times, with a period of two weeks between the measurements. This data was used for the AVEC2013 and AVEC2014 Challenge.

Distress Analysis Interview Corpus – Wizard of Oz (DAIC-WOZ) contains clinical interviews designed to support the diagnosis of psychological distress conditions such as anxiety, depression, and post-traumatic stress disorder. Data collected include audio and video recordings and extensive questionnaire responses [54]. This data was used for the AVEC2016 and AVEC2017 Challenge.

Extended DAIC Database(E-DAIC) is the extended version of DAIC-WOZ database for depression and PTSD assessment, developed by ICT [54]. This data was used for the AVEC2019 Challenge.

Emotional Audio-Textual Depression Corpus (EATD-Corpus) contains audios and extracted transcripts of responses from 162 depressed and non-depressed volunteers. EATD-Corpus is the first and only public depression dataset that includes audio and text data in Chinese.

C. Metrics

In many studies, depression detection was seen as a binary classification problem. Subject with the minimal degree of the scale were labeled to have non-depressed and others were labeled to have depressed. And the experimental evaluation methods were some common classification metrics such as precision, accuracy, recall, F1-score, specificity, sensitivity and Area Under the Curve (AUC).

Some studies used regression analysis to predict the degree of depression according to the score of scale, and the experimental evaluation methods were some common regression metrics as: mean absolute error (MAE) and root mean squared error (RMSE).

1) Classification indicators: Confusion metrics is the basic of classification metrics. And confusion metrics can be split to 4 parts: True Positive (TP), False Positive (FP), False Negative (FN) and True Negative (TN). TP means a positive sample and predicted as positive. False Positive FP means a positive sample but predicted as negative. False Negative FN means a negative sample but predicted as positive. True Negative TN means a negative sample and predicted as negative. Therefore, the total positive samples can be calculated as $TP + FP$, and the total negative samples can be calculated as $FN + TN$. The overall of confusion metrics is shown in Fig 3.

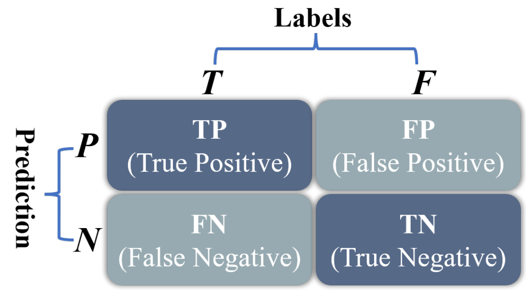


Fig. 3. Confusion Matrix: each row represents the forecast category, and the total number of each row represents the number of data predicted for that category. Each column represents the real category of data, and the total data in each column represents the number of data instances of this category.

Precision reflects the ability of the model to correctly predict positive samples, it calculates how many positive samples are predict as *positive*. The calculation formula as follows:

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

Accuracy reflects the ability of the model to correctly predict the overall sample, it calculates how many positive samples are predict as *positive* and negative samples as *negative*. The calculation formula as follows:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} = \frac{TP + TN}{P + N} \tag{2}$$

Recall reflects the ability of the model to correctly predict the full degree of positive samples, it calculates the proportion of positive samples predicted as *positive* to the total positive samples. The calculation formula as follows:

$$Recall = \frac{TP}{TP + FN} = \frac{TP}{P} \tag{3}$$

F1-score is a statistical measure of the accuracy of model. It defined as the harmonic mean of precision and recall. The formula as follows:

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{4}$$

Specificity reflects the ability of the model to correctly predict the full extent of negative samples, it calculates the proportion of negative samples predicted to be *negative* to the total negative samples. The calculation formula as follows:

$$Specificity = \frac{TN}{FP + TN} = \frac{TN}{N} \tag{5}$$

Sensitivity reflects the ability of the model to correctly predict the full extent of positive samples, it calculates the proportion of positive samples predicted to be *positive* to the total positive samples. The calculation formula as follows:

$$Sensitivity = \frac{TP}{TP + FN} = \frac{TP}{P} \tag{6}$$

TABLE II
COMMONLY USED DATASETS

Dataset	Subjects(D/H)	Type	Task
RSDD [55]	116484 (9210:107274)	Depression	Text (Social Network)
eRisk [56]	887	Depression	
MODMA [57]	53 (24/29) *EEG* 52 (23/29) *Audio*	Mental-disorder	EEG+ Audio
AVEC2013/2014 & AViD-Corpus [58]	100 (46/54)	Depression	Audio+ Facial expressions
AVEC2016/2017 & DAIC-WOZ [59]	193 (60/133)	Depression	
AVEC2019 & E-DAIC [60]	275	Depression	
EATD-Corpus [61]	162 (30/132)	Depression	Audio+ Text

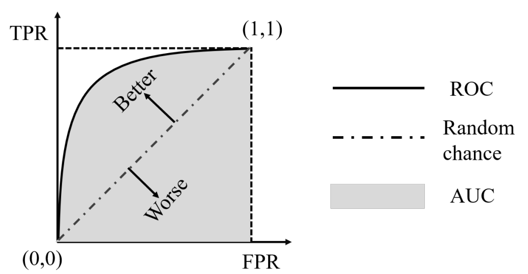


Fig. 4. ROC curve is obtained based on confusion matrix. Its abscissa is FPR, and ordinate is TPR. Draw the coordinates of the same model (FPR, TPR) in ROC space, and it becomes the ROC curve. AUC is defined as the area under the ROC curve.

ROC represents a probability graph to show the performance of a classification model at different threshold levels. The curve is plotted between two parameters, which are: True Positive Rate(TPR) and False Positive Rate (FPR). TPR is a synonym for Recall, which can be calculated as:

$$TPR = \frac{TP}{TP + FN} \tag{7}$$

FPR can be calculated as:

$$FPR = \frac{FP}{FP + TN} \tag{8}$$

AUC is known for Area Under the ROC curve. AUC calculates the two-dimensional area under the entire ROC curve ranging from (0,0) to (1,1), as shown in Fig 4. In the ROC curve, AUC computes the performance of the binary classifier across different thresholds and provides an aggregate measure. The value of AUC ranges from 0 to 1, which means an excellent model will have AUC near 1.

2) Regression indicators: MAE and RMSE are often used to quantify how well a model fits a dataset.

MAE tells us the mean absolute difference between the predicted values and the actual values in a dataset. The lower the MAE, the better a model fits a dataset. The calculation formula as follows:

$$MAE = \frac{1}{n}\sum_{i=0}^{n}|y_i - \breve{y}_i| \tag{9}$$

Where y_i is the observed value for the i^{th} observation, $\breve{y}_i$ is the predicted value for the i^{th} observation, and n is the sample size.

RMSE tells us the square root of the average squared difference between the predicted values and the actual values in a dataset. The lower the RMSE, the better a model fits a dataset. The calculation formula as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(y_i - \breve{y}_i)^2} \tag{10}$$

Many studies have shown that depressed and healthy people are very different in many aspects, such as brain imaging, electrophysiological signals, audiovisual data, text, and other multi-modal data. By analyzing the abnormal performance, these data can auxilia in the depression diagnose.

III. DEPRESSION RECOGNITION METHOD BASED ON BRAIN IMAGING

Compared with healthy people, the brain structure and function of patients with depression may have some abnormalities [62], which will inhibit the patient's thinking speed. In severe cases, it will lead to cognitive impairment and even a state of mental illness. In addition, depression can atrophy the hippocampus of the patient's brain, alter neurotransmitters in the brain, and cause chronic inflammation, resulting in memory loss and frequent fatigue, which greatly reduces the patient's energy level and motivation, making it difficult for the patient to socialize. These findings found a new breakthrough point for the study of the pathogenesis of depression. With the development of neuroimaging technology, great progress has been made in the exploration of neurobiological mechanisms of depression.

Brain structure imaging studies mainly focus on the frontal lobe, temporal lobe, amygdala, basal ganglia, cingulate gyrus,

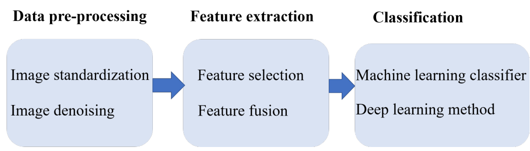


Fig. 5. Flow diagram of depression detection based on brain imaging.

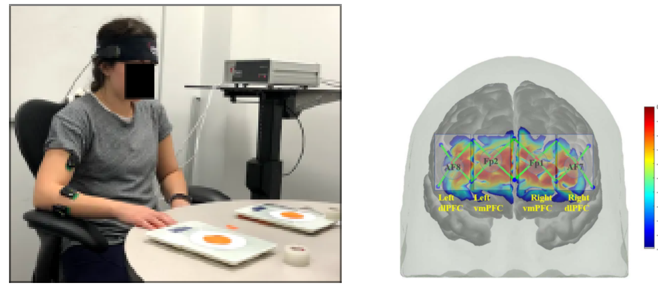
hippocampus and other brain regions of the cerebral cortex. Depression research based on brain imaging data is mainly divided into brain structure, brain function and brain metabolic imaging. Brain function imaging studies of depression mainly include Functional Nearinfrared Spectroscopy (fNIRS), nuclear Magnetic Resonance Imaging (MRI), etc. Brain metabolic imaging studies to explore depression mainly include Positron Emission computed Tomography (PET) and Magnetic Resonance Spectroscopy (MRS), etc.

As shown in Fig 5, brain images are acquired by scanning the subject's brain with various precision instruments, and there are significant differences in brain volume and structure among different people, as well as a large amount of interference and noise in the acquisition data due to the slight head movement of the subject or the scanning equipment during the acquisition process, so image pre-processing is needed to eliminate individual differences and noise interference. Among them, image normalization includes head movement correction, brain tissue separation, and segmentation alignment, this step can remove individual differences and head movement interference. Image denoising is mainly used to remove the noise generated by the external environment and equipment, and the common methods include Gaussian filtering, band-pass filtering, wavelet transform, etc. In the feature extraction process, linear models, wavelet packet decomposition and other methods are usually applied to extract time and frequency domain features, or the features of each brain region are extracted according to standard brain templates, and then redundant and irrelevant features are removed using feature selection algorithms, and the effective features are fused for subsequent classification.

A. fNIRS

fNIRS utilizes the scattering properties of blood to near-infrared light to obtain changes in oxyhemoglobin and deoxyhemoglobin during brain activity. It is an important tool for the study of brain function imaging. It can be used to compare the differences in brain function between healthy and mentally ill patients, and to achieve non-invasive detection of brain function [63]–[67].

In recent years, machine learning techniques have shown great potential in the diagnosis and treatment of mental health disorders including depression [68] [69]. The most common machine learning algorithms for predictive classification such as Support Vector Machines (SVM), Decision Trees (DT) and naïve bayes,etc, have been applied to fNIRS data to classify Major Depressive Disorder (MDD) patients and healthy individuals. All of these methods have yielded good results in depression identification [70]. Hong et al. [71] designed



(a) Data Collection (b) fNIRS

Fig. 6. Participant instrumented and sitting upright. PFC activation sensitivity map of fNIRS probe. Red and blue dots indicate sources and detectors, respectively , green lines indicate channels. Source-detector separation distance is 2.5 cm [72].

a SVM based classifier to classify the features extracted from fNIRS. There are also some integrated machine learning methods that have been applied in depression recognition, such as Random Forest (RF) and Extreme Gradient Boosting (XGBoost),etc. Zhu et al. [72] extracted ten features from HbO signals, from each channel served as inputs to XGBoost and RF algorithms developed for each block and combination of successive blocks.

Traditional machine learning classification methods require a priori feature preprocessing to train the model, extensive reliance on manual feature extraction and selection. By using deep Learning methods like Artificial Neural Network (ANN), Deep Neural Networks (DNNs) and Long Short-Term Memory (LSTM), etc, records can be fed directly to the algorithm for training, avoiding the need for feature selection. As a result, more and more researchers have applied deep learning algorithms to multimodal fNIRS classification tasks [73],and have also achieved high accuracy in depression recognition and diagnosis [74]. Ma et al. [75] was able to distinguish bipolar depression from major depressive disorder in adults during a verbal fluency task by using an LSTM . Chao et al. [76] used a cascade forward neural network to classify depression and achieved excellent results.

B. MRI

MRI is a new examination technology based on the principle that atomic nuclei with magnetic distance can produce transitions between energy levels under the action of a magnetic field. It helps to check the energy state and cerebral blood flow of the patient's brain. In MRI techniques, the MRI images obtained with different scanning parameters include structural MRI (sMRI) and functional MRI (fMRI).

1) sMRI: The sMRI images can clearly show the anatomical structure of the brain with high spatial resolution and can be used to objectively reflect the structural morphological changes within the brain. The detection of depression from sMRI scans usually includes processes such as image acquisition and preprocessing, feature extraction and selection, and classification. Machine learning methods have also been widely used, such as One Rule (OneR), SVM, Information Gain (IG) and ReliefF, etc [77] [78]. Deep learning methods have brought out remarkable results in medical imaging analysis, such as Mousavian et al. [79] found that the pre-trained VGG 16 and fine- tuning produced good results.

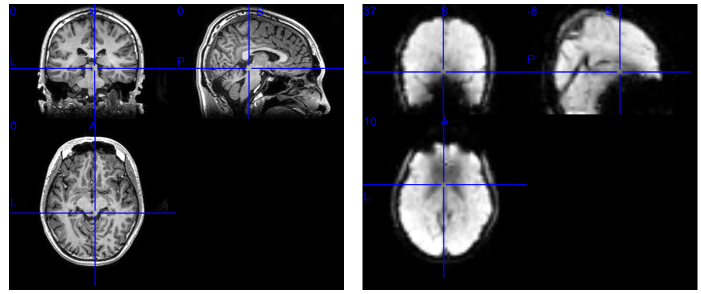
2) fMRI: fMRI uses magnetic resonance imaging to measure changes in hemodynamics caused by neuronal activity. It can observe the changes of the brain through the small changes in the magnetic resonance signal caused by the oxygenation state of the brain, thereby revealing the relationship between brain activity and thinking. According to different experimental conditions, fMRI is mainly divided into Resting State fMRI (RS-fMRI) and Task State fMRI (TS-fMRI).

RS-fMRI collects data when the subjects are in a relaxed state. Since there is no task interference, it can help doctors understand brain development, diseases, etc [80]. RS-fMRI is mainly aimed at the analysis of regional brain activity and functional connectivity of brain regions. Many researchers have built machine learning classification models to predict the accuracy of depression [81]–[83]. Zeng et al. [84] built an unsupervised machine learning model for MDD. Bhaumik et al. [85] extracted the features of the left posterior cingulate cortex and the right dorsolateral prefrontal cortex and input it to the SVM classifier for depression recognition. Yan et al. [86] exploited the hidden information embedded in dynamic functional connectivity (DFC) and developed an accurate and objective image-based diagnosis system for MDD. To improve the generalization ability of classifiers, multicenter studies [87] have been conducted and cover deep learning algorithms. Zhao et al. [88] established a generative adversarial network depression classification model based on functional brain network connections in a multicenter sample. Jun rt al. [89] distinguished drug-naïve MDD patients from healthy controls using Graph Convolutional Networks (GCNs).

TS-fMRI refers to the process of collecting data, subjects need to perform specific tasks, such as exercise, cognitive activities. Current research combining machine learning with TS fMRI is mainly related to emotional face stimulation, speech and music listening tasks, and this method enables individual-level analysis and reduces subjective misjudgments. Machine learning classifiers have achieved good results in related research, such as Gaussian Process Classifiers(GPC), RF, SVM, etc [90] [91]. Because deep learning has the advantages of reducing human intervention, deep analysis and feature extraction, selection and classification in the same optimal deep structure. Gui trained a Deep Learning Network (DLN) model with fMRI data of listening to positive and negative music [92].

3) Brain metabolic imaging: Brain metabolic imaging provides an accurate understanding of nerve cell activity and metabolic changes under normal conditions and disease states, as well as the metabolic situation of the cerebral cortex under different physiological conditions of stimulation and thinking activity. Brain metabolic imaging such as PET and MRS can visualize the metabolic activity and various physiological or pathological metabolic changes in human brain and reflect them in the form of images.

PET uses the decay law or distribution characteristics of substances related to nuclear radiation, such as glucose, proteins, nucleic acids, etc., in the research object to obtain detailed information to reflect their metabolic activities to achieve the purpose of diagnosis. PET is mainly based on the annihilation effect of positrons and electrons and can be used to explore characteristic pathological states and indicators in depressed patients. By investigating the correlation between changes in metabolic function in MDD brain regions and depressive symptoms, it was found that compared to normal healthy subjects, depressed patients usually show such phenomena as decreased regional cerebral blood flow values and standardized glucose uptake values in the prefrontal lobe, increased total distribution of translocated proteins, and increased inflammatory markers [94]–[97].



(a) sMRI images (b) fMRI images

Fig. 7. Three View of MRI images [93].

MRS uses nuclear chemical shifts to study molecular structure and can be used to detect metabolite concentrations in brain regions. MRS can detect biochemical changes in the injured tissue of the body by magnetic resonance hydrogen spectroscopy, which can effectively observe the metabolism of the injured area and has some clinical value for early diagnosis of the depression. It was found that N-acetylaspartate (NAA) levels were significantly lower in the left hippocampus, glutamine and glutamate (Glx) levels were significantly increased in the right hippocampus, and choline complex (Cho) and creatine (Cr) levels were significantly higher in depressed patients compared to controls, and correlated with the severity of depression in patients [98]–[100].

Brain imaging technology has revealed abnormalities in brain structure, brain function and brain metabolism in patients with depression, providing new ideas for early diagnosis and optimization of treatment plans.

4) Performance Comparison: Table III shows a summary of the experimental results based on brain imaging to detect depression, including the source and name of the method, the input brain imaging modality, the data set used (- represents that the data used were collected by ourselves), and the evaluation criteria of the experimental results, including Accuracy, FScore Precision, Recall, Sensitivity, Specificity, and Area Under Curve (AUC).

SVM dominates the machine learning classifiers and usually achieves good results when compared with other classifiers. And deep learning algorithms usually achieve higher recognition accuracy than machine learning classifiers.

TABLE III
EXPERIMENTAL RESULTS BASED ON BRAIN IMAGING

ID	Method	Brain imaging	Dataset	Classification						
				Accuracy	F Score	Precision	Recall	Sensitivity	Specificity	AUC
1	SVM. [71]	fNIRS	-	86.76	-	-	-	-	-	-
2	RF+XGBoost [72]	fNIRS	-	92.58	-	-	-	84.78	91.05	-
3	AlexNet [74]	fNIRS	-	90.00	-	91.00	-	-	-	-
4	ALSTMIT [75]	fNIRS	-	96.20	-	-	-	-	-	-
5	SVM [70]	fNIRS	-	85.15	-	-	-	-	-	-
6	CNN-LSTM [101]	RS-fMRI	NKI online dataset	100.00	-	-	-	86.00	0.00	50.00
7	ST-CNN [101]	RS-fMRI	NKI online dataset	100.00	-	-	-	86.00	0.00	50.00
8	3D CNN [101]	RS-fMRI	NKI online dataset	90.00	-	-	-	79.00	10.00	50.00
9	3D VGG-16 [101]	RS-fMRI	NKI online dataset	100.00	-	-	-	86.00	0.00	50.00
10	FCN [101]	RS-fMRI	NKI online dataset	94.00	-	-	-	79.00	96.00	88.00
11	FCN_woT-test [101]	RS-fMRI	NKI online dataset	86.00	-	-	-	0.00	100.00	50.00
12	FCN_CanICA [101]	RS-fMRI	NKI online dataset	87.00	-	-	-	70.00	90.00	80.00
13	DCF+SVM [86]	RS-fMRI	-	95.59	-	-	-	96.77	94.68	99.13
14	SVM [85]	RS-fMRI	-	76.10	-	-	-	81.50	68.90	-
15	Unsupervised ML [84]	RS-fMRI	-	92.50	-	-	-	-	-	-
16	FCN+GAN [88]	RS-fMRI	-	70.10	70.30	-	-	73.50	66.50	70.30
17	SVM [81]	RS-fMRI	-	90.00	-	-	-	-	-	-
18	GAN [88]	RS-fMRI	-	80.70	-	-	-	-	-	-
19	GCNs [89]	RS-fMRI	-	74.10	-	-	-	56.60/46.6	86.90	79.10
20	SVM-FoBa [102]	RS-fMRI	-	84.78	83.72	91.30	78.26	-	-	-
21	MVPA [103]	RS-fMRI	-	90.22	89.89	93.02	86.96	-	-	-
22	LASSO [87]	RS-fMRI	-	70.00	-	-	-	-	-	-
23	sLASSO [91]	fMRI	-	83.82	84.00	-	-	84.61	83.03	-
24	gLASSO [91]	fMRI	-	91.95	92.00	-	-	91.35	92.55	-
25	sgLASSO [91]	fMRI	-	90.08	90.00	-	-	88.45	91.73	-

IV. Depression Recognition Method Based on Electrophysiological Signals

Electrophysiological signals, which mainly include EEG, ECG, EOG, galvanic skin response, and body temperature, have been used extensively to diagnose depression because of their non-invasive detection and ease of use [104]. Among them, EEG [105]–[107] is used most frequently, followed by ECG [108], and EOG [28]. It is interesting to note that practically all research based on EOG employ eye movement signals (EM) to identify depression. Furthermore, as other types of signals are used too infrequently, we focus on EEG, HRV, and EM-based depression recognition.

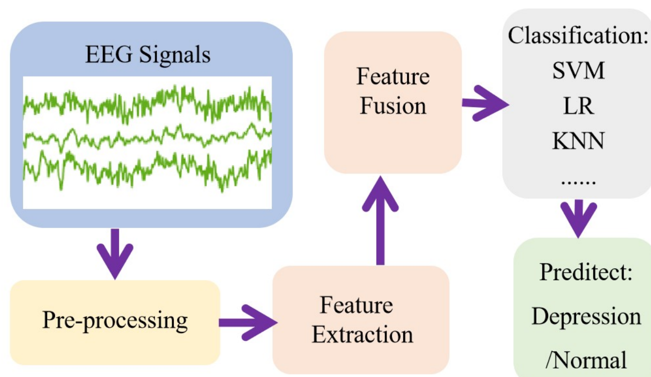


Fig. 8. Flow diagram of EEG depression recognition using traditional machine learning.

A. EEG

EEG is the most direct and objective representation of the neurological activity of the human brain. It is closely related to brain activity and emotional state, provides a timely reflection of the body's emotional changes, and provides a high temporal resolution for assessing brain dynamics. So the EEG is one of the most promising signals that can be employed to help with depression diagnosis. Current research about EEG-based depression recognition is still primarily based on traditional machine learning algorithms. Meanwhile, a few researchers have experimented with a subfield of machine learning-deep learning algorithms, which have made breakthroughs for related research.

1) EEG depression recognition based on traditional machine learning: The main process of EEG depression identification using traditional machine learning is shown in Fig 8. There are two key points to emphasize: feature extraction and classifiers. According to the characteristics of EEG signals, feature extraction can be roughly further summarized into two categories: linear and non-linear. However, brain activity is inherently non-linear, and the EEG signal is non-stationary [109]–[111], which lead to limitations in the linear approach to assessing the dynamics of depressive EEG. In contrast, the non-linear approach demonstrates a relatively clear advantage. The non-linear dynamics theory suggests that a self-organizing, low-dimensional, deterministic system can generate complicated and non-periodic behavior that is inherent to the system rather than generated from external sources. For such non-linear systems, numerous parameters have been discovered, including the correlation dimension [112], the Lyapunov exponent [113]–[115], and entropy [116]–[118]. The correlation dimension describes the system's complexity; the largest Lyapunov exponent reflects the pace of uncertainty growth and is sensitive to the system's beginning value; and the entropy depicts the level of order in the system. Researchers also employ a combination of multiple nonlinear indictors because they can each reflect the distinctive properties of EEG from a different angle, exposing information that may not be picked up by other approaches. For example, Acharya et al. [119] presented a novel method for automated EEG-based diagnosis of depression using nonlinear methods: fractal dimension, largest Lyapunov exponent, sample entropy, detrended fluctuation analysis, Hurst's exponent, higher order spectra, and recurrence quantification analysis. The rational combination of the nonlinear features is used to diagnose depression.

It was determined that depression had an effect on patients' EEG activity by comparing the distribution of linear and non-linear indicators between depressed patients and healthy controls. Based on this, the researchers chose a few markers with strong discriminatory power to build a series of features for depression identification. They then created a depression identification model based on different classifiers to achieve initial detection and diagnosis of depression. Although the indicators utilized vary, the popular classifiers are usually the same, such as k-Nearest Neighbor (KNN), linear discriminant analysis, linear regression (LR), support vector machines (SVM), decision trees, Bayesian classifiers, random forests and probabilistic neural networks. Hosseinifard et al. [120] used associative dimensional features as input to a linear regression classifier and obtained 90% recognition accuracy. Ahmadlou et al. [121] fed fractal dimensional features of depressed patients into an augmented probabilistic neural network and obtained an accuracy of 91.3%. Cukic et al. [117] obtained an accuracy of 90.24%-97.56% by feeding Higuchi fractal dimension and sample entropy into seven classifiers.

2) EEG depression recognition based on Deep learning: In recent years, deep learning has been utilized more and more to detect depression with the help of EEG. All adopted models in these studies can be categorized into three main groups, namely Convolutional Neural Network (CNN) models [122], Long-short Term Memory (LSTM) models, and hybrid neural networks [123]–[125]. And their detailed flow is shown in Fig 9.

In specific, the CNN-based model can automatically and adaptively extract useful information directly from the input data instead of manually selecting features. Acharya et al. [107] presented the first application of the deep neural network concept and CNN for diagnosis of depression. The model's inputs were the EEG signals from the left and right hemispheres of the brain, respectively. 5 convolutional layers were responsible for providing important feature obtained from the input EEG signals to train the algorithm, 5 pooling layers were accountable for reducing of the feature map's size, and 3 fully-connected layers established connection between neurons of a layer with the next one. Finally, results produced by the last fully-connected layer was applied to a soft-max function to detect depressive cases. Similarly, Li et al. [126] analyzed different aspects of EEG (spectral, spatial, and temporal information) for the diagnosis of of mild depression. Results showed that spectral information of EEG signals play major role whereas temporal information showed significant improvement in diagnostic performance. The uitlized the pre-trained ConvNet architectures on EEG-

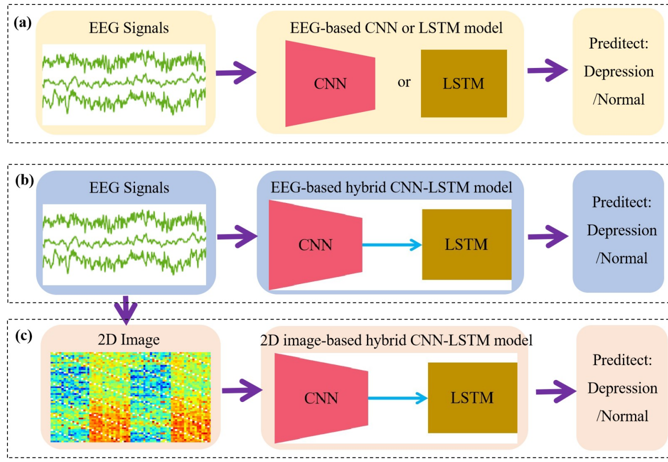


Fig. 9. Flow diagram of EEG depression recognition based on deep learning. (a) Raw signal-based deep feature extraction and classification using CNN or LSTM; (b) raw signal-based deep feature extraction and classification using hybrid neural networks; (c) 2D image-based deep feature extraction and classification using hybrid neural networks.

based mental load classification task and achieved accuracy of 85.62% for recognition of mild depression and normal controls.

The benefit of the LSTM when processing temporal signals such as EEG is that it has the ability to read and modify the long-term dependent information of a temporal sequence at will [127], [128]. Only a little portion of LSTM research, nevertheless, have focused on detecting depression through physiological signals. Kumar et al. [129] researched depression prediction by the LSTM model and the help of feature extraction. To generate input data, time-domain analysis with moving window segmentation was employed to extract the statistical mean feature. LSTM block comprised of one LSTM layer with ten hidden neurons, a dropout layer of 0.1, and a dense layer. They also compared the introduced method with two other models that were CNN-LSTM and ConvLSTM. Among them, LSTM had the best performance over those structures since it owned the smallest RMSE values (Root mean square error) as model evaluator rather than them.

The majority of the hybrid neural network are combined models of both CNN and LSTM blocks. Additionally, it has two significant groups of data input formats: time series and feature images. Concretely, Thoduparambil et al. [125] designed a deep model in which an integration of CNN and LSTM is implemented for the detection of depression. Three CNN layers and three MaxPooling1D layers formed the first section of the model which exploited to extract features. The second unit included two LSTM layers whose duty was to generate the feature maps by discovering different patterns in EEG signals and then retain the sequence of these learning. Similarly, Sharma et al. [130] also proposed a new hybrid neural network called DepHNN. CNN and LSTM are two deep learning algorithms used to capture the temporal dependencies in the time-series EEG input signals and to process the sequence learning, respectively. Besides, Saeedi et al. [123] used brain effective connectivity method to convert 1D EEG signal into 2D image. Then, they developed a classifier using state of the art deep learning methods (CNN-LSTM) as a novel approach for automated diagnosis of the depression patients from EEG signals. Finally, their experiments showed that the spatial and temporal characteristics of the EEG signals are captured by 1DCNN-LSTM. Relying on the results, deep learning model is also capable of effectively analyzing the brain connectivity.

3) Performance Comparison: We evaluate the depression recognition method based on EEG. The performance comparisons based on traditional machine learning and deep learning are summarized in Table IV. We would like to note that, we cannot comprehensively compare all studies because the dataset utilized vary, but by comparing the work from the same researchers, we can draw the following conclusion:

(1) Single-channel EEG analysis, employing the combination of measures, can provide the accuracy for discrimination of depression not lower than reported in other studies where multichannel EEG signals were analysed [131].

(2) The nonlinear analysis of EEG can be a useful method for discriminating depressed patients and normal subjects. According to [120], the accuracy of three classifiers are higher for all nonlinear features as the input in compare to power bands features. Brain system is best-characterized non-linear dynamical process. The nonlinearity of brain limits the ability of linear analysis to provide full description of underlying dynamics.

(3) In contrast, the novelty of the deep learning model is that it does not require the employment of feature extraction, selection, and reduction. The model has the ability to self-learn and pick up distinctive features during training without a separate feature extraction or feature selection step. Acharya [107] presented the first application of the deep neural network concept for diagnosis of depression. Even though the accuracy is not as great as the previous one [119], deep learning technology development is anticipated to lead to new discoveries.

B. ECG

Depression is associated with the autonomic nervous system (ANS) dysfunction [135]. Heart rate variability (HRV), levels of variability of the heart beat-to-beat interval over time, has been known to provide an index of ANS functioning including the sympathetic and parasympathetic system [136]. Many studies have demonstrated automated detection using HRV in patients with major depression using machine learning methods.

1) HRV depression recognition based on traditional machine learning: In terms of classic approaches, the fundamental steps of EEG-based depression identification and HRV-based depression identification are similar. Particularly, the signal's multidimensional features are first extrated, then a classification model is built, and finally this model is applied to diagnosis depression. The most commonly used features to quantify HRV signals are time-domain and frequency-domain features. Time-domain HRV features can be calculated directly from the time series of R-peak to R-peak interval (RRI) [137]. The mean and standard deviation of RRIs (SDNN) [138], the root mean square of successive RR interval differences (RMSSD), and the percentage of successive RRIs differing by more than 50ms (PNN50) are all commonly used time domain features [139]. Whereas the mean of RRIs measures the intensity of HRV [140], the SDNN measures overall HRV and reflects both sympathetic and parasympathetic activity in the autonomic nervous system [141], and the RMSSD and PNN50 are more sensitive to parasympathetic modulation.

Frequency-domain provides an assessment of vagal modulation of the RRI, extracted from the ECG. It is mostly commonly acquired by fast Fourier transformation (FFT). The RRI can be separated into three components depending on the frequency band: very low frequency (VLF, 0.003-0.04 Hz), low frequency (LF, 0.04-0.15 Hz), and high frequency (HF, 0.15-0.4 Hz) band [142]. According to previous studies, HF is primarily affected by parasympathetic activities while LF is influenced by both sympathetic and parasympathetic activity. The ratio of LF to HF indicates that sympathetic activity predominates as compared to parasympathetic activity [143].

The typical HRV markers mentioned above are frequently used to detect depression as well. Kemp [144] summarized the achievements of numerous researchers and discovered that depressive individuals had much lower SDNN and RMSDD values in the time domain, HF values in the frequency domain, and HF/LF values in the frequency domain than healthy controls. The differences in HRV between individuals with severe depression and healthy controls have been further investigated in numerous subsequent research. Additionally to these markers, it was discovered that compared to healthy participants, depressive patients exhibited considerably lower PNN50 in the temporal domain and greater LF in the frequency domain [145].

Although HRV has been analyzed traditionally using linear methods, such as time-domain and frequency-domain analyses, growing evidence has demonstrated that linear HRV measures may not correctly represent the complex dynamics of heartbeat regulation modulated by the ANS [146]and that linear HRV features show a relatively higher inter-subject variability than nonlinear HRV indices, suggesting the importance of nonlinear HRV analysis. Schulz et al. [147] have shown that nonlinear HRV indices allow more reliable discrimination of major depressive disorder (MDD) patients from controls than linear HRV features, as the latter exhibit high inter-subject variability. In particular, the non-linear features used for EEG: fractal dimension, Lyapunov's index, entropy, etc., also apply to HRV. Greco et al. [148] found that non-linear indicators such as fractal dimension, sample entropy and recursive graph analysis demonstrated significantly higher heart rate variability even in subclinical depression compared to healthy controls. Daniel et al. [149]found that depressed patients over 60 years of age had significantly lower sample entropy values compared to healthy controls.

In order to recognize depression using HRV, a variety of classifiers can be used. Kang et al. [150] used the bayesian network to recognition depressed patients from a healthy people, the recognition results demonstrate the significant association between depression and HRV. Byun et al. [141] demonstrated the HRV-based diagnosis of MDD using SVM classifier. Monitoring the changes in linear and nonlinear

TABLE IV
OVERVIEW OF MACHINE LEARNING BASED METHODS FOR DEPRESSION ASSESSMENT FROM EEG.

Methods	Paper	Dataset	Feature	Classification	Metrics	Value
EEG depression recognition based on traditional machine learning	Acharya [119]	15D+15C	FD, LLX, Hurst,HOS, SampEn, DFA and RQA	SVM	Accuracy	98.00
	Bachmnn [131]	13D+13C	SAI, APV and RGP HFD, DFA and Lempel-Ziv	LR	Accuracy	Single-channel:92.00 Multil-channel:90.00
	Hosseinifard [120]	45D+45C	Power DFA, HFD, CD and lyapunov exponent	KNN	Accuracy	Linear/Nonlinear:73.3/80
				LDA		Linear/Nonlinear:76.6/86.6
				LR		Linear/Nonlinear:76.6/90
	Cukic [117]	23D+20C	HFD,SampEn	Multilayer perceptron	Accuracy	95.12
				LR		97.56
				SVM with linear kernel		95.12
				SVM with polynomial		95.12
				Decision tree		95.12
				Random forest		92.68
				Naive Bayes		92.68
	Ahmadlou [121]	12D+12C	FD	EPNN	Accuracy	91.30
	Faust [132]	30D+30C	ApEn, SampEn, REN and Ph	PNN	Accuracy	99.7
	Liao [105]	12D+12C	KEFBCS	SVM	Accuracy	80.00
EEG depression recognition based on deep learning	Wan [122]	23D+12C	Raw signals	HybridEEGNet	Accuracy	79.08
	Saeedi [123]	34D+30C	Effective connectivity	1DCNN-LSTM	Accuracy	99.24
	Quayyam [124]		Raw signals	1DCNN-GRU-RF	F1-score	99.94
	Thoduparambil [125]		Raw signals	CNN-LSTM	Accuracy	Left hemisphere:98.84 Right hemisphere:99.07
	Acharya [107]	15D+15C	Raw signals	CNN	Accuracy	Left hemisphere:93.5 Right hemisphere:96.0
	Kang [133]	34D+30C	Asymmetry matrix image	CNN	Accuracy	98.85
	Kumar [129]	30D	time domain feature	LSTM	RMSE	0.005
	Sharma [130]	21D+24C	Raw signals	DepHNN	Accuracy	99.1
	Li [126]	24D+24C	spectral, spatial, and temporal information	ConvNet	Accuracy	85.62
	Ay [134]	15D+15C	Raw signals	CNN-LSTM	Accuracy	Left hemisphere:99.12 Right hemisphere:97.66

Largest Lyapunov exponent: LLX; Higher order spectra: HOS; Recurrence quantification analysis: RQA; Spectral asymmetry index: SAI;
Alpha power variability: APV; Relative gamma power: RGP; Detrended fluctuation analysis: DFA; Linear discriminant analysis: LDA; Sample Entropy: SampEn; Fractal Dimension: FD; Enhanced probabilistic neural network: EPNN; approximate entropy: ApEn; Renyi entropy: REN; Bispectral phase entropy: Ph; Kernel eigen-filter-bank common spatial pattern: KEFBCS.

HRV features for various autonomic nervous system states can facilitate the more objective identification of MDD patients.

2) HRV depression recognition based on deep learning: Although traditional machine learning techniques perform well, they nevertheless have several shortcomings, which are dependence on feature extraction and feature selection to a large extent. It indicates that the procedure consequently becomes both time and computationally-intensive. Unlike traditional machine learning, deep learning can automatically learn features from input data, therefore researchers are trying to use it to diagnose depression. Noor et al. [151] presented a model that used Recurrent Neural Network (RNN)and LSTM to predict the risk of depression based on HRV. zang et al. [152] proposed a CNN network containing 2 convolutional layers, 2 max pooling layers and 1 fully connected layer to diagnose depression, and the results showed that the network achieved high classification performance. In conclusion, the end-to-end deep learning approach can identify depression from HRV signals, and possess high diagnostic performance.

3) Performance Comparison: Previous studies in which ECG/HRV features were used to discriminate depression patients have yielded promising results obtained via various machine learning methods, which are summarized in Table V. In these studies, Kuang et al. [150] achieved an accuracy of 86% through selected HRV features, but only female subjects participated in this study; Zhang et al. [153] used relatively small sample sizes. Traditional studies have used HRV features to discriminate patients with depression, and HRV is usually expressed from the RR interval collected from ECG data. It is inevitable that the QRS wave algorithm will be utilized to locate R-peak. The R-peak positioned by different QRS wave algorithms also has biases. On the contrary, the deep learning method directly takes the ECG signal as input without extracting the HRV sequence. In addition, traditional studies used simple classifiers to train features manually extracted from HRV sequences. The performance of this type of classifier mainly depends on the feature selection process, which is laborious and time-consuming. The deep model can overcome the shortcomings of manual feature extraction and selection.

C. EM

EM refers to the voluntary and involuntary movements of the eyes that assist with obtaining, fixating and following visual stimuli. EM information, which includes object gaze, gaze duration, gaze displacement, and pupil size, provides a clear indicator of the information-processing mechanisms at work in the brain and can be used to characterize how emotions are perceived by individuals. Therefore, EM is often used as the primary measure of emotion and mood.

The researchers observed, measured, and recorded the experimental participants' EM characteristics through a variety of task paradigms, which can be categorised into saccade, gaze, memory and free-viewing tasks. EM measures used on saccade tasks include saccade gain, saccade error rate and task mistake rate. These metrics are related to cognitive ability. Sweeney et al. [158] found that depressed patients demonstrated increased rates of response suppression errors on an antisaccade task. The gaze-type task examines the subject's capacity to suppress eye-hopping and is linked to brain function in the dorsolateral prefrontal cortex (DLPFC). In the memory task, subjects must first fixate on a central point while being prohibited from moving their gaze towards the goal, which appears in the peripheral vision. After the peripheral target disappears they are asked to eye skip towards the target area from memory, while eye skip latencies are recorded and error rates are used to assess damage to the basal ganglia and frontal regions. Studies have revealed that depressed patients have higher error rates for eye skipping than normal controls. A free-viewing task, which measures attentional bias, is usually achieved by viewing pictures of emotional faces. According to Duque et al. [159], major depressive patients had a negative affective bias, or a longer duration of first look and total gaze on negative emotional faces than normal controls, which was positively connected with the severity of depressed symptoms.

New developments in the diagnosis of depression using eye-movement measurements have been made possible by the development of machine learning. li et al. [160] used statistical methods (FDR correction and t-test) and PCA to screen behavioral features as well as eye-movement features, and finally achieved a 91% depression recognition rate using a kernel-limit learning machine. Pan et al. [161] extracted reaction time features reflecting subjects' attentional bias as well as eye-movement features and used SVM to achieve 86% classification accuracy for both depressed and normal categories. Lu et al. [162] used a particle swarm algorithm to optimise the regularisation and kernel parameters in an extreme learning machine and used eye-movement data to classify both normal and depressed people, achieving a classification accuracy of 88.55%.

V. Depression recognition method based on Audiovisual data

Depression is now the most common type of psychological disorder, manifested as a single or repeated multiple depressive episodes, slow clinical behavior, passive and lazy life, do not want to do anything, do not want to contact and interact with people around, often sitting alone, or lying in bed all day, living alone behind closed doors, alienating friends and relatives, avoiding social interaction, etc.. Second, the patient mainly shows significant and persistent depression, depression and pessimism. The patient's thinking is slow, unresponsive, and blocked. Clinically, there is a decrease in active speech, a significant slowdown in speech, a low voice, and difficulty in answering questions, and in severe cases, communication is not smooth. These characteristics are all abnormalities of depressed patients compared to healthy people and can be captured visually or audibly, which can be used to identify depression.

In recent years, there has been an increasing interest in identifying depression from behavioral signals and analyzing abnormal expressive behaviors that may result from depression, such as facial expressions that become sluggish, frequent avoidance of eye contact with others, and use of short sentences with a flat tone when speaking, and many researchers

TABLE V
OVERVIEW OF MACHINE LEARNING BASED METHODS FOR DEPRESSION ASSESSMENT FROM HRV.

Methods	Paper	Dataset	Feature	Classification	Metrics	Value
HRV depression recognition based on traditional machine learning	Zhang [153]	10D+10C	VLF,LF,HF,LF/HF,SDNN,RMSSD	Neuro-fuzzy network	Accuracy	95.00
	Byuna [141]	31D+41C	RRI,SDNN,RMSSD,PNN50,TRI,TINN, VLF,LF,HF,LF/HF,Tot	SVM	Accuracy	74.4
			ApEn,SampEn,DFA,CD, SD1,SD2	Statistical filter	Accuracy	73.1
	Roh [154]	23D	HR,SDNN,PNN50, LF/HF,peakLF,peakHF, LLE,SampEn,ApEn	SVM	Accuracy	71
	Matsui [155]	13D+28C	LF,HF	LDA	Accuracy	88
	Sun [156]	44D+47C	LF,HF,LF/HF	Logistic regression	Accuracy	79
	Kuang [150]	38D+38C	SDNN,RMSSD,SDSD,PNN50,MEAN VLF,LF,HF, SampEn,DFA	Bayesian networks	Accuracy	86
HRV depression recognition recognition based on deep learning	Noor [151]	5000D	Raw signals	RNN	Accuracy	97.24
	zang [152]	37D+37C	Raw signals	CNN	Accuracy	93.96
	Mohanraj [157]	15D+15C	Raw signals	DesNN	Accuracy	90

Standard deviation of the successive difference of RR intervals: SDSD; Mean value of RR intervals: MEAN; Integral of the histogram of the RR interval divided by its height: TRI; Baseline width of the RR interval histogram:TINN; Total power: Tot; Standard deviation of the Poincar plot perpendicular to (SD1) and along (SD2) the line of identity: SD1, SD2.

have found that depression identification can be well achieved by comparative analysis of these characteristics.

AVEC is an expression recognition challenge held every year since 2011, jointly organized by Imperial College London, University of Nottingham, Queen Mary University, USC and University of Passau, Germany, etc. It is recognized as the top international competition in the field of affective computing. AVEC 2013 started to introduce the task of depression recognition, considering the auditory-visual based analysis of depression as a regression problem or classification problems.

A. Depression recognition method based on facial expression

A person's personality and mood can be seen from his face, and some studies have found that the appearance and temperament of patients with depression are also different from ordinary people. Zhu et al. [163] research team of the institute of psychology of the Chinese Academy of Sciences studied 100 people with mental diseases. They asked 100 patients to read neutral articles through a special technology. From the research process, it was found that these patients frowned and drooped corners of their mouths when reading aloud, and a few tears appeared in their eyes, just like crying, looking very sad.

In order to study the relationship between patients' depression severity and facial expression over time, Girard et al. [164] designed a study to track the subjects for two years, and collected the data of 36 patients with severe depression at

Fig. 10. Samples from AVEC datasets.

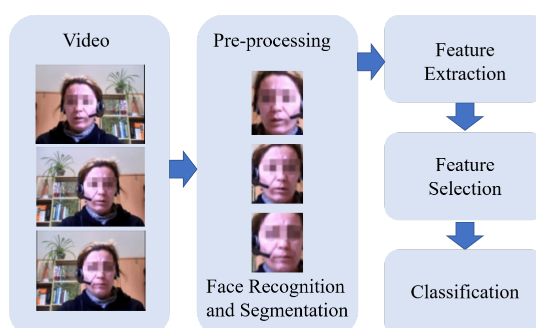


Fig. 11. Flow chart of expression-based depression recognition model.

the beginning and reduced symptoms after two years. Through the analysis of facial expressions in video data by manual and automatic systems, it showed that the automatic coding of Facial Action Coding System (FACS) action units is highly consistent with the manual coding, and showed a similar effect in the change of depression severity over time.

The results showed that when the severity of patients' symptoms was high, participants made more facial expressions related to contempt and smiled less, and those smiles were more likely to be accompanied by facial actions related to contempt. These results were consistent with the performance of Exchange-Oriented Withdrawal in the "Social Risk Hypothesis" of depression, indicating that patients with depression show more withdrawal in social intercourse. According to this hypothesis, when patients have severe symptoms, they will stay away from others to protect themselves from expected rejection, contempt and social exclusion [165]–[167].

1) Facial expression depression recognition based on machine learning: To achieve depression recognition from facial images, traditional solutions for vision-based automatic depression diagnosis systems usually include several steps as shown in Fig 11: first, facial images are recognized and segmented from the raw video data, then features are extracted, and finally the features are fed into a classifier for classification or regression.

Machine learning can be divided into traditional machine based on the depth of the model structure learning and deep learning, which are now often applied to tasks such as vision and speech on tasks such as vision and speech. While traditional machine learning algorithms are suitable for small amounts of data, deep learning has higher performance on larger data sets.

Traditional machine learning methods such as SVM, naive Bayes (NB), RF and logistic regression (LR) are the most commonly used classifiers in the study of facial features of depressed patients [168]–[171].

2) Facial expression depression recognition based on deep learning: Convolutional neural networks (CNNs) are the most commonly used deep learning networks for facial recognition research in recent years. Many researches based on CNN and its innovative architectures such as three-dimensional CNN (C3D), modality separation networks (MSN), deep residual regression convolutional neural networks (DRR-CNN) for recognizing, classifying, and predicting human emotions, as well as exploring how facial action intensity changes from low to high levels of emotion [172]–[177]. There are also studies that build on CNNs by embedding expectation loss into ResNet-50, a residual neural network, for distributional learning, which allows exploring the sequential relationship between facial images and depression levels to better predict depression levels [178].

Recurrent neural networks (RNN) [179] are suitable for learning for time series, better simulate feature changes to improve classification accuracy, and when combined with CNNs can also handle computer vision problems that contain sequential inputs. The LSTM [180] [181] is also commonly used in facial recognition research, and is suitable for processing and predicting important events with very long intervals and delays in time series, which is more in line with the detection of continuously changing emotions and close to the clinical reality.

Depression alters many behaviors, among which the face presents most of people's nonverbal information. Therefore, facial expressions are highly informative characteristic indicators in the diagnosis of depression, providing more support for clinical studies of depression and offering the possibility of automated detection of depression.

3) Performance Comparison: Table VI summarizes the results of the experiments on the detection of depression based on facial expressions, including the source and name of the method, the data set used (- means the data used were collected by ourselves), the evaluation criteria of the experimental results, including the classification criteria Accuracy, F1 Score Precision, Recall, the evaluation criteria of depression severity by depression scale mean absolute error (MAE) and root mean square error (RMSE), and Average Error for regression scale scores.

The AVEC competition has greatly advanced research on depression detection based on behavioral performance, and more and more deep algorithms have been applied to depression recognition tasks, all with good results.

B. Speech

Speech is a non-invasive signal that is low-cost and easily accessible. The study of speech-based depression recognition began with the examination of clinical aspects of depressed individuals' speech. Clinical observations have revealed that there are phonetic differences in speech between healthy and depressed populations. These differences can be seen in the fact that depressed people typically speak more smoothly and monotonously, whereas healthy people speak more rhythmically, with less pauses and fewer times than depressed people [50], [193]. Furthermore, the researchers designed characteristics based on the variations between the two groups. These characteristics are objective in nature and are able to discriminate between populations that are healthy and those that are depressed. But with the rapid advancement of deep learning techniques in recent years, speech-based depression recognition has transitioned from hand-crafted acoustic features to a deep learning-based framework.

1) Speech depression recognition based on traditional machine learning: So far, automatic speech-based depression analysis has primarily relied on classic machine learning methods supplemented with hand-crafted feature, as shown in Fig.12. In terms of hand-crafted feature engineering, the major work is to learn acoustic features related to depression and experiment with feature sets to improve performance. Researchers have also employed classic machine learning algorithms, such as support vector machines, as classifiers for depression diagnosis.

In the study of speech-based depression recognition, acoustic characteristics principally include prosodic, voice quality, formant, and spectral features [194], [195]. The prosodic features were computed from the speech waveform and included fundamental frequency (F0), log of energy, jitter and shimmer

TABLE VI
EXPERIMENTAL RESULTS BASED ON FACIAL

ID	Method	Dataset	Classification				depression severity estimation		Regression	
			Accuracy	F1 Score	Precision	Recall	RMSE	MAE	Scale	Average Error
1	AVEC2013 baseline [182]	AVEC2013	-	-	-	-	13.61	10.88	-	-
2	AVEC2014 baseline [183]	AVEC2014	-	-	-	-	10.86	8.86	-	-
3	LPQ+Geo [184]	AVEC2013	-	-	-	-	9.72	7.86	-	-
4	MHH+PLS [185]	AVEC2014	-	-	-	-	10.5	8.44	-	-
5	LPQ-TOP+MFA [186]	AVEC2013	-	-	-	-	10.27	8.22	-	-
6	AVEC2016 baseline [187]	DAIC-WOZ	-	50.00	60.00	42.80	7.13	5.88	PHQ	7.13
7	Williamson et al. [188]	DAIC-WOZ	-	53.00	-	-	-	-	PHQ	5.33
8	AVEC2017 baseline [189]	AVEC2017	-	-	-	-	6.97	6.12	-	-
9	LSTM [180]	-	67.7	-	-	-	-	-	-	-
10	Two DCNN [190]	AVEC2013	-	-	-	-	9.82	7.58	-	-
11	Two DCNN [190]	AVEC2014	-	-	-	-	9.55	7.47	-	-
12	SVM+NB+RF [169]	-	69.10	-	-	-	-	-	-	-
13	AVEC2019 baseline [60]	AVEC201	-	-	-	-	8.01	-	-	-
14	C3D [172]	AVEC2013	-	-	-	-	8.26	6.40	-	-
15	C3D [172]	AVEC2014	-	-	-	-	8.31	6.59	-	-
16	Ray et al. [191]	-	-	-	-	-	8.95	-	-	-
17	ResNet-50 [178]	AVEC2013	-	-	-	-	8.25	6.30	-	-
18	ResNet-50 [178]	AVEC2014	-	-	-	-	8.23	6.15	-	-
19	ResNet-50+pool [192]	AVEC2014	-	-	-	-	8.43	6.37	-	-
20	MSN [175]	AVEC2013	-	-	-	-	7.90	5.98	-	-
21	MSN [175]	AVEC2014	-	-	-	-	7.61	5.82	-	-
22	MR-DepressNet [176]	AVEC2013	-	-	-	-	8.28	6.20	-	-
23	MR-DepressNet [176]	AVEC2014	-	-	-	-	8.39	6.21	-	-
24	Two-Stream model [174]	AVEC2013	-	-	-	-	7.97	5.96	-	-
25	Two-Stream model [174]	AVEC2014	-	-	-	-	7.94	6.20	-	-
26	CNN [170]	-	66.45	-	-	-	-	-	-	-
27	SS-LSTM-MIL [181]	DAC-WOZ	-	78.3	81.80	75.00	-	-	-	-
28	RNN-C3D Tight-Face & Loose-Face weighted merge [179]	AVEC2013	-	-	-	-	9.28	7.37	-	-
29	RNN-C3D Tight-Face & Loose-Face weighted merge [179]	AVEC2014	-	-	-	-	9.20	7.22	-	-

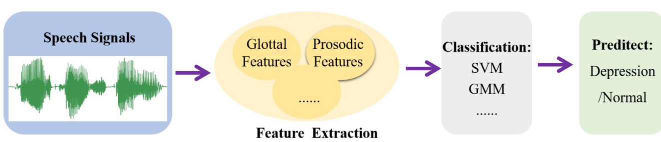


Fig. 12. Flow diagram of speech depression recognition based on traditional machine learning.

[194]. Where F0 is one of the most prevalent prosodic features, and its range of variation and decrease in mean value may be related to the severity of depression. The parameters related to voice quality features are frequency perturbation, amplitude perturbation, and glottal parameters. The glottal spectrum shows the condition of vocal function, while frequency and amplitude perturbation combined describe the stability of vocal fold vibration [196]. Formant features provide information regarding vocal tract resonance and pronunciation efforts, which represent physical vocal tract properties. As formant features, the first three formants are commonly employed. Spectral features are a reflection of the relationship between the shape change of the vocal tract and the articulator movement, and the most widely utilized features are PSD and Mel-scaleFrequency Cepstral Coefficients (MFCC). The four types of features mentioned above are regularly extracted in frames to generate frame-level Low Level Descriptors (LLDs). Meanwhile, they can also be extracted directly using feature extraction tools such as OpenSMILE, COVAREP. A single acoustic feature cannot adequately describe depressive symptoms due to their complexity and variety. As a result, many researchers have developed and mixed various acoustic features to build higher-performance depression recognition models. Shankayi et al. [197] extracted three categories of features from speech signals: prosodic, vocal tract spectrum, and glottal source. The results showed that using features all together leads to better results than using each category alone.

In traditional speech-based depression recognition studies, SVM [59], [198]–[200] and Gaussian Mixture Model (GMM) [200]–[202] are the two most popularly utilized modeling and classification methods. Jiang et al. [203] studied 170 subjects and proposed a computational methodology based on SVM (STEDD). They documented accuracies of 75.96% for females and 80.30% for males. Ooi et al. [204] studied 30 subjects (15 were at risk of depression and 15 were not at risk) and presented an ensemble method using GMM classifiers that used prosodic and glottal features. They reported a classification result of 74%. Alghowinem et al. [202] summarized low-level descriptors and statistical features of 60 subjects (30 controls and 30 depressed patients) and compared four classifiers: GMM, SVM, Hierarchical Fuzzy Signature, and Multilayer Perceptron Neural Network. They concluded that GMM and SVM performed better.

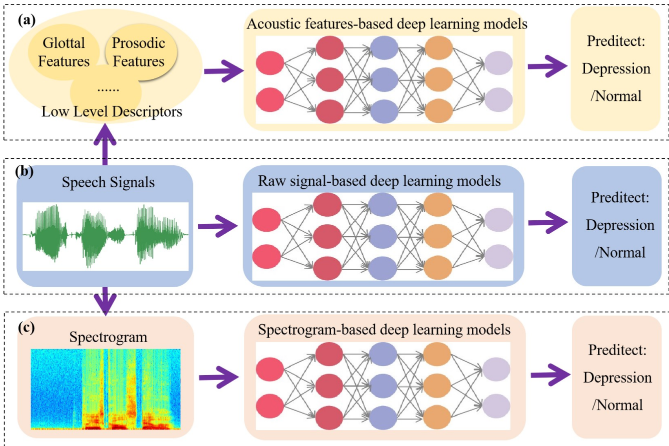


Fig. 13. Flow diagram of speech depression recognition based on deep learning, with the first to third rows representing acoustic feature-based, spectrogram-based, and end-to-end deep learning models, respectively.

2) Speech depression recognition based on deep learning: More recently deep learning methods have showed their capacity in many audio based applications. These methods learn discriminative features through multiple layers and performed better than traditional methods. There are three ways to employ deep learning in this area, depending on the format of the input data.(1)Acoustic features based deep learning model: Traditional acoustic features are put into the deep classifier for training, recognition or prediction; (2)Spectrogram based deep learning models: Spectrograms are given as input to CNN and low level features are extracted from spectrograms where spectrogram is log scale plot of Short time fourier transform. (3)End to end deep learning models: Pushing raw signal into deep architecture to let model learn high-level features by itself. Architecture of these models is shown in Fig.13 .

CNN could capture spatial properties of features and has the ability of parallel computing. Therefore, CNN can be used as a classifier for traditional acoustic features. Well known features, like MFCCs and logMel are fairly simple to extract and have a small number of dimensions which might be more suitable to a low-resource setting than raw signal. Du et al. [205] presented a novel audio-based approach, called IncepLSTM, which effectively integrated Inception module and LSTM on the 16-dimensional MFCCs to capture multi-scale temporal information for Bipolar Disorder recognition. What's more, experiments were conducted on the AVEC 2018 dataset and the results demonstrated the effectiveness of their proposed approach.

The spectrogram converts the speech signal from a 1-dimensional to a 2-dimensional signal, and it not only represents the dynamic spectral properties of the speech signal but also visualizes the speech. There are observable differences between the speech spectrograms of depressed and non-depressed people. In the spectrograms of non-depressed samples, intensity of speech is concentrated more at lower frequencies and low at higher frequencies, whereas, in depressed speech samples, high frequency components also exists with higher intensities and intensity is presented more in short periods of time intervals. Therefore, CNN can use these kind of features from spectrograms to identify depression. Srimadhur et al. [206] carried out an investigation on depression detection using spectrogram based CNN. And speech samples from audio visual emotion challenge (AVEC) 2016 DAIC-Woz dataset were utilized for validating the models. However, most of the input speech features of these studies are based on the amplitude spectrogram, which loses the phase spectrogram information. Therefore, these speech features may lose some important information related to depression. In order to make full use of speech information, Fan et al. [207] proposed a complex squeeze-and-excitation network (CSENet) for SDLP, which used the complex spectrogram as the input feature. The complex spectrogram contains all of the speech information both amplitude and phase spectrogram.

End to end deep architecture have advantages like that it does not require scholars to have a priori knowledge, deep networks can learn better features and give better classification result. Srimadhur et al. [206] proposed spectrogram based CNN and end to end CNN models to estimate the severity level of depression on AVEC 2016 DAIC-woz dataset. Experimental analysis has shown that performance of end to end model is ahead of spectrogram based model and baseline models by an efficiency of 13%.

3) Performance Comparison: We evaluate the depression recognition method based on speech. The performance comparisons based on traditional machine learning and deep learning are summarized in Table VII and Table VIII respectively. From the table we obtain the following three observations:

(1) Comparative analysis of the performances of several classifiers in depression assessment and prediction indicate that the use of an hybrid classifier using GMM and SVM model gave the best overall classification results [202]. Different fusion methods, namely feature, score and decision fusion have been also investigated in [202] and it has been demonstrated that : first, amongst the fusion methods, score fusion performed better when combined with GMM, HFS and MLP classifiers. Second, decision fusion worked best for SVM (both for raw data and GMM models) and finally, feature fusion exhibited weak performance compared to other fusion methods.

(2) Performance of end to end model is better than spectrogram based convolutional neural network model. Srimadhur [208] conducted an experiment on depression detection using spectrogram based CNN and end to end deep models. Parameter tuning has been performed and comparative analysis has been carried out between two models and best model has been chosen for categorizing the depression state. The results indicated that performance of end to end model was better than the baseline models and spectrogram based convolutional neural network model on DAIC-woz dataset. Main reason is because of variability in the volume during recording of data samples and normalized the data samples to reduce this affect but there is no changes in the spectrogram of data samples so this spectrogram based convolutional neural networks for depression detection is ineffective to variance of speaker volume.

(3)So far, the most popular method is still the combination of acoustic features and deep classifiers. As it could solve

Fig. 14. Text on social network .

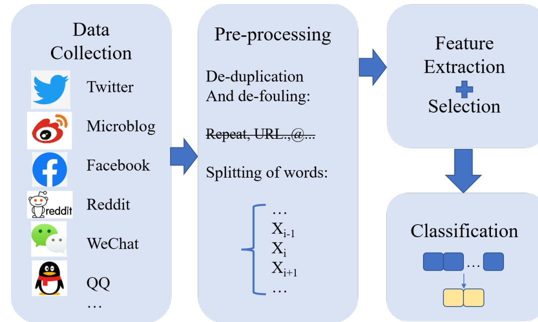


Fig. 15. Flow chart of text-based depression recognition model.

the problems encountered in hand-crafted features, such as high threshold, labour cost and low feature utilization rate, deep learning slowly becomes the leader in the field of machine learning. Although the end-to-end model also has better performance, it is difficult to determine the contribution of each module in the architecture due to its end to end characteristics, limiting further performance improvement.

VI. Depression Recognition Method Based on Text

Although depression has become one of the most concerned psychological problems of human beings, but due to the limited public awareness of depression, and many people do not pay attention to or even reject mental and psychological diseases, they will hide their true inner feelings, resulting in long-term repression of negative emotions can not find a suitable way to vent. The rapid development of Internet technology has built a suitable platform for people to vent their psychological feelings.

In recent years, the popularity of Internet technology has made social media platforms such as Microblog, Twitter and Facebook an important platform for people to express their psychological emotions. Studies have shown that people tend to express their true emotions online more than other ways, and the development of social networks not only provides people with a more convenient way to communicate, but also provides a new window for people to vent their emotions [221]. People can record their life status in real time through social networks and interact with their friends to express their emotions to relieve stress. Several researchers have studied data from users on social networking platforms and found that depressed patients differ significantly from normal users in terms of linguistic attributes and social behavior [222]–[225]. For example, patients suffering from depression use first-person pronouns and past tense verbs more frequently, as well as adjectives with derogatory meanings [226], [227]. The aforementioned study conducted a comparative analysis of language use and social behavior characteristics of depressed and normal individuals under various different social networking platforms and confirmed a strong correlation between social networking activity records and users' depressive status. Therefore, the development of social networks also provides a new way to detect depressed users: through the current computer technology to analyze the user's social network data to detect the user's depression status [228], [229].

As shown in Fig 15, studies related to depression detection based on social media texts usually collect users' behavioral data on social network platforms such as Twitter, Weibo, Facebook, Reddit, or publicly published text content for analysis, while some other studies use relatively private social network data such as WeChat Moments and Qzone. Firstly, the raw data is pre-processed, such as removing other non-target language user data, removing deactivated words, URLs and special characters, etc., and then the sentences are divided into words. The next step is feature extraction and selection of the processed data. For the data selection and related feature engineering aspects, they can be mainly divided into the following aspects: linguistic features, behavioral features, emotional and cognitive features, demographic features, image features, etc. Finally, the attributes obtained by feature selection will be used to identify depressed users in social networks and to detect users with depression from normal users.

1) Text depression recognition based on machine learning: The researchers extracted features such as sentiment, mood and writing behavior of users from different social networking platforms and used various machine learning models for depression prediction. The most applied traditional machine learning method is SVM [230], [231], Peng [232] et al. used a multi-core SVM model for depression identification based on social media data. aldarwish [233] et al. used a plain Bayesian model , SVM models, etc. for depression rank identification and verified the utility of social media sites for depression rank identification. Secondly LR [234], RF [235], [236], etc. are also widely used and Eichstaedt [234] et al. used LR methods to predict depressed users on Facebook. Finally, other classical machine learning classification algorithms such as NB, DT and XGBoost have also been used in related studies.

2) Text depression recognition based on deep learning: Deep learning methods are another approach that has been widely used by researchers in recent years. Among the frequently used methods are DNN, CNN, and RNN [55], [237]–[239]. Shen [240] et al. proposed a cross-domain deep

TABLE VII
OVERVIEW OF TRADITIONAL MACHINE LEARNING BASED METHODS FOR DEPRESSION ASSESSMENT FROM SPEECH.

Paper	Dataset	Classifiers	Metrics	Performance
Ringeval [189]	SEWA	Random forest	RMSE/MAE	7.78/5.72
Low [209]	hand-craftes dataset	SVM+GMM		
Alghowinem [202]	hand-craftes dataset	SVM+GMM+decision fusion	Accuracy	91.67%
Valster [210]	AVid-Corpus	SVR	RMSE/MAE	14.12/10.35
Cummins [211]	AVEC2013	SVM	Accuracy	82%
Meng [212]	AVEC2013	Partial least square regression	RMSE/MAE	11.54/9.78
Williamson [213]	AVEC2014	GMM	RMSE/MAE	8.50/6.52
Nasir [59]	DAIC-WOZ	SVM	F1	63%
Gong [198]	DAIC-WOZ	SVM	RMSE/MAE	4.99/3.96
Jayawardena [214]	DAIC-WOZ	LR	RMSE	6.84
Valstar [187]	DAIC-WOZ	SVM + grid search +random forest	RMSE/MAE	7.78/5.72

TABLE VIII
OVERVIEW OF DEEP LEARNING BASED METHODS FOR DEPRESSION ASSESSMENT FROM SPEECH.

Paper	Dataset	Features	Classification	Metrics	Performance
Kang [215]	AVEC2014	LLDs	DNN	RMSE/MAE	7.37/5.87
Yang [35]	DAIC-WOZ	LLDs	DCNN	RMSE/MAE	5.97/5.16
Al Hannai [216]	DAIC-WOZ	LLDs	LSTM-RNN	RMSE/MAE	10.03/7.60
Dham [217]	DAIC-WOZ	LLDs	FF-NN	RMSE/MAE	7.63/6.28
Salekin [218]	DAIC-WOZ	LLDs	NN2Vec + BLSTMMIL	F1-score	85.44%
Zhang [219]	DAIC-WOZ	Spectrogram	Transformer	RMSE/MAE	5.73/4.75
Othmani [220]	DAIC-WOZ	LLDs+ spectrogram	LSTM	F1-score	82%
Srimadhur [208]	DAIC-WOZ	Raw signal	CNN	F1-score	78%
Srimadhur [208]	DAIC-WOZ	Spectrogram	CNN	F1-score	66%
Ma [29]	DAIC-WOZ	Spectrogram	Depaudionet	F1-score	52%

neural network model with feature adaptive transformation and combination strategy (DNN-FATC) to transfer relevant information to a heterogeneous domain, using sufficient Twitter data as the source domain and enhanced detection in some other target domain (e.g., Weibo). Rao [241] et al. built a multilayer MGL-CNN to further identify depressed individuals in online forums by modeling them separately at sentence level and user level. LSTM is a special type of RNN that learns long-term dependent information and has also been widely used in depression detection based on social media texts. Hu [242] et al. proposed a Bi-LSTM-based depressive tendency detection model for microblog users. The content features of the microblog text were mined by bi-directional transmission and capturing the semantic dependencies of the context.

3) Performance Comparison: .

TableIX summarizes the experimental results of the social platform text-based depression detection, including the source and name of the method, the type of data used and the source of the dataset (- indicates that the data used were collected by themselves), and the evaluation criteria of the experimental results including Precision, Recall, F1 Score, Accuracy, and AUC.

Text-based depression detection mostly uses social network text. In machine learning methods, using related methods such as logistic regression and Bayes can achieve better results than using SVM alone, and with the rise of deep learning algorithms, better results than the above traditional machine learning are usually achieved.

A. Gait

Human gait, as a daily movement, occurs in parallel with the development of higher brain structures and functions (prefrontal cortex, basal ganglia, and cerebellum) and reflects the integrity of the higher brain systems [243]; thus, it is a good indicator of mental status. Compared to traditional mental illness detection biometrics, such as facial expression, speech, and physiological parameters, gait is remotely observable, more difficult to imitate, and requires less cooperation from the subject [244]. These advantages make gait a promising source for depression recognition.

Furthermore, researchers have also studied the gait characteristics of depression patients through motion analysis [245]–[250]. Specifically, Lemake et al. [251] calculated the spatiotemporal gait parameters of patients with major depressive disorder and found that the patients show significant reductions in stride length, cycle time, and lower limb support. Moreover, by using the motion capture system with video cameras, Michalak [252] found reduced walking velocity, arm swing, vertical body movement, increased body sway, and more slumped posture in patients. Thanks to the development of consumer-grade depth cameras [253], [254] (e.g., Kinect [255], the data collection scheme is shown in Fig 16), human

TABLE IX
EXPERIMENTAL RESULTS BASED ON TEXT

ID	Method	Data		Metrics				
		Media	Dataset	Precision	Recall	F1-Score	Accuracy	AUC
1	SVM [228]	Twitter	-	0.74	0.63	-	0.70	-
2	Naïve Bayes [227]	Twitter	CLPsych 2015	0.82	0.82	0.81	0.86	0.94
3	Logistic Regression [227]	Twitter	CLPsych 2015	0.86	0.82	0.84	0.82	0.91
4	Multi-kernel SVM [232]	Weibo	-	0.76	0.77	0.76	0.83	-
5	SVM+ Naïve Bayes [233]	Facebook, LiveJournal, Twitter	-	1.00	0.57	-	0.63	-
6	LDA(latent dirichlet allocation) [234]	Facebook	-	-	-	-	-	0.72
7	MGL-CNN [241]	Reddit	RSDD	0.63	0.48	0.54	-	-
8	MGL-CNN [241]	Online media	eRisk 2017	0.63	0.57	0.60	-	-
9	CNN [55]	Reddit	RSDD	0.75	0.57	0.65	-	-
10	MDL [237]	Twitter	-	0.8+	0.8+	0.85	0.8+	-
11	DNN-FATC [240]	Twitter+ Weibo	-	-	-	0.79	-	-
12	CNNWithMax [238]	Twitter	CLPSych 2015	0.87	0.87	0.87	0.88	0.95
13	MultiChannelCNN [238]	Facebook	Bell Let's Talk	0.81	0.84	0.82	0.83	0.92
14	Bi-LSTM [242]	Weibo	-	-	-	-	0.95	-
15	LIWC [225]	Reddit	-	-	-	0.68	0.79	0.75

body dynamics can be traced accurately in a three-dimensional manner without the assistance of the attached markers or the requirement of a specifically designed environment. The recorded depth and skeleton data have been widely and successfully used in depression recognition. Therefore, in the current studies of depression recognition, the main work is to learn gait features from the skeleton coordinates recorded by Kinect related with depression and explore feature sets for better performance. In the meantime, traditional machine learning algorithms are employed to classify such as SVM, KNN, RF, etc. However, deep learning methods have made breakthroughs in Computer Vision (CV) research. Therefore, many studies are shifting from the traditional hand-crafted gait features to the framework based on deep learning for gait depression recognition.

The two gait representations - the skeleton and the silhouette - that serve as the foundation of the study must first be explained before introducing the two categories of approaches indicated above. However, the skeleton-based structural model is currently more widely used in gait depression recognition. It can estimate the spatial location information of human joint points based on the human body structure [256]. Moreover, it is insensitive to changes in clothes, changes in carrying things, etc [257]. The researchers proposed skeleton features describing the walking process for the skeleton modality, including spatial-temporal, time-domain, and frequency-domain features [258]. On the other hand, gait energy image (GEI) is a widely used silhouette feature [259]. Benefiting from the cycle length-based time and space normalization, GEI has strong anti-noise capabilities while reducing calculation and memory without loss of performance [260].

1) Gait depression recognition based on traditional machine learning: Gait depression recognition is still in its infancy, with most approaches focusing on traditional machine learning. As shown in Fig 17, researchers extracted the spatial-temporal, time-domain, and frequency-domain features from the skeleton and input them into machine learning classifiers (such as SVM) for depression recognition.

In traditional research of gait depression recognition, hand-crafted features are regularly used together with statistical features. The commonly used gait features are as follows:

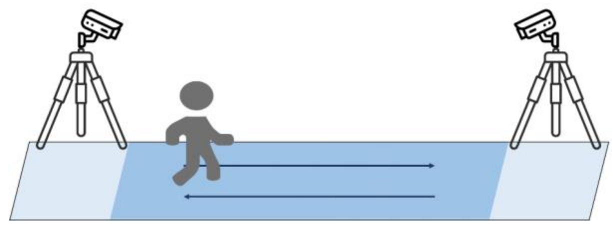


Fig. 16. Video based gait acquisition system.

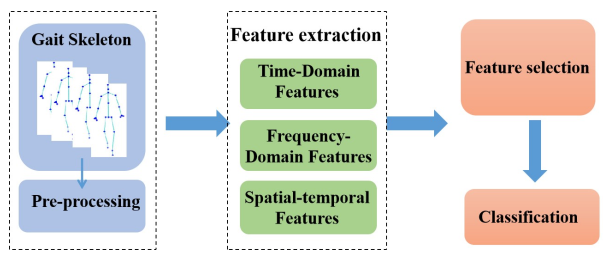


Fig. 17. Flow chart of gait depression recognition based on traditional machine learning.

(1) Time-Domain Features: Gait time-Domain features include mean, standard deviation, skewness, and kurtosis of the original data. Specifically, the mean is a measure of the central tendency of the random variable characterized by that distribution. Standard deviation measures the amount of dispersion of a dataset. Kurtosis measures outliers present in the probability distribution. (2)Frequency-Domain Features: Continuous motion trajectories can be considered as signals. By performing frequency analysis method on them, the frequency features of gait can be obtained. The discrete Fourier transform, PSD [261], and Hilbert-Huang transform are commonly used techniques for translating time domain data to frequency domain signals. (3)spatial-temporal Features: Spatiotemporal gait parameters, including gait velocity, stride length, cadence, stance phase, swing phase etc., provide basic data characterizing a subjects gait.

Previous studies showed that gait-related features could predict depression accurately, and Kinect provided objective and easily accessible data. Wang et al. [262] used skeleton

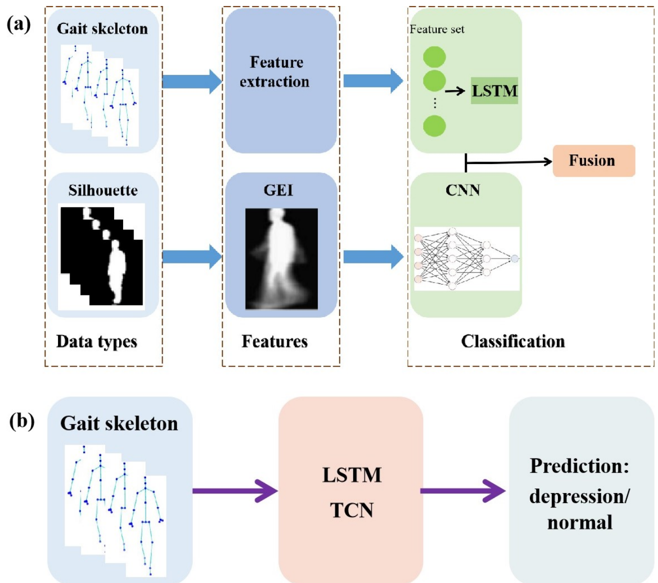


Fig. 18. Two different approaches for using deep learning models in gait depression recognition: (a) Building a structure combined skeleton feature set and silhouette features with deep learning method; (b) Building an end-to-end deep architecture.

data estimated by Kinect devices to predict depression risk in the dataset of 43 scored-depressed and 52 non-depressed individuals. They combined the time domain information, frequency domain information, and spatial geometric features of gait information. The experimental results show that spatial features help a lot in evaluating depression.

Traditional classification or regression algorithms are employed after feature extraction such as SVM, LR, RF, Decision Tree, GMM, KNN, etc. Li [263] used the Kinect V2 device to record kinematic skeleton data of the participants' 25 body joints, the presented spatial features and low-level features is directly extracted from the record original Kinect-3D coordinates. The scored-depressed and non-depressed individuals can be well classified by computational models which were import processed data directly. The proposed experiment demonstrated four strong machine learning tools: SVM, LR, RF and Gradient Boosting (GB).

2) Gait depression recognition based on deep learning: Due to the successful application in CV, deep learning is introduced to gait depression recognition. Unlike traditional methods, once the model and parameters are established, no further human involvement is required. The essence of deep learning is to learn high-level features automatically by building more hidden layer models to improve the accuracy of classification or score prediction.

According to Fig 18, there are two ways to employ deep learning in this field: (1) Build a structure that combines skeleton feature set and silhouette features with the deep learning method. The key to the skeleton study is the data quality of joint coordinates, which is easily affected by distance, light, and other environments. The skeleton estimation is not accurate in complex scenes. The difficulty of silhouettes study is the cross-view recognition, and the recognition results vary greatly under different views. Multimodal fusion can improve the recognition accuracy and generalization performance of the model by exploiting the correlation between each modality. (2) Build an end-to-end deep architecture and then push skeleton sequence into deep architecture to let model learn high-level features by itself.

By merging skeleton characteristics with silhouette features(GEI) for depth feature identification, significant performance is achieved in the detection of gait depressions, and the generalization performance of the model can be enhanced. Shao [264] proposed a multi-modal depression recognition method by combining skeleton data and silhouette data of gait. Firstly, for skeleton data, the skeleton feature set for depression recognition was proposed, which included spatiotemporal features and kinematics features. For silhouette data, a CNN model with a new loss function was designed to extract silhouette features. Finally, they merged the silhouette features of the front and side views; then in the decision level, they fused the classification results of skeleton data and silhouette data.

Compared with methods of performing feature extraction and classification separately, an end to end deep architecture pushes raw signal into its model to learn and give results. End to end deep architectures have advantages like that it does not require scholars to have a priori knowledge, deep networks can learn better features and give better classification result. However, there are a few issues which limit these architectures, such as large-scale data supporting, over-fitting easily, and poor interpretability. Specifically, there is often a lack of relevant experimental data or difficulty in collecting large amounts of data in the actual research process. Data size is gradually emerging as one of the constraints restricting further research on automatic depression recognition because models perform poorly with less data. Data augmentation is a widely used technique in image processing that, to a certain extent, can address the over-fitting issue brought on by insufficient data [265]–[267]. It includes a set of techniques that can augment the size and quality of training datasets, and it has been proved to be adequate for improving the performance of the deep learning models [268]. Yang [269] proposed a skeleton data augmentation method based on Kinect V2 to evaluate the risk of depression. First, Yang proposed five techniques to enhance skeleton data and applied it to depression and emotion datasets. Then, according to mutual information and classification accuracy, the augmentation methods are divided into two categories (non-noise augmentation and noise augmentation). The results showed that this data augmentation method greatly improves the recognition effect of the classifier.

3) Performance Comparison: To shed more light on the performance of gait depression recognition methods, we summarize the performance of the methods tested in Tabel X. To perform a fair comparison, the experimental classification accuracy results (%) are directly derived from the corresponding original papers. It should be noted that we concentrate on contrasting the performance of various approaches on the same data set. Several observations can be summarized as follows:

(1) For traditional gait depression recognition methods, scholars have been working on exploring novel gait features and using traditional machine learning models for classifi-

cation with increasing accuracy. Specifically, first, time- and frequency-domain features are more efficient in constructing computational models to identify depression than spatiotemporal features [262]. It is worth noting that the models built of time- domain features and spatiotemporal and time- domain features have the same performances, which suggests that spatiotemporal features had very few contributions to recognize depression. It is also reflected in the fact that the model comprised of time- and frequency-domain features has the same performance as the model consisted of all features. Second, different classifiers display varying performance when given the same feature vector as input.

(2) For deep learning-based gait depression recognition methods, the multi-modal models are more accurate than the single-modal models [264]. Researchers found that the skeleton can help to locate the spatial position of joints of the human body, and the silhouette can describe the detailed information of the human body shape so that the constructed 3D model is more vivid. Shao obtained better classification performance using skeleton and silhouette fusion, which agrees that the skeleton and silhouette information are related and complementary.

(3) The gait depression recognition method based on deep learning outperforms traditional machine learning for the same dataset [269]–[271]. In previous studies, scholars have been committed to exploring new gait features and using traditional machine learning models for classification, and the accuracy has been continuously improved. However, traditional machine learning technologies were limited in their ability to process natural data in their raw form, and lack the ability to express complex functions, making it difficult to solve more complex natural signal processing problems. Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction, and demonstrate a powerful ability to learn the essential characteristics of a data set from a small number of samples.

VII. Depression Recognition Method Based on Multi-Modal Data

As described in Section 3, most studies have been conducted to detect depression based only on single physiological indicators, such as EEG, EM, speech signals, facial expressions, text, etc. Although single indicators can identify depression to some extent, they do not give a comprehensive representation. For example, in case of a depressed person, when a stimulus is presented, speech signals showed a longer response time and lower pronunciation rate [275], while EM showed decreased eyebrow movement and elevated blink rates [276]–[278]; thus, it is considerably likely that both modalities are correlated. It seems obvious that the signal from single modality provided only partial information, while a combination of different modality signals can be used to form a more realistic model for recognizing depression than the former [279]. So, there is increasing interest in using different modalities to handle information.

In short, a more accurate and robust depression recognition model can be constructed by integrating complementary information from different modalities that reflect different aspects of depression. In this section, we will generally divide depression detection based on multi-modal data into three categories: depression recognition method based on multiple electrophysiological signals, depression recognition method based on multiple brain imaging, and depression recognition method based on multiple audio and video data.

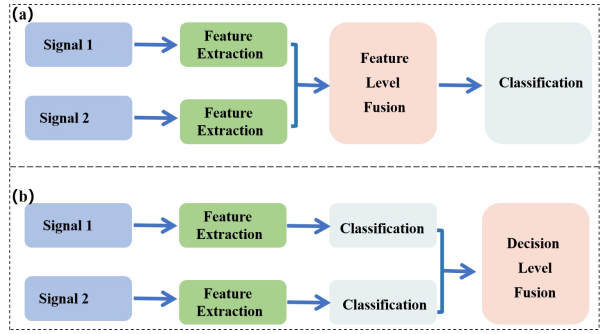


Fig. 19. Flow chart of fusion based on multiple electrophysiological signals (taking two modalities as an example): subfigures a and b represent feature level fusion and decision level fusion, respectively.

A. *Depression recognition method based on multiple electrophysiological signals*

Various studies have shown that electrophysiological signals such as EEG, ECG, EOG, and EMG can reflect a person's physiological and emotional state to varying degrees, which is helpful for clinically assisted diagnosis of depression [280]. Further, electrophysiological signals not only have the benefits of good timing and convenient operation, but also have higher generalizability and stability of the developed depression recognition model [281]. Therefore, with the continuous development and update of machine learning, research into the depression recognition using multiple physiological signals has increasingly attracted attention. The researches generally adopt feature level fusion or decision level fusion in depression recognition.

1) Feature level fusion: Feature level fusion, which can be divided into serial and parallel forms, is the fusion of the data that results from feature extraction [282], [283]. Feature level fusion aims to find correlations between data from the extracted features or to evaluate the modalities, which can reduce noise interference and improve model efficiency to some extent. Taking depression recognition based on multiple electrophysiological signals as an example, the specific process of feature level fusion is shown in Fig 19. In order to identify depressions, Zhu et al. [284] proposed a model based on feature level fusion. And they validated the presented classification model by collecting two physiological signals, EEG and EM. The experimental result indicated that the feature fusion method slightly improved the recognition accuracy by 1.88%, compared with the unimodal classification approach that uses only EEG or EM. Thus, it was concluded that the feature level fusion methods can improve the mild depression recognition accuracy, demonstrating the complementary nature of the modalities.

TABLE X
OVERVIEW OF MACHINE LEARNING BASED METHODS FOR DEPRESSION ASSESSMENT FROM GAIT.

Methods	Paper	Dataset	Feature	Classification	Metrics	Value
Gait depression recognition based on traditional machine learning	Li [263]	85SD+85C	skeleton sequential data	SVM	Accuracy	53.85
				RF		61.54
				LR		73.08
				GB		76.92
	Wang [270]	43SD+52C	Velocity	FC+Softmax	Accuracy	88.84+7.13
			Angle Velocity			83.93+7.33
			TF(Pseudo-Velocity Model)			91.96+5.33
			SG			70.98+11.85
			TF+SG			93.75+2.98
	Lu [272]	43SD+52C	the joint energy feature	KNN	Accuracy	80.34
				SVM		91.21
				RF		85.71
	Yuan [273]	54SD+47C	FF	SVM	Accuracy	91.09
	Fang [271]	43SD+52C	walking speed arm swing,stride length, vertical head movement, body sway,step width, joint rom,stance duration	SVM	Accuracy	90.53
				RF		91.58
				KNN		87.37
				LR		88.42
				LDA		88.42
	Wang [262]	126SD+121C	SF	SVM	AUC	0.58
			TF			0.83
			FF			0.87
			SF+TF			0.83
			SF+FF			0.83
			TF+FF			0.93
			ALL			0.93
	Zhao [274]	179	FFT	SVR	Accuracy	64
Gait depression recognition recognition based on deep learning	Shao [264]	86SD+114C	Skeleton features set,GEI	LSTM+CNN	Accuracy	85.45
			GEI			66.67
			skeleton feature set			80.28
	Yang [269]	43SD+52C	skeleton data	LSTM	Accuracy	92.15

Score-depressed: SD; Time-Domain Features: TF; Frequency-Domain Features: FF; Spatial geometric feature: SG; spatial-temporal Features: SF

2) Decision level fusion: Decision level fusion is the process of taking the information obtained from each modality and making independent decisions, and then fusing the outcomes of these decisions in some ways. Taking depression recognition based on multiple electrophysiological signals as an example, the specific process of decision level fusion is shown in Fig 19. Among the different fusion strategies, decision level fusion has unique advantages to fuse the output of various classifiers and getting an effective result. It can also synthesize multi-source information and avoid mutual interference. However, the classification results of each modality are usually not completely reliable, and there are misclassifications. Moreover, the classification results from different modalities of one object may have high conflict, and this is very unfavorable to the fusion result. To solve these problems, Zhu et al. [285] proposed a content-based multiple evidence fusion (CBMEF) method, which fused EEG and EM data at decision level. The method mainly included two modules, the classification performance matrix module and the dual-weight fusion module. The classification performance matrices of different modalities were estimated by Bayesian rule based on confusion matrix and Mahalanobis distance, and the matrices were used to correct the classification results. Then the relative conflict degree of each modality was calculated, and different weights were assigned to the above modalities at the decision fusion layer according to this conflict degree. The idea of introducing the classification performance matrix and the dual-weight model to multimodal biosignals fusion casts a new light on the researches of depression recognition.

B. Others

In addition to the above three types of multimodal fusions, there are a number of other modal fusions that have also attracted extensive attention from researchers. For example, Bruder et al. [286] combined high temporal resolution EEG signals with MRI data to increase the precision of identifying depression. Zhang et al. [287] explored from physiological and behavioral perspectives simultaneously and fused pervasive EEG and speech signals to make the detection of depression more objective, effective and convenient. Jing et al. [288] built a supervised regression model based on personal language and natural gait data to predict depression and anxiety in patients. Their results could be a basis of both applications and future studies on the multi-source data fusion in anxiety and depression recognition.

C. Depression recognition method based on multiple brain imaging

The onset of depression is related to biochemical, genetic, social and environmental factors and involves abnormalities in a variety of neurotransmitters, brain regions and loops. Brain imaging can directly reflect the abnormalities in the brain of depressed patients, and the analysis of brain imaging data can be used as an important tool to detect depression. Each type of brain imaging can reflect the depressed patient's brain, such as sMRI can clearly demonstrate the anatomical structure of the brain with high spatial resolution and reflect the structural morphology within the brain, while fMRI detects changes in the blood-oxygen-dependent level signal of the brain and can reflect the active condition of the neural regions of the brain. fMRI has a higher temporal resolution compared to sMRI, and the corresponding spatial sub The corresponding spatial representation is reduced.

In recent years, most studies have been based on single-modality brain imaging data for depression detection, and few studies have combined multiple brain imaging data from different modalities for analysis. However, combining multi-modal brain imaging data can compensate for each other's deficiencies caused by different imaging reasons and can better detect depression. For example, Mousavian et al. [289] considered the similarity of spatial cubes from sMRI and RS-fMRI data, from which features were extracted and fed into a unified machine learning classifier framework for depression detection.

D. Depression recognition method based on multiple audio-visual data

Depressed patients usually show some abnormal behavioral expressions, such as becoming sluggish in facial expressions,

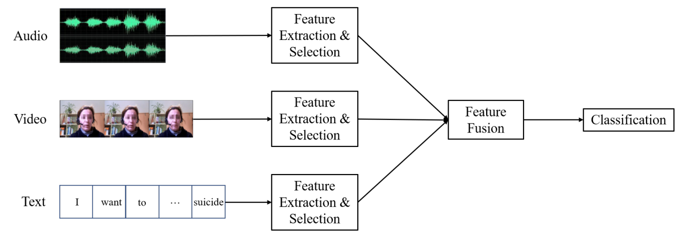


Fig. 20. Flow chart of depression detection based on multimodal behavioral data.

frequently avoiding eye contact with others, using short sentences with a flat tone when speaking, and having slow movements of the head and body parts. However, these depressive features may also appear when a healthy person is in a bad mood, and not all depressed patients directly show the above abnormal depressive symptoms, which means that depressed patients do not have a certain characteristic unique to them, so it is difficult to accurately detect depression by relying on a single modal characteristic. In recent years, there have also been many mature studies on audio-video based multimodal fusion methods in depression detection [290], which have more comprehensive feature coverage and enhanced performance in depression detection compared with unimodal audio and visual cue detection methods.

Machine learning methods play a key role in multimodal depression detection. He et al. [291] introduced the median robust LBP-TOP (MRLBP-TOP) descriptor that can learn patterns on different scales from image sequences for depression detection. In addition, Sadari et al. [292] used ordinal logistic regression for depression identification and proposed a new approach for the task.

With the development of deep learning, more and more researchers have applied deep learning-based methods to the task of depression detection based on multimodal data, such as CNN, DNN, LSTM, etc [**?**], [293], [294]. Yang et al. [35] designed a hybrid multimodal depression detection model based on audio-video and text data, containing an audiovisual multimodal depression recognition framework based on deep convolutional neural networks (DCNN) and deep neural networks (DNN) to predict depression severity defined by depression scales. Furthermore, He et al. [295] proposed a new temporal attention (STA) architecture and multimodal attentional feature fusion (MAFF) method for learning and extracting information and features between audio and video cues to predict subjects' scores on depression scales.

To advance the development of depression detection, AVEC started to introduce the task of depression identification in 2013 to detect depression through collected audio, video, and text data, among others. In 2018, AVEC introduced the bipolar disorder (BD) sub-challenge for researchers to detect bipolar depression, based on which, Yang et.al [296] proposed a new architecture incorporating DNN and random forest for bipolar depression analysis. Du et al. [205] designed the IncepLSTM model, which combines the initial module of feature sequences and LSTM with learning multi-scale temporal patterns for BD analysis.

TABLE XI
EXPERIMENTAL RESULTS BASED ON MULTI-MODAL

ID	Multi-modal	Method	Dataset	Metrics	
				RMSE	MAE
1		Y ang et al. [301]	DAIC-WOZ	5.97	5.16
2		Y ang et al. [35]	DAIC-WOZ	5.40	4.35
3		Y ang et al. [293]	DAIC-WOZ	6.34	5.38
4		Y ang et al. [294]	DAIC-WOZ	6.34	5.39
5		Shi et al. [299]	E-DAIC	5.50	-
6	Audio+Video+Text	Makiuchi et al. [300]	E-DAIC	6.11	-
7		Fan et al. [297]	E-DAIC	5.91	4.39
8		Zhang et al. [298]	E-DAIC	6.85	5.84
9		Zhao et al. [302]	DAIC-WOZ	5.51	4.20
10		Williamson et al. [188]	6th	5.31	3.34
11		Gong et al. [198]	DAIC-WOZ	4.99	3.96
12		AVEC2016(baseline)	DAIC-WOZ	6.62	5.52
13	Audio+Text	Alhanai et al. [216]	-	6.27	4.97
14		AVEC2013(baseline)	AVEC2013	13.61	10.88
15	Audio+Video	AVEC2014(baseline)	AVEC2014	10.86	8.86
16		Niu et al. [303]	AVEC2013	7.03	5.21
17		Jan et al. [304]	AVEC2014	7.43	6.14
18		Melo et al. [178]	AVEC2013	8.25	6.30
19		Melo et al. [178]	AVEC2014	8.23	6.15

Moreover, audio and video data can also be combined with textual data want for depression detection [188], [198]. Hanai et al. [216] simulated the interaction with audio and text features in an LSTM neural network model to detect depression. Albert et al. [208] combined data from three modalities: audio, 3D video of key points of the face, and text transcriptions of patients speaking in clinical interviews. Sentences were summarized into individual embedding models using causal convolutional networks (C-CNN) for depression classification.

In the Detecting Depression with AI SubChallenge (DDS) of A VEC2019, predict the presence and severity of depression in individuals through digital biomarkers such as vocal acoustics, verbal content of speech, and facial expressions. Provides additional opportunities for multimodal depression detection. [297], [298] Yin et al. [299] proposed a multimodal approach with hierarchical recurrent neural structure to integrate visual, audio, and text features for depression detection. Makiuchi et al. [300] designed separate models for speech and text data for analysis, and for speech patterns used deep spectral features extracted from a pre-trained VGG-16 network with a gated convolutional neural network (GCNN) followed by an LSTM layer. For text embedding BERT text features are extracted and Convolutional Neural Network (CNN) and LSTM layers are used. Finally the two modalities are combined using feature fusion and the experiments show that the multimodal fusion approach is better than the unimodal one.

E. Performance Comparison

Table XI summarizes the results of the experiments on the detection of depression based on multi-modal data, including the source and name of the method, the data set used (- means the data used were collected by ourselves), the evaluation criteria of the experimental results, including the evaluation criteria of depression severity by depression scale mean absolute error (MAE) and root mean square error (RMSE).

VIII. CONCLUSION AND FUTURE WORK

A. Conclusion

1) Brain imaging: Multimodal neuroimaging techniques have revealed abnormalities in brain structure, brain function, and brain metabolism in depressed patients, providing new ideas for the early diagnosis of depression and optimization of treatment options. Currently, most studies are still based on unimodal structural and functional brain imaging, such as MRI, fMRI, fNIRS, etc., and the results of these studies demonstrate the usefulness of various types of brain imaging for classifying depressive symptoms. However, studies based on PET, MRS and other brain metabolism and multimodal brain imaging are still in the minority, but these aspects still have a great potential for application in the diagnosis of depression, waiting for researchers to explore.

Current research on brain imaging-based detection of depression still has some limitations: (1) The limitations of the current studies are the small number of training samples and unbalanced data sets. (2) The studies on brain structure, function and metabolism are relatively independent, and too much differentiation in the studies should be avoided; they can be combined with each other as an overall study to explore the important connections of important brain regions in structure, function and metabolism. (3) At present, there are few studies on the imaging differences of brain structure, brain function and brain metabolism between patients with non-refractory depression and patients with refractory depression, and most of them are cross-sectional studies, and there is an urgent need for in-depth longitudinal studies with large samples.

2) Facial expressions: A person's personality and mood can be seen in their facial features, and some studies have found that the looks and temperament of people with depression can differ from those of the general population. The increasing social concern about depression and the valuable data provided by visual challenge competitions such as AVEC have both accelerated to some extent the development of studies based on facial features of depressed patients in aiding the identification and diagnosis of depression. As a result, research on depression recognition based on facial expression behavior has become a hot topic. At the same time, because facial features are so rich in variation, analysis of data from only a single point in time cannot achieve accurate diagnosis of depression, so most studies focus on the analysis of the temporal dimension to improve detection accuracy. In future studies, more facial features can be combined with other physiological and behavioral features to improve the detection performance.

3) Text: With the development and popularity of the Internet, it has made various social media platforms an important platform for people to express their psychological emotions. By comparing and analyzing the effects of detecting depressed users based on Twitter, Weibo, Facebook, Reddit and other data in related literature, it was demonstrated that it is feasible to detect depressed users using text-based social network data. From the experiments and comparison with related work, it is clear that dataset, features, and detection models are all key to the detection of depressed users based on social network data, and there are large differences in these aspects across the literature. In addition, the higher accuracy rates in the existing work are obtained on balanced samples, which differ significantly from the actual distribution of depressed users and imply that there are still many challenges to be faced in the practical application.

There are still some limitations in the current research on social text-based depression detection: (1) most of the studies are based on public social network data, due to privacy issues, few studies are using private social network data such as Qzone and WeChat friend circle for depressed user detection, while compared to public social networks these private data contain more user emotions and have great potential for depressed user detection. The private data contains more user emotions than public social networks, and has great potential for depressed user detection. (2) Most of the literature removes expressions and special symbols in preprocessing to reduce the noise of data, but many users are used to express their emotions with these symbols, and this information is also effective for user psychological detection, so the model can be further improved and performance enhanced by the fusion and utilization of multidimensional features. (3) Similarly, combining information such as images, audio and video posted by users in social networks might be better for depressive tendency prediction.

4) Multi-modal data: Depression is a common and highly prevalent mental disorder, and the existing assessment process is complex and relatively subjective, and its effective diagnosis needs to be addressed. Depression is not only traumatic to the patient's mental health, but also to his or her appearance and behavior, and the accuracy of depression detection can be further improved by combining these abnormalities. For example, combining structural and functional brain imaging can yield high-resolution and well-structured brain data, while audio and video data can be combined with patients' facial expressions and voice expressions to better analyze the emotions expressed by patients, etc.

B. Future Work

Depression is the fourth largest disease in the world, but the medical prevention and treatment of depression is still in a situation of low recognition rate in China. The recognition rate of hospitals above prefecture level cities is less than 20%, and less than 10% of patients receive relevant drug treatment; Moreover, at the same time, the incidence of depression (and suicides) has begun to show a trend of younger age (universities, and even primary and secondary school students). How to find an efficient and rapid diagnostic method becomes imminent. This paper summarizes the research and application of various modal data in the diagnosis of depression, and provides a reference for designing a more convenient and fast clinical auxiliary diagnosis method.

Great progress has been made in the detection of depression based on each modality, but many studies still have certain limitations: (1) they ignore the changes of patients in the temporal dimension and do not take into account the effects of medication and other factors on the patient's body. (2) Most studies use scales to determine depression labels for diagnosed by professional physicians and lack reliability. (3) The data are usually collected by machine, recording the communication between human and machine, rather than collected from human-to-human conversation, which may not match the responses in real scenarios. (4) There is no more breakdown of depression categories.

In the future, in order to obtain more reliable results of depression identification, researchers should consider collecting more data from more angles to analyze the diagnosis of depression. It is also hoped that multimodal depression identification can help more patients with depression get early intervention and treatment, and provide more effective help for the rehabilitation of patients with depression.

References

[1] W. H. Organization *et al.*, "Depression and other common mental disorders: global health estimates," World Health Organization, Tech. Rep., 2017.

[2] D. Bhugra and A. Mastrogianni, "Globalisation and mental disorders: overview with relation to depression," *The British Journal of Psychiatry*, vol. 184, no. 1, pp. 10–20, 2004.

[3] T. S. Rao, M. Asha, B. Ramesh, and K. J. Rao, "Understanding nutrition, depression and mental illnesses," *Indian journal of psychiatry*, vol. 50, no. 2, p. 77, 2008.

[4] Q. Wang, "The social determinants of depressive disorders in china," *Lancet psychiatry*, 2021.

[5] R. Peveler, A. Carson, and G. Rodin, "Depression in medical patients," *Bmj*, vol. 325, no. 7356, pp. 149–152, 2002.

[6] A. Singh and N. Misra, "Loneliness, depression and sociability in old age," *Industrial psychiatry journal*, vol. 18, no. 1, p. 51, 2009.

[7] D. Eisenberg, S. E. Gollust, E. Golberstein, and J. L. Hefner, "Prevalence and correlates of depression, anxiety, and suicidality among university students," *American journal of orthopsychiatry*, vol. 77, no. 4, pp. 534–542, 2007.

[8] I. O. Bergfeld, M. Mantione, M. Figee, P. R. Schuurman, A. Lok, and D. Denys, "Treatment-resistant depression and suicidality," *Journal of affective disorders*, vol. 235, pp. 362–367, 2018.

[9] L. Hemming, G. Haddock, J. Shaw, and D. Pratt, "Alexithymia and its associations with depression, suicidality, and aggression: an overview of the literature," *Frontiers in psychiatry*, vol. 10, p. 203, 2019.

[10] Y. Huang, Y. Wang, H. Wang, Z. Liu, X. Yu, J. Yan, Y. Yu, C. Kou, X. Xu, J. Lu *et al.*, "Prevalence of mental disorders in china: a cross-sectional epidemiological study," *The Lancet Psychiatry*, vol. 6, no. 3, pp. 211–224, 2019.

[11] D. Faust and J. Ziskin, "The expert witness in psychology and psychiatry," *Science*, vol. 241, no. 4861, pp. 31–35, 1988.

[12] N. Cheng and S. Mohiuddin, "Addressing the nationwide shortage of child and adolescent psychiatrists: determining factors that influence the decision for psychiatry residents to pursue child and adolescent psychiatry training," *Academic psychiatry*, vol. 46, no. 1, pp. 18–24, 2022.

[13] T. Butryn, L. Bryant, C. Marchionni, F. Sholevar *et al.*, "The shortage of psychiatrists and other mental health providers: causes, current state, and potential solutions," *International Journal of Academic Medicine*, vol. 3, no. 1, p. 5, 2017.

[14] B. K. Finch, B. Kolody, and W. A. Vega, "Perceived discrimination and depression among mexican-origin adults in california," *Journal of health and social behavior*, pp. 295–313, 2000.

[15] C. Lauber, C. Nordt, and W. Rössler, "Recommendations of mental health professionals and the general population on how to treat mental disorders," *Social psychiatry and psychiatric epidemiology*, vol. 40, no. 10, pp. 835–843, 2005.

[16] M. C. Angermeyer, A. Holzinger, M. G. Carta, and G. Schomerus, "Biogenetic explanations and public acceptance of mental illness: systematic review of population studies," *The British Journal of Psychiatry*, vol. 199, no. 5, pp. 367–372, 2011.

[17] B. A. Pescosolido, T. R. Medina, J. K. Martin, and J. S. Long, "The "backbone" of stigma: identifying the global core of public prejudice associated with mental illness," *American journal of public health*, vol. 103, no. 5, pp. 853–860, 2013.

[18] J. Luo, M. Wu, D. Gopukumar, and Y. Zhao, "Big data application in biomedical research and health care: a literature review," *Biomedical informatics insights*, vol. 8, pp. BII–S31 559, 2016.

[19] D. M. Schnyer, P. C. Clasen, C. Gonzalez, and C. G. Beevers, "Evaluating the diagnostic utility of applying a machine learning algorithm to diffusion tensor mri measures in individuals with major depressive disorder," *Psychiatry Research: Neuroimaging*, vol. 264, pp. 1–9, 2017.

[20] J. R. Sato, J. Moll, S. Green, J. F. Deakin, C. E. Thomaz, and R. Zahn, "Machine learning algorithm accurately detects fmri signature of vulnerability to major depression," *Psychiatry Research: Neuroimaging*, vol. 233, no. 2, pp. 289–291, 2015.

[21] R. Ramasubbu, M. R. Brown, F. Cortese, I. Gaxiola, B. Goodyear, A. J. Greenshaw, S. M. Dursun, and R. Greiner, "Accuracy of automated classification of major depressive disorder as a function of symptom severity," *NeuroImage: Clinical*, vol. 12, pp. 320–331, 2016.

[22] B. Vai, L. Parenti, I. Bollettini, C. Cara, C. Verga, E. Melloni, E. Mazza, S. Poletti, C. Colombo, and F. Benedetti, "Predicting differential diagnosis between bipolar and unipolar depression with multiple kernel learning on multimodal structural neuroimaging," *European Neuropsychopharmacology*, vol. 34, pp. 28–38, 2020.

[23] Y. Wei, Q. Chen, A. Curtin, L. Tu, X. Tang, Y. Tang, L. Xu, Z. Qian, J. Zhou, C. Zhu *et al.*, "Functional near-infrared spectroscopy (fnirs) as a tool to assist the diagnosis of major psychiatric disorders in a chinese population," *European archives of psychiatry and clinical neuroscience*, vol. 271, no. 4, pp. 745–757, 2021.

[24] G. Nilsonne and F. E. Harrell, "Eeg-based model and antidepressant response," *Nature Biotechnology*, vol. 39, no. 1, pp. 27–27, 2021.

[25] M. Shim, M. J. Jin, C.-H. Im, and S.-H. Lee, "Machine-learning-based classification between post-traumatic stress disorder and major depressive disorder using p300 features," *NeuroImage: Clinical*, vol. 24, p. 102001, 2019.

[26] H. Jiang, T. Popov, P. Jylänki, K. Bi, Z. Yao, Q. Lu, O. Jensen, and M. Van Gerven, "Predictability of depression severity based on posterior alpha oscillations," *Clinical Neurophysiology*, vol. 127, no. 4, pp. 2108–2114, 2016.

[27] X. Li, T. Cao, S. Sun, B. Hu, and M. Ratcliffe, "Classification study on eye movement data: Towards a new approach in depression detection," in *2016 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 2016, pp. 1227–1232.

[28] A. Y. Kim, E. H. Jang, S. Kim, K. W. Choi, H. J. Jeon, H. Y. Yu, and S. Byun, "Automatic detection of major depressive disorder using electrodermal activity," *Scientific reports*, vol. 8, no. 1, pp. 1–9, 2018.

[29] X. Ma, H. Yang, Q. Chen, D. Huang, and Y. Wang, "Depaudionet: An efficient deep model for audio based depression classification," in *Proceedings of the 6th international workshop on audio/visual emotion challenge*, 2016, pp. 35–42.

[30] X. Zhou, K. Jin, Y. Shang, and G. Guo, "Visually interpretable representation learning for depression recognition from facial images," *IEEE Transactions on Affective Computing*, vol. 11, no. 3, pp. 542–552, 2018.

[31] B. Miao, X. Liu, and T. Zhu, "Automatic mental health identification method based on natural gait pattern," *PsyCh Journal*, vol. 10, no. 3, pp. 453–464, 2021.

[32] M. Trotzek, S. Koitka, and C. M. Friedrich, "Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 3, pp. 588–601, 2018.

[33] S. Park, I. Kim, S. W. Lee, J. Yoo, B. Jeong, and M. Cha, "Manifestation of depression and loneliness on social networks: a case study of young adults on facebook," in *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, 2015, pp. 557–570.

[34] X. Yang, R. McEwen, L. R. Ong, and M. Zihayat, "A big data analytics framework for detecting user-level depression from social networks," *International Journal of Information Management*, vol. 54, p. 102141, 2020.

[35] L. Yang, H. Sahli, X. Xia, E. Pei, M. C. Oveneke, and D. Jiang, "Hybrid depression classification and estimation from audio video and text information," in *Proceedings of the 7th annual workshop on audio/visual emotion challenge*, 2017, pp. 45–51.

[36] A. Lenartowicz and R. Poldrack, "Brain imaging," 2017.

[37] E. J. Nestler, M. Barrot, R. J. DiLeone, A. J. Eisch, S. J. Gold, and L. M. Monteggia, "Neurobiology of depression," *Neuron*, vol. 34, no. 1, pp. 13–25, 2002.

[38] A. Widmann, E. Schröger, and B. Maess, "Digital filter design for electrophysiological data–a practical approach," *Journal of neuroscience methods*, vol. 250, pp. 34–46, 2015.

[39] S. Thomée, A. Härenstam, and M. Hagberg, "Mobile phone use and stress, sleep disturbances, and symptoms of depression among young adults-a prospective cohort study," *BMC public health*, vol. 11, no. 1, pp. 1–11, 2011.

[40] P. A. Lalousis, S. J. Wood, L. Schmaal, K. Chisholm, S. L. Griffiths, R. L. Reniers, A. Bertolino, S. Borgwardt, P. Brambilla, J. Kambeitz *et al.*, "Heterogeneity and classification of recent onset psychosis and depression: a multimodal machine learning approach," *Schizophrenia bulletin*, vol. 47, no. 4, pp. 1130–1140, 2021.

[41] Y. Meng, W. Speier, M. K. Ong, and C. W. Arnold, "Bidirectional representation learning from transformers using multimodal electronic health record data to predict depression," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 8, pp. 3121–3129, 2021.

[42] Y. J. Toenders, A. Kottaram, R. Dinga, C. G. Davey, T. Banaschewski, A. L. Bokde, E. B. Quinlan, S. Desrivières, H. Flor, A. Grigis *et al.*, "Predicting depression onset in young people based on clinical, cognitive, environmental, and neurobiological data," *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, vol. 7, no. 4, pp. 376–384, 2022.

[43] A. Beck, "An inventory for measuring depression," *Archives of General Psychiatry*, vol. 4, no. 6, pp. 561–, 1961.

[44] A. T. Beck, R. A. Steer, and G. Brown, "Beck depression inventory–ii," *Psychological assessment*, 1996.

[45] W. W. K. Zung, "A self-rating depression scale," *Arch. Gen. Psychiatry*, vol. 12, no. 1, pp. 63–70, 1965.

[46] J. B. Williams, K. Kroenke, and R. L. Spitzer, "The phq-9: validity of a brief depression severity measure," 2013.

[47] L. S. Radloff, "The ces-d scale : A self-report depression scale for research in the general population," *Appl Psychol Meas*, vol. 1, 1977.

[48] T. L. Brink, J. A. Yesavage, O. Lum, P. H. Heersema, M. Adey, and T. L. Rose, "Screening tests for geriatric depression," *Clinical gerontologist*, vol. 1, no. 1, pp. 37–43, 1982.

[49] M. Hamilton, "A rating scale for depression," *J. Neurol. Neurosurg*, vol. 23, no. 1, p. 56, 1960.

[50] S. A. Montgomery and M. Åsberg, "A new depression scale designed to be sensitive to change," *The British journal of psychiatry*, vol. 134, no. 4, pp. 382–389, 1979.

[51] W. H. Organization *et al.*, *The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines*. World Health Organization, 1992.

[52] D. L. Segal, "Diagnostic and statistical manual of mental disorders (dsm-iv-tr)," *The Corsini Encyclopedia of Psychology*, pp. 1–3, 2010.

[53] Y.-F. Chen, "Chinese classification of mental disorders (ccmd-3): towards integration in international classification," *Psychopathology*, vol. 35, no. 2-3, pp. 171–175, 2002.

[54] J. Gratch, R. Artstein, G. Lucas, G. Stratou, S. Scherer, A. Nazarian, R. Wood, J. Boberg, D. DeVault, S. Marsella *et al.*, "The distress analysis interview corpus of human and computer interviews," UNIVERSITY OF SOUTHERN CALIFORNIA LOS ANGELES, Tech. Rep., 2014.

[55] A. Yates, A. Cohan, and N. Goharian, "Depression and self-harm risk assessment in online forums," *arXiv preprint arXiv:1709.01848*, 2017.

[56]

[57] H. Cai, Y. Gao, S. Sun, N. Li, and B. Hu, "Modma dataset: a multimodel open dataset for mental- disorder analysis," 2020.

[58] M. Valstar, B. Schuller, K. Smith, T. Almaev, F. Eyben, J. Krajewski, R. Cowie, and M. Pantic, "Avec 2014: 3d dimensional affect and depression recognition challenge," in *Proceedings of the 4th international workshop on audio/visual emotion challenge*, 2014, pp. 3–10.

[59] M. Nasir, A. Jati, P. G. Shivakumar, S. Nallan Chakravarthula, and P. Georgiou, "Multimodal and multiresolution depression detection from speech and facial landmark features," in *Proceedings of the 6th international workshop on audio/visual emotion challenge*, 2016, pp. 43–50.

[60] F. Ringeval, B. Schuller, M. Valstar, N. Cummins, R. Cowie, L. Tavabi, M. Schmitt, S. Alisamir, S. Amiriparian, E.-M. Messner *et al.*, "Avec 2019 workshop and challenge: state-of-mind, detecting depression with ai, and cross-cultural affect recognition," in *Proceedings of the 9th International on Audio/visual Emotion Challenge and Workshop*, 2019, pp. 3–12.

[61] Y. Shen, H. Yang, and L. Lin, "Automatic depression detection: An emotional audio-textual corpus and a gru/bilstm-based model," 2022.

[62] A. S. Fischer, M. C. Camacho, T. C. Ho, S. Whitfield-Gabrieli, and I. H. Gotlib, "Neural markers of resilience in adolescent females at familial risk for major depressive disorder," *Jama Psychiatry*, 2018.

[63] L. Fu, X. Dan, J. Xiao, L. Yao, W. Ying, L. Xiao, H. Wang, G. Wang, and Z. Liu, "Reduced prefrontal activation during the tower of london and verbal fluency task in patients with bipolar depression: A multi-channel nirs study," *Frontiers in Psychiatry*, vol. 9, pp. 214–, 2018.

[64] L. Holper, M. J. Lan, P. J. Brown, E. M. Sublette, and J. J. Mann, "Brain cytochrome, xidase as a marker of mitochondrial function: A pilot study in major depression using nirs," *Depression and Anxiety*, no. 7, 2019.

[65] X. Liu, G. Sun, X. Zhang, B. Xu, C. Shen, L. Shi, X. Ma, X. Ren, K. Feng, and P. Liu, "Relationship between the prefrontal function and the severity of the emotional symptoms during a verbal fluency task in patients with major depressive disorder: A multi-channel nirs study," *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, vol. 54, pp. 114–121, 2014.

[66] M. Kawano, T. Kanazawa, H. Kikuyama, A. Tsutsumi, and H. Yoneda, "Correlation between frontal lobe oxy-hemoglobin and severity of depression assessed using near-infrared spectroscopy," *Journal of Affective Disorders*, vol. 205, pp. 154–158, 2016.

[67] A. Manelis, T. J. Huppert, E. Rodgers, H. A. Swartz, and M. L. Phillips, "The role of the right prefrontal cortex in recognition of facial emotional expressions in depressed individuals: fnirs study," *Journal of Affective Disorders*, vol. 258, pp. 151–158, 2019.

[68] A. Shatte, D. M. Hutchinson, and S. J. Teague, "Machine learning in mental health: A scoping review of methods and applications," *Psychological Medicine*, vol. 49, no. 9, pp. 1–23, 2019.

[69] Y. Zhu and M. Mehta, "Machine learning approach on frontal lobe activity to assess depression in adults: Implications for rehabilitation outcomes," 2017, pp. 1–2.

[70] C. Lei, S. Hong, and G. Ruiqi, "Automatic depression discrimination on fnirs by using general linear model and svm," *IEEE*, 2017.

[71] S. Hong, W. Du, and Q. Zhao, "Automatic depression discrimination on fnirs by using fastica/wpd and svm," 2015.

[72] Y. Zhu, J. K. Jayagopal, R. K. Mehta, M. Erraguntla, and S. H. Chang, "Classifying major depressive disorder using fnirs during motor rehabilitation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. PP, no. 99, pp. 1–1, 2020.

[73] C. Eastmond, A. Subedi, S. De, and X. Intes, "Deep learning in fnirs: A review," 2022.

[74] R. Wang, Y. Hao, Q. Yu, M. Chen, I. Humar, and G. Fortino, "Depression analysis and recognition based on functional near-infrared spectroscopy." *IEEE journal of biomedical and health informatics*, vol. 25, no. 12, pp. 4289–4299, 2021.

[75] T. Ma, H. Lyu, J. Liu, Y. Xia, and S. He, "Distinguishing bipolar depression from major depressive disorder using fnirs and deep neural network," *Progress In Electromagnetics Research*, vol. 169, pp. 73–86, 2020.

[76] J. Chao, S. Zheng, H. Wu, D. Wang, X. Zhang, H. Peng, and B. Hu, "fnirs evidence for distinguishing patients with major depression and healthy controls," *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 29, pp. 2211–2221, 2021.

[77] K. Kipli, A. Z. Kouzani, and M. Joordens, "Evaluation of feature selection algorithms for detection of depression from brain smri scans," in *Icme International Conference on Complex Medical Engineering*, 2013.

[78] P. Chagué, B. Marro, S. Fadili, M. Houot, and O. Colliot, "Radiological classification of dementia from anatomical mri assisted by machine learning-derived maps," *Journal of Neuroradiology*, 2020.

[79] M. Mousavian, J. Chen, and S. Greening, "Depression detection using feature extraction and deep learning from smri images," in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, 2019.

[80] J. X. Zheng, Y. C. Chen, H. Chen, J. Liang, B. Fan, F. Yuan, W. W. Tang, X. Yin, and J. P. Gu, "Disrupted spontaneous neural activity related to cognitive impairment in postpartum women," *Frontiers in Psychology*, vol. 9, pp. 624–, 2018.

[81] M. Wei, J. Qin, R. Yan, H. Li, Z. Yao, and Q. Lu, "Identifying major depressive disorder using hurst exponent of resting-state brain networks," *Psychiatry Research: Neuroimaging*, vol. 214, no. 3, pp. 306–312, 2013.

[82] D. Hirshfeld-Becker, J. D. Gabrieli, B. Shapero, J. Biederman, S. Whitfield-Gabrieli, and X. J. Chai, "Intrinsic functional brain connectivity predicts onset of major depression disorder in adolescence: A pilot study," *Brain Connectivity*, 2019.

[83] H. Guo, Y. Li, G. K. Mensah, Y. Xu, and D. Chen, "Resting-state functional network scale effects and statistical significance-based feature selection in machine learning classification," *Computational and Mathematical Methods in Medicine*, vol. 2019, no. 2, pp. 1–18, 2019.

[84] L. L. Zeng, H. Shen, L. L. And, and D. Hu, "Unsupervised classification of major depression using functional connectivity mri," *Human Brain Mapping*, 2014.

[85] R. Bhaumik, L. M. Jenkins, J. R. Gowins, R. H. Jacobs, A. Barba, D. K. Bhaumik, and S. A. Langenecker, "Multivariate pattern analysis strategies in detection of remitted major depressive disorder using resting state functional connectivity," *Neuroimage Clinical*, p. 390, 2016.

[86] B. Yan, X. Xu, M. Liu, K. Zheng, and B. Li, "Quantitative identification of major depression based on resting-state dynamic functional connectivity: A machine learning approach," *Frontiers in Neuroscience*, vol. 14, 2020.

[87] A. Yamashita, Y. Sakai, T. Yamada, N. Yahata, A. Kunimatsu, N. Okada, T. Itahashi, R. Hashimoto, H. Mizuta, and N. Ichikawa, "Generalizable brain network markers of major depressive disorder across multiple imaging sites," *PLOS Biology*, vol. 18, 2020.

[88] J. Zhao, J. Huang, D. Zhi, W. Yan, and J. Sui, "Functional network connectivity (fnc)-based generative adversarial network (gan) and its applications in classification of mental disorders," *Journal of Neuroscience Methods*, vol. 341, p. 108756, 2020.

[89] E. Jun, K. Na, W. Kang, J. Lee, H. Suk, and B. Ham, "Identifying resting-state effective connectivity abnormalities in drug-nave major depressive disorder diagnosis via graph convolutional networks," *Human Brain Mapping*, vol. 41, no. 17, 2020.

[90] O. Leticia, C. D. Ladouceur, M. L. Phillips, B. Michael, M. M. Janaina, and B. J. Harrison, "What does brain response to neutral faces tell us about major depression? evidence from machine learning and fmri," *PLoS ONE*, vol. 8, no. 4, pp. e60 121–, 2013.

[91] S. Yu, Y. Junichiro, T. Shigeru, T. Masahiro, Y. Shinpei, O. Yasumasa, Y. Shigeto, D. Kenji, and B. Michael, "Toward probabilistic diagnosis and understanding of depression based on functional mri data analysis with logistic group lasso," *PLOS ONE*, vol. 10, no. 5, pp. e0 123 524–, 2015.

[92] R. Gui, T. Chen, and H. Nie, "The impact of emotional music on active roi in patients with depression based on deep learning: A task-state fmri study," *Computational Intelligence and Neuroscience*, vol. 2019, no. 6, pp. 1–14, 2019.

[93] K. Marek, D. Jennings, S. Lasch, A. Siderowf, C. Tanner, T. Simuni, C. Coffey, K. Kieburtz, E. Flagg, S. Chowdhury *et al.*, "The parkinson progression marker initiative (ppmi)," *Progress in neurobiology*, vol. 95, no. 4, pp. 629–635, 2011.

[94] C. Fu, D. Shi, Y. Gao, and J. Xu, "Functional assessment of prefrontal lobes in patients with major depression disorder using a dual-mode technique of 3d-arterial spin labeling and 18f-fluorodeoxyglucose positron emission tomography/computed tomography," *Experimental & Therapeutic Medicine*, 2017.

[95] S. E. Holmes, R. Hinz, S. Conen, C. J. Gregory, and P. S. Talbot, "Elevated translocator protein in anterior cingulate in major depression and a role for inflammation in suicidal thinking: a pet study," *Biological Psychiatry*, vol. 83, no. 1, 2017.

[96] H. Li, A. P. Sagar, and S. Kéri, "Microglial markers in the frontal cortex are related to cognitive dysfunctions in major depressive disorder," *Journal of Affective Disorders*, vol. 241, pp. 305–310, 2018.

[97] L. Su, Y. Cai, Y. Xu, A. Dutt, S. Shi, and E. Bramon, "Cerebral metabolism in major depressive disorder: a voxel-based meta-analysis of positron emission tomography studies," *Bmc Psychiatry*, vol. 14, no. 1, pp. 1–7, 2014.

[98] J. D. Coplan, D. Lu, A. E. Sehamy, C. Tang, A. P. Jackowski, C. G. Abdallah, C. B. Nemeroff, M. J. Owens, S. J. Mathew, and J. M. Gorman, "Early life stress associated with increased striatal n-acetyl-aspartate: Cerebrospinal fluid corticotropin-releasing factor concentrations, hippocampal volume, body mass and behavioral correlates." *Chronic stress (Thousand Oaks, Calif.)*, vol. 2, 2018.

[99] S. Rooney, E. Roman, A. O'Callaghan, D. Roddy, and V. O'Keane, "1h (proton) mrs in depression - a review of literature and future directions," 2018.

[100] L. Zhou, L. Min, S. H. Zheng, and L. Jie, "[analysis of h-mrs in patients with depression after basal ganglia infarction]." *Journal of Zhejiang University*, vol. 42, no. 2, pp. 205–211, 2013.

[101] M. Mousavian, J. Chen, and S. Greening, "Depression detection using atlas from fmri images," in *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2020, pp. 1348–1353.

[102] N.-F. Jie, E. A. Osuch, M.-H. Zhu, M. Wammes, X.-Y. Ma, T.-Z. Jiang, J. Sui, and V. D. Calhoun, "Discriminating bipolar disorder

from major depression using whole-brain functional connectivity: a feature selection analysis with svm-foba algorithm," *Journal of signal processing systems for signal, image, and video technology*, vol. 90, no. 2, pp. 259–271, 1 2018.

[103] L. Gu, L. Huang, F. Yin, and Y. Cheng, "Classification of depressive disorder based on rs-fmri using multivariate pattern analysis with multiple features," in *2017 4th IAPR Asian Conference on Pattern Recognition (ACPR)*, 2017, pp. 61–66.

[104] G. E. Bruder, J. W. Stewart, and P. J. McGrath, "Right brain, left brain in depressive disorders: clinical and theoretical implications of behavioral, electrophysiological and neuroimaging findings," *Neuroscience & Biobehavioral Reviews*, vol. 78, pp. 178–191, 2017.

[105] S.-C. Liao, C.-T. Wu, H.-C. Huang, W.-T. Cheng, and Y.-H. Liu, "Major depression detection from eeg signals using kernel eigen-filter-bank common spatial patterns," *Sensors*, vol. 17, no. 6, p. 1385, 2017.

[106] G. M. Bairy, O. S. Lih, Y. Hagiwara, S. D. Puthankattil, O. Faust, U. Niranjan, and U. R. Acharya, "Automated diagnosis of depression electroencephalograph signals using linear prediction coding and higher order spectra features," *Journal of Medical Imaging and Health Informatics*, vol. 7, no. 8, pp. 1857–1862, 2017.

[107] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, H. Adeli, and D. P. Subha, "Automated eeg-based screening of depression using deep convolutional neural network," *Computer methods and programs in biomedicine*, vol. 161, pp. 103–113, 2018.

[108] X. Ding, X. Yue, R. Zheng, C. Bi, D. Li, and G. Yao, "Classifying major depression patients and healthy controls using eeg, eye tracking and galvanic skin response data," *Journal of affective Disorders*, vol. 251, pp. 156–161, 2019.

[109] J. Walleczek, *Self-organized biological dynamics and nonlinear control: toward understanding complexity, chaos and emergent function in living systems*. Cambridge University Press, 2006.

[110] A. Scott, *Encyclopedia of nonlinear science*. Routledge, 2006.

[111] W. Klonowski, "Application of nonlinear dynamics in biosignal analysis," in *Topical Problems of Nonlinear Wave Physics*, vol. 5975. SPIE, 2006, pp. 335–344.

[112] S. A. Akar, S. Kara, S. Agambayev, and V. Bilgiç, "Nonlinear analysis of eegs of patients with major depression during different emotional states," *Computers in biology and medicine*, vol. 67, pp. 49–60, 2015.

[113] A. Ortiz, K. Bradler, M. Mowete, S. MacLean, J. Garnham, C. Slaney, B. H. Mulsant, and M. Alda, "The futility of long-term predictions in bipolar disorder: mood fluctuations are the result of deterministic chaotic processes," *International journal of bipolar disorders*, vol. 9, no. 1, pp. 1–7, 2021.

[114] A. Kustubayeva, M. Zholdassova, G. Matthews, and E. Nelson, "Lyapunov exponent of theta rhythm as a marker of depression during attentional network test," *Biological Psychiatry*, vol. 89, no. 9, p. S164, 2021.

[115] G. Chopra and P. WHIG, "Using machine learning algorithms classified depressed patients and normal people," *International Journal of Machine Learning for Sustainable Development*, vol. 4, no. 1, pp. 31–40, 2022.

[116] L. Zhao, L. Yang, Z. Su, and C. Liu, "Cardiorespiratory coupling analysis based on entropy and cross-entropy in distinguishing different depression stages," *Frontiers in physiology*, vol. 10, p. 359, 2019.

[117] M. Cukic, D. Pokrajac, M. Stokic, V. Radivojevic, M. Ljubisavljevic *et al.*, "Eeg machine learning with higuchi fractal dimension and sample entropy as features for successful detection of depression," *arXiv preprint arXiv:1803.05985*, 2018.

[118] L. Zhao, L. Yang, B. Li, Z. Su, and C. Liu, "Frontal alpha eeg asymmetry variation of depression patients assessed by entropy measures and lemple–ziv complexity," *Journal of Medical and Biological Engineering*, vol. 41, no. 2, pp. 146–154, 2021.

[119] U. R. Acharya, V. K. Sudarshan, H. Adeli, J. Santhosh, J. E. Koh, S. D. Puthankatti, and A. Adeli, "A novel depression diagnosis index using nonlinear features in eeg signals," *European neurology*, vol. 74, no. 1-2, pp. 79–83, 2015.

[120] B. Hosseinifard, M. H. Moradi, and R. Rostami, "Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from eeg signal," *Computer methods and programs in biomedicine*, vol. 109, no. 3, pp. 339–345, 2013.

[121] M. Ahmadlou, H. Adeli, and A. Adeli, "Fractality analysis of frontal brain in major depressive disorder," *International Journal of Psychophysiology*, vol. 85, no. 2, pp. 206–211, 2012.

[122] Z. Wan, J. Huang, H. Zhang, H. Zhou, J. Yang, and N. Zhong, "Hybrideegnet: A convolutional neural network for eeg feature learning and depression discrimination," *IEEE Access*, vol. 8, pp. 30 332–30 342, 2020.

[123] A. Saeedi, M. Saeedi, A. Maghsoudi, and A. Shalbaf, "Major depressive disorder diagnosis based on effective connectivity in eeg signals: A convolutional neural network and long short-term memory approach," *Cognitive Neurodynamics*, vol. 15, no. 2, pp. 239–252, 2021.

[124] A. Qayyum, I. Razzak, and W. Mumtaz, "Hybrid deep shallow network for assessment of depression using electroencephalogram signals," in *International conference on neural information processing*. Springer, 2020, pp. 245–257.

[125] P. P. Thoduparambil, A. Dominic, and S. M. Varghese, "Eeg-based deep learning model for the automatic detection of clinical depression," *Physical and Engineering Sciences in Medicine*, vol. 43, no. 4, pp. 1349–1360, 2020.

[126] X. Li, R. La, Y. Wang, J. Niu, S. Zeng, S. Sun, and J. Zhu, "Eeg-based mild depression recognition using convolutional neural network," *Medical & biological engineering & computing*, vol. 57, no. 6, pp. 1341–1352, 2019.

[127] S. Alhagry, A. A. Fahmy, and R. A. El-Khoribi, "Emotion recognition based on eeg using lstm recurrent neural network," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 10, 2017.

[128] P. Wang, A. Jiang, X. Liu, J. Shang, and L. Zhang, "Lstm-based eeg classification in motor imagery tasks," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 26, no. 11, pp. 2086–2095, 2018.

[129] S. D. Kumar and D. Subha, "Prediction of depression from eeg signal using long short term memory (lstm)," in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*. IEEE, 2019, pp. 1248–1253.

[130] G. Sharma, A. Parashar, and A. M. Joshi, "Dephnn: a novel hybrid neural network for electroencephalogram (eeg)-based screening of depression," *Biomedical signal processing and control*, vol. 66, p. 102393, 2021.

[131] M. Bachmann, L. Päeske, K. Kalev, K. Aarma, A. Lehtmets, P. Ööpik, J. Lass, and H. Hinrikus, "Methods for classifying depression in single channel eeg using linear and nonlinear signal analysis," *Computer methods and programs in biomedicine*, vol. 155, pp. 11–17, 2018.

[132] O. Faust, P. C. A. Ang, S. D. Puthankattil, and P. K. Joseph, "Depression diagnosis support system based on eeg signal entropies," *Journal of mechanics in medicine and biology*, vol. 14, no. 03, p. 1450035, 2014.

[133] M. Kang, H. Kwon, J.-H. Park, S. Kang, and Y. Lee, "Deep-asymmetry: Asymmetry matrix image for deep learning method in pre-screening depression," *Sensors*, vol. 20, no. 22, p. 6526, 2020.

[134] B. Ay, O. Yildirim, M. Talo, U. B. Baloglu, G. Aydin, S. D. Puthankattil, and U. R. Acharya, "Automated depression detection using deep representation and sequence learning with eeg signals," *Journal of medical systems*, vol. 43, no. 7, pp. 1–12, 2019.

[135] V. K. Jandackova, A. Britton, M. Malik, and A. Steptoe, "Heart rate variability and depressive symptoms: a cross-lagged analysis over a 10-year period in the whitehall ii study," *Psychological medicine*, vol. 46, no. 10, pp. 2121–2131, 2016.

[136] H. R. Variability, "Standards of measurement, physiological interpretation, and clinical use. task force of the european society of cardiology and the north american society of pacing and electrophysiology," *Eur Heart J*, vol. 17, no. 3, pp. 354–381, 1996.

[137] A. Sgoifo, L. Carnevali, M. d. l. A. Pico Alfonso, and M. Amore, "Autonomic dysfunction and heart rate variability in depression," *Stress*, vol. 18, no. 3, pp. 343–352, 2015.

[138] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in public health*, p. 258, 2017.

[139] F. Schneider, J. Martin, A. Hapfelmeier, D. Jordan, G. Schneider, and C. M. Schulz, "The validity of linear and non-linear heart rate metrics as workload indicators of emergency physicians," *PloS one*, vol. 12, no. 11, p. e0188635, 2017.

[140] J. Colombo, R. Arora, N. L. DePace, and A. I. Vinik, "Clinical autonomic dysfunction," *Switzerland: Springer International Publishing*, 2015.

[141] S. Byun, A. Y. Kim, E. H. Jang, S. Kim, K. W. Choi, H. Y. Yu, and H. J. Jeon, "Detection of major depressive disorder from linear and nonlinear heart rate variability features during mental task protocol," *Computers in biology and medicine*, vol. 112, p. 103381, 2019.

[142] M. Kidwell and B. A. Ellenbroek, "Heart and soul: heart rate variability and major depression," *Behavioural pharmacology*, vol. 29, no. 2, pp. 152–164, 2018.

[143] G. E. Billman, "The lf/hf ratio does not accurately measure cardiac sympatho-vagal balance," p. 26, 2013.

[144] A. H. Kemp, D. S. Quintana, K. L. Felmingham, S. Matthews, and H. F. Jelinek, "Depression, comorbid anxiety disorders, and heart rate variability in physically healthy, unmedicated patients: implications for cardiovascular risk," *PloS one*, vol. 7, no. 2, p. e30777, 2012.

[145] T. Shinba, "Altered autonomic activity and reactivity in depression revealed by heart-rate variability measurement during rest and task conditions," *Psychiatry and Clinical Neurosciences*, vol. 68, no. 3, pp. 225–233, 2014.

[146] A. L. Goldberger, L. A. Amaral, J. M. Hausdorff, P. C. Ivanov, C.-K. Peng, and H. E. Stanley, "Fractal dynamics in physiology: alterations with disease and aging," *Proceedings of the national academy of sciences*, vol. 99, no. suppl_1, pp. 2466–2472, 2002.

[147] S. Schulz, M. Koschke, K.-J. Bär, and A. Voss, "The altered complexity of cardiovascular regulation in depressed patients," *Physiological measurement*, vol. 31, no. 3, p. 303, 2010.

[148] A. Greco, S. M. Benvenuti, C. Gentili, D. Palomba, E. P. Scilingo, and G. Valenza, "Assessment of linear and nonlinear/complex heartbeat dynamics in subclinical depression (dysphoria)," *Physiological measurement*, vol. 39, no. 3, p. 034004, 2018.

[149] D. E. Vigo, L. N. Siri, M. S. L. De Guevara, J. A. Martínez-Martínez, R. D. Fahrer, D. P. Cardinali, O. Masoli, and S. M. Guinjoan, "Relation of depression to heart rate nonlinear dynamics in patients 60 years of age with recent unstable angina pectoris or acute myocardial infarction," *The American journal of cardiology*, vol. 93, no. 6, pp. 756–760, 2004.

[150] D. Kuang, R. Yang, X. Chen, G. Lao, F. Wu, X. Huang, R. Lv, L. Zhang, C. Song, and S. Ou, "Depression recognition according to heart rate variability using bayesian networks," *Journal of psychiatric research*, vol. 95, pp. 282–287, 2017.

[151] S. T. Noor, S. T. Asad, M. M. Khan, G. S. Gaba, J. F. Al-Amri, and M. Masud, "Predicting the risk of depression based on ecg using rnn," *Computational Intelligence and Neuroscience*, vol. 2021, 2021.

[152] X. Zang, B. Li, L. Zhao, D. Yan, and L. Yang, "End-to-end depression recognition based on a one-dimensional convolution neural network model using two-lead ecg signal," *Journal of Medical and Biological Engineering*, vol. 42, no. 2, pp. 225–233, 2022.

[153] Z.-X. Zhang, X.-W. Tian, and J. S. Lim, "New algorithm for the depression diagnosis using hrv: A neuro-fuzzy approach," in *International Symposium on Bioelectronics and Bioinformations 2011*. IEEE, 2011, pp. 283–286.

[154] T. Roh, S. Hong, and H.-J. Yoo, "Wearable depression monitoring system with heart-rate variability," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2014, pp. 562–565.

[155] T. Matsui, K. Kakisaka, and T. Shinba, "Impaired parasympathetic augmentation under relaxation in patients with depression as assessed by a novel non-contact microwave radar system," *Journal of Medical Engineering & Technology*, vol. 40, no. 1, pp. 15–19, 2016.

[156] G. Sun, T. Shinba, T. Kirimoto, and T. Matsui, "An objective screening method for major depressive disorder using logistic regression analysis of heart rate variability data obtained in a mental task paradigm," *Frontiers in Psychiatry*, vol. 7, p. 180, 2016.

[157] S. Mohanraj, S. Balasubramaniyam, V. Kannan, and D. Jeeva, "A deep convolution neural network framework for detecting depression," in *2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS)*. IEEE, 2022, pp. 1061–1068.

[158] J. A. Sweeney, M. H. Strojwas, J. J. Mann, and M. E. Thase, "Prefrontal and cerebellar abnormalities in major depression: evidence from oculomotor studies," *Biological psychiatry*, vol. 43, no. 8, pp. 584–594, 1998.

[159] A. Duque and C. Vázquez, "Double attention bias for positive and negative emotional faces in clinical depression: Evidence from an eye-tracking study," *Journal of behavior therapy and experimental psychiatry*, vol. 46, pp. 107–114, 2015.

[160] M. Li, L. Cao, Q. Zhai, P. Li, S. Liu, R. Li, L. Feng, G. Wang, B. Hu, and S. Lu, "Method of depression classification based on behavioral and physiological signals of eye movement," *Complexity*, vol. 2020, 2020.

[161] Z. Pan, H. Ma, L. Zhang, and Y. Wang, "Depression detection based on reaction time and eye movement," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 2184–2188.

[162] S. Lu, S. Liu, M. Li, X. Shi, and R. Li, "Depression classification model based on emotionally related eye-movement data and kernel extreme learning machine," *Journal of Medical Imaging and Health Informatics*, vol. 10, no. 11, pp. 2668–2674, 2020.

[163] T. Zhu, "Identifying psychological symptoms based on facial movements," *International Journal of Psychophysiology*, vol. 168.

[164] J. M. Girard, J. F. Cohn, M. H. Mahoor, S. Mavadati, and D. P. Rosenwald, "Social risk and depression: Evidence from manual and automatic facial expression analysis," in *International Conference on Automatic Face and Gesture Recognition*, 2013.

[165] J. Rottenberg, J. J. Gross, and I. H. Gotlib, "Emotion context insensitivity in major depressive disorder." *Journal of Abnormal Psychology*, vol. 114, no. 4, pp. 627–639, 2005.

[166] L. M. Bylsma, B. H. Morris, and J. Rottenberg, "A meta-analysis of emotional reactivity in major depressive disorder," *Clinical Psychology Review*, vol. 28, no. 4, pp. 676–691, 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0272735807001626

[167] A. Pampouchidou, O. Simantiraki, C. M. Vazakopoulou, C. Chatzaki, and M. Tsiknakis, "Facial geometry and speech analysis for depression detection," *IEEE*, 2017.

[168] S. Bhatia, M. Hayat, M. Breakspear, G. Parker, and R. Goecke, "A video-based facial behaviour analysis approach to melancholia," in *IEEE International Conference on Automatic Face & Gesture Recognition*, 2017.

[169] J. Li, Z. Liu, Z. Ding, and G. Wang, "A novel study for mdd detection through task-elicited facial cues," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2018.

[170] A. Mulay, A. Dhekne, R. Wani, S. Kadam, and P. Deshpande, "Automatic depression level detection through visual input," in *2020 Fourth World Conference on Smart Trends in Systems Security and Sustainablity (WorldS4)*, 2020.

[171] Z. Jiang, S. Harati, A. Crowell, H. Mayberg, S. Nemati, and G. Clifford, "Classifying major depressive disorder and response to deep brain stimulation over time by analyzing facial expressions." *IEEE transactions on biomedical engineering*, vol. 68, no. 2, pp. 664–672, 2021.

[172] W. Melo, E. Granger, and A. Hadid, "Combining global and local convolutional 3d networks for detecting depression from facial expressions," in *14th IEEE International Conference on Automatic Face and Gesture Recognition*, 2019.

[173] L. Yang, "Multi-modal depression detection and estimation," in *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, 2019.

[174] W. Melo, E. Granger, and M. B. Lopez, "Encoding temporal information for automatic depression recognition from facial analysis," in *Int'l Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2020)*, 2020.

[175] W. Melo, E. Granger, and A. Hadid, "A deep multiscale spatiotemporal network for assessing depression from facial dynamics," *IEEE Transactions on Affective Computing*, vol. PP, no. 99, 2020.

[176] X. Zhou, K. Jin, Y. Shang, and G. Guo, "Visually interpretable representation learning for depression recognition from facial images," *IEEE Transactions on Affective Computing*, vol. 11, no. 3, pp. 542–552, 2020.

[177] X. Li, W. Guo, and H. Yang, "Depression severity prediction from facial expression based on the drr_depressionnet network," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2020.

[178] W. Melo, E. Granger, and A. Hadid, "Depression detection based on deep distribution learning," in *IEEE International Conference on Image Processing (ICIP 2019)*, 2019.

[179] M. Al Jazaery and G. Guo, "Video-based depression level analysis by encoding deep spatiotemporal features," *IEEE Transactions on Affective Computing*, vol. 12, no. 1, pp. 262–268, 2021.

[180] M. H. Su, C. H. Wu, K. Y. Huang, Q. B. Hong, and H. M. Wang, "Exploring microscopic fluctuation of facial expression for mood disorder classification," in *2017 International Conference on Orange Technologies (ICOT)*, 2017.

[181] Y. Wang, J. Ma, B. Hao, P. Hu, and S. Li, "Automatic depression detection via facial expressions using multiple instance learning," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 2020.

[182] M. Valstar, B. Schuller, K. Smith, F. Eyben, B. Jiang, S. Bilakhia, S. Schnieder, R. Cowie, and M. Pantic, "Avec 2013: The continuous audio/visual emotion and depression recognition challenge," in *Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge*, ser. AVEC '13. New York, NY, USA: Association for Computing Machinery, 2013, pp. 3–10. [Online]. Available: https://doi.org/10.1145/2512530.2512533

[183] M. Valstar, B. Schuller, K. Smith, T. Almaev, F. Eyben, J. Krajewski, R. Cowie, and M. Pantic, "Avec 2014: 3d dimensional affect and depression recognition challenge," in *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge*, ser. AVEC '14. New York, NY, USA: Association for Computing

Machinery, 2014, pp. 3–10. [Online]. Available: https://doi.org/10.1145/2661806.2661807

[184] H. Kaya and A. A. Salah, "Eyes whisper depression: A cca based multimodal approach," in *Acm International Conference on Multimedia*, 2014.

[185] A. Jan, H. Meng, Y. Gaus, F. Zhang, and S. Turabzadeh, "Automatic depression scale prediction using facial expression dynamics and regression," in *International Workshop on Audio/visual Emotion Challenge*, 2014.

[186] L. Wen, X. Li, G. Guo, and Y. Zhu, "Automated depression diagnosis based on facial dynamic analysis and sparse coding," *IEEE Transactions on Information Forensics & Security*, vol. 10, no. 7, pp. 1432–1441, 2015.

[187] M. Valstar, J. Gratch, B. Schuller, F. Ringeval, D. Lalanne, M. Torres Torres, S. Scherer, G. Stratou, R. Cowie, and M. Pantic, "Avec 2016: Depression, mood, and emotion recognition workshop and challenge," in *Proceedings of the 6th international workshop on audio/visual emotion challenge*, 2016, pp. 3–10.

[188] J. R. Williamson, E. Godoy, M. Cha, A. Schwarzentruber, and T. F. Quatieri, "Detecting depression using vocal, facial and semantic communication cues," in *the 6th International Workshop*, 2016.

[189] F. Ringeval, B. Schuller, M. Valstar, J. Gratch, R. Cowie, S. Scherer, S. Mozgai, N. Cummins, M. Schmitt, and M. Pantic, "Avec 2017: Real-life depression, and affect recognition workshop and challenge," in *Proceedings of the 7th annual workshop on audio/visual emotion challenge*, 2017, pp. 3–9.

[190] Z. Yu, Y. Shang, Z. Shao, and G. Guo, "Automated depression diagnosis based on deep networks to encode facial appearance and dynamics," *IEEE Transactions on Affective Computing*, vol. PP, no. 99, pp. 1–1, 2017.

[191] A. Ray, S. Kumar, R. Reddy, P. Mukherjee, and R. Garg, "Multi-level attention network using text, audio and video for depression prediction," in *Proceedings of the 9th international on audio/visual emotion challenge and workshop*, 2019, pp. 81–88.

[192] X. Zhou, P. Huang, H. Liu, and S. Niu, "Learning content-adaptive feature pooling for facial depression recognition in videos," *Electronics Letters*, vol. 55, no. 11, 2019.

[193] A. Pampouchidou, O. Simantiraki, C.-M. Vazakopoulou, C. Chatzaki, M. Pediaditis, A. Maridaki, K. Marias, P. Simos, F. Yang, F. Meriaudeau *et al.*, "Facial geometry and speech analysis for depression detection," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2017, pp. 1433–1436.

[194] D. J. France, R. G. Shiavi, S. Silverman, M. Silverman, and M. Wilkes, "Acoustical properties of speech as indicators of depression and suicidal risk," *IEEE transactions on Biomedical Engineering*, vol. 47, no. 7, pp. 829–837, 2000.

[195] E. Moore, M. Clements, J. Peifer, and L. Weisser, "Analysis of prosodic variation in speech for clinical depression," in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No. 03CH37439)*, vol. 3. IEEE, 2003, pp. 2925–2928.

[196] E. Moore II, M. A. Clements, J. W. Peifer, and L. Weisser, "Critical analysis of the impact of glottal features in the classification of clinical depression in speech," *IEEE transactions on biomedical engineering*, vol. 55, no. 1, pp. 96–107, 2007.

[197] R. Shankayi, M. Vali, M. Salimi, and M. Malekshahi, "Identifying depressed from healthy cases using speech processing," in *2012 19th Iranian Conference of Biomedical Engineering (ICBME)*. IEEE, 2012, pp. 242–245.

[198] Y. Gong and C. Poellabauer, "Topic modeling based multi-modal depression detection," in *Proceedings of the 7th annual workshop on Audio/Visual emotion challenge*, 2017, pp. 69–76.

[199] N. Cummins, J. Epps, and E. Ambikairajah, "Spectro-temporal analysis of speech affected by depression and psychomotor retardation," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 7542–7546.

[200] B. S. Helfer, T. F. Quatieri, J. R. Williamson, D. D. Mehta, R. Horwitz, and B. Yu, "Classification of depression state based on articulatory precision." in *Interspeech*, 2013, pp. 2172–2176.

[201] J. R. Williamson, T. F. Quatieri, B. S. Helfer, R. Horwitz, B. Yu, and D. D. Mehta, "Vocal biomarkers of depression based on motor incoordination," in *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge*, 2013, pp. 41–48.

[202] S. Alghowinem, R. Goecke, M. Wagner, J. Epps, T. Gedeon, M. Breakspear, and G. Parker, "A comparative study of different classifiers for detecting depression from spontaneous speech," in *2013 IEEE international conference on acoustics, speech and signal processing*. IEEE, 2013, pp. 8022–8026.

[203] H. Jiang, B. Hu, Z. Liu, L. Yan, T. Wang, F. Liu, H. Kang, and X. Li, "Investigation of different speech types and emotions for detecting depression using different classifiers," *Speech Communication*, vol. 90, pp. 39–46, 2017.

[204] K. E. B. Ooi, M. Lech, and N. B. Allen, "Prediction of major depression in adolescents using an optimized multi-channel weighted speech classification system," *Biomedical Signal Processing and Control*, vol. 14, pp. 228–239, 2014.

[205] Z. Du, W. Li, D. Huang, and Y. Wang, "Bipolar disorder recognition via multi-scale discriminative audio temporal representation," in *Proceedings of the 2018 on Audio/Visual Emotion Challenge and Workshop*, 2018, pp. 23–30.

[206] N. Srimadhur and S. Lalitha, "An end-to-end model for detection and assessment of depression levels using speech," *Procedia Computer Science*, vol. 171, pp. 12–21, 2020.

[207] C. Fan, Z. Lv, S. Pei, and M. Niu, "Csenet: Complex squeeze-and-excitation network for speech depression level prediction," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 546–550.

[208] A. Haque, M. Guo, A. S. Miner, and L. Fei-Fei, "Measuring depression symptom severity from spoken language and 3d facial expressions," *arXiv preprint arXiv:1811.08592*, 2018.

[209] L.-S. A. Low, N. C. Maddage, M. Lech, L. B. Sheeber, and N. B. Allen, "Detection of clinical depression in adolescents' speech during family interactions," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 574–586, 2010.

[210] M. Valstar, B. Schuller, K. Smith, F. Eyben, B. Jiang, S. Bilakhia, S. Schnieder, R. Cowie, and M. Pantic, "Avec 2013: the continuous audio/visual emotion and depression recognition challenge," in *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge*, 2013, pp. 3–10.

[211] N. Cummins, J. Epps, M. Breakspear, and R. Goecke, "An investigation of depressed speech detection: Features and normalization," in *Twelfth Annual Conference of the International Speech Communication Association*, 2011.

[212] H. Meng, D. Huang, H. Wang, H. Yang, M. Ai-Shuraifi, and Y. Wang, "Depression recognition based on dynamic facial and vocal expression features using partial least square regression," in *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge*, 2013, pp. 21–30.

[213] J. R. Williamson, T. F. Quatieri, B. S. Helfer, G. Ciccarelli, and D. D. Mehta, "Vocal and facial biomarkers of depression based on motor incoordination and timing," in *Proceedings of the 4th international workshop on audio/visual emotion challenge*, 2014, pp. 65–72.

[214] S. Jayawardena, J. Epps, and E. Ambikairajah, "Ordinal logistic regression with partial proportional odds for depression prediction," *IEEE Transactions on Affective Computing*, 2020.

[215] Y. Kang, X. Jiang, Y. Yin, Y. Shang, and X. Zhou, "Deep transformation learning for depression diagnosis from facial images," in *Chinese conference on biometric recognition*. Springer, 2017, pp. 13–22.

[216] T. Al Hanai, M. M. Ghassemi, and J. R. Glass, "Detecting depression with audio/text sequence modeling of interviews." in *Interspeech*, 2018, pp. 1716–1720.

[217] S. Dham, A. Sharma, and A. Dhall, "Depression scale recognition from audio, visual and text analysis," *arXiv preprint arXiv:1709.05865*, 2017.

[218] A. Salekin, J. W. Eberle, J. J. Glenn, B. A. Teachman, and J. A. Stankovic, "A weakly supervised learning framework for detecting social anxiety and depression," *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, vol. 2, no. 2, pp. 1–26, 2018.

[219] P. Zhang, M. Wu, H. Dinkel, and K. Yu, "Depa: Self-supervised audio embedding for depression detection," in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 135–143.

[220] A. Othmani, D. Kadoch, K. Bentounes, E. Rejaibi, R. Alfred, and A. Hadid, "Towards robust deep neural networks for affect and depression recognition from speech," in *International Conference on Pattern Recognition*. Springer, 2021, pp. 5–19.

[221] S. Chancellor and M. De Choudhury, "Methods in predictive techniques for mental health status on social media: a critical review," *NPJ digital medicine*, vol. 3, no. 1, pp. 1–11, 2020.

[222] S. Chancellor, Z. Lin, E. L. Goodman, S. Zerwas, and M. De Choudhury, "Quantifying and predicting mental illness severity in online pro-eating disorder communities," in *Proceedings of the 19th ACM confer-*

ence on computer-supported cooperative work & social computing, 2016, pp. 1171–1184.
[223] M. De Choudhury and S. De, "Mental health discourse on reddit: Self-disclosure, social support, and anonymity," in *Eighth international AAAI conference on weblogs and social media*, 2014.
[224] T. Nguyen, D. Phung, B. Dao, S. Venkatesh, and M. Berk, "Affective and content analysis of online depression communities," *IEEE Transactions on Affective Computing*, vol. 5, no. 3, pp. 217–226, 2014.
[225] J. Wolohan, M. Hiraga, A. Mukherjee, Z. A. Sayyed, and M. Millard, "Detecting linguistic traces of depression in topic-restricted text: Attending to self-stigmatized depression with nlp," in *Proceedings of the first international workshop on language cognition and computational models*, 2018, pp. 11–21.
[226] S. Rude, E.-M. Gortner, and J. Pennebaker, "Language use of depressed and depression-vulnerable college students," *Cognition & Emotion*, vol. 18, no. 8, pp. 1121–1133, 2004.
[227] M. Nadeem, "Identifying depression on twitter," *arXiv preprint arXiv:1607.07384*, 2016.
[228] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting depression via social media," in *Seventh international AAAI conference on weblogs and social media*, 2013.
[229] F. Magami and L. A. Digiampietri, "Automatic detection of depression from text data: A systematic literature review," in *XVI Brazilian Symposium on Information Systems*, 2020, pp. 1–8.
[230] H.-C. Shing, S. Nair, A. Zirikly, M. Friedenberg, H. Daumé III, and P. Resnik, "Expert, crowdsourced, and machine assessment of suicide risk via online postings," in *Proceedings of the fifth workshop on computational linguistics and clinical psychology: from keyboard to clinic*, 2018, pp. 25–36.
[231] S. Smys and J. S. Raj, "Analysis of deep learning techniques for early detection of depression on social media network-a comparative study," *Journal of trends in Computer Science and Smart technology (TCSST)*, vol. 3, no. 01, pp. 24–39, 2021.
[232] Z. Peng, Q. Hu, and J. Dang, "Multi-kernel svm based depression recognition using social media data," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 1, pp. 43–57, 2019.
[233] M. M. Aldarwish and H. F. Ahmad, "Predicting depression levels using social media posts," in *2017 IEEE 13th international Symposium on Autonomous decentralized system (ISADS)*. IEEE, 2017, pp. 277–280.
[234] J. C. Eichstaedt, R. J. Smith, R. M. Merchant, L. H. Ungar, P. Crutchley, D. Preoţiuc-Pietro, D. A. Asch, and H. A. Schwartz, "Facebook language predicts depression in medical records," *Proceedings of the National Academy of Sciences*, vol. 115, no. 44, pp. 11 203–11 208, 2018.
[235] A. E. Tate, R. C. McCabe, H. Larsson, S. Lundström, P. Lichtenstein, and R. Kuja-Halkola, "Predicting mental health problems in adolescence using machine learning techniques," *PloS one*, vol. 15, no. 4, p. e0230389, 2020.
[236] S. Kwakernaak, K. van Mens, W. Cahn, R. Janssen, G. Investigators *et al.*, "Using machine learning to predict mental healthcare consumption in non-affective psychosis," *Schizophrenia research*, vol. 218, pp. 166–172, 2020.
[237] G. Shen, J. Jia, L. Nie, F. Feng, C. Zhang, T. Hu, T.-S. Chua, and W. Zhu, "Depression detection via harvesting social media: A multimodal dictionary learning solution." in *IJCAI*, 2017, pp. 3838–3844.
[238] A. H. Orabi, P. Buddhitha, M. H. Orabi, and D. Inkpen, "Deep learning for depression detection of twitter users," in *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, 2018, pp. 88–97.
[239] J. Ive, G. Gkotsis, R. Dutta, R. Stewart, and S. Velupillai, "Hierarchical neural model with attention mechanisms for the classification of social media text related to mental health," in *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, 2018, pp. 69–77.
[240] T. Shen, J. Jia, G. Shen, F. Feng, X. He, H. Luan, J. Tang, T. Tiropanis, T. S. Chua, and W. Hall, "Cross-domain depression detection via harvesting social media." International Joint Conferences on Artificial Intelligence, 2018.
[241] G. Rao, Y. Zhang, L. Zhang, Q. Cong, and Z. Feng, "Mgl-cnn: a hierarchical posts representations model for identifying depressed individuals in online forums," *IEEE Access*, vol. 8, pp. 32 395–32 403, 2020.
[242] X. Hu, J. Shu, and Z. Jin, "Depression tendency detection model for weibo users based on bi-lstm," in *2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*. IEEE, 2021, pp. 785–790.
[243] P. L. Sheridan and J. M. Hausdorff, "The role of higher-level cognitive function in gait: executive dysfunction contributes to fall risk in alzheimer's disease," *Dementia and geriatric cognitive disorders*, vol. 24, no. 2, pp. 125–137, 2007.
[244] S. Xu, J. Fang, X. Hu, E. Ngai, Y. Guo, V. Leung, J. Cheng, and B. Hu, "Emotion recognition from gait analyses: Current research and future directions," *arXiv preprint arXiv:2003.11461*, 2020.
[245] L. Sloman, M. Pierrynowski, M. Berridge, S. Tupling, and J. Flowers, "Mood, depressive illness and gait patterns," *The Canadian Journal of Psychiatry*, vol. 32, no. 3, pp. 190–193, 1987.
[246] T. C. Brandler, C. Wang, M. Oh-Park, R. Holtzer, and J. Verghese, "Depressive symptoms and gait dysfunction in the elderly," *The American Journal of Geriatric Psychiatry*, vol. 20, no. 5, pp. 425–432, 2012.
[247] M. B. Van Iersel, A. Haitsma, M. Olde Rikkert, and C. Benraad, "Quantitative gait analysis to detect gait disorders in geriatric patients with depression," *J Am Geriatr Soc*, vol. 53, no. 8, pp. 1441–2, 2005.
[248] J. Michalak, N. Troje, and T. Heidenreich, "The effects of mindfulness-based cognitive therapy on depressive gait patterns," *Journal of Cognitive and Behavioral Psychotherapies*, vol. 11, no. 1, pp. 13–27, 2011.
[249] S. Radovanović, M. Jovičić, N. P. Marić, and V. Kostić, "Gait characteristics in patients with major depression performing cognitive and motor tasks while walking," *Psychiatry research*, vol. 217, no. 1-2, pp. 39–46, 2014.
[250] A. S. Naidu, A. Vasudev, A. M. Burhan, E. Ionson, and M. Montero-Odasso, "Does dual-task gait differ in those with late-life depression versus mild cognitive impairment?" *The American Journal of Geriatric Psychiatry*, vol. 27, no. 1, pp. 62–72, 2019.
[251] M. R. Lemke, T. Wendorff, B. Mieth, K. Buhl, and M. Linnemann, "Spatiotemporal gait patterns during over ground locomotion in major depression compared with healthy controls," *Journal of psychiatric research*, vol. 34, no. 4-5, pp. 277–283, 2000.
[252] J. Michalak, N. F. Troje, J. Fischer, P. Vollmar, T. Heidenreich, and D. Schulte, "Embodiment of sadness and depression—gait patterns associated with dysphoric mood," *Psychosomatic medicine*, vol. 71, no. 5, pp. 580–587, 2009.
[253] M. Karg, K. Kühnlenz, and M. Buss, "Recognition of affect based on gait patterns," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, no. 4, pp. 1050–1061, 2010.
[254] G. Venture, H. Kadone, T. Zhang, J. Grèzes, A. Berthoz, and H. Hicheur, "Recognizing emotions conveyed by human gait," *International Journal of Social Robotics*, vol. 6, no. 4, pp. 621–632, 2014.
[255] F. Gholami, D. A. Trojan, J. Kövecses, W. M. Haddad, and B. Gholami, "A microsoft kinect-based point-of-care gait assessment framework for multiple sclerosis patients," *IEEE journal of biomedical and health informatics*, vol. 21, no. 5, pp. 1376–1385, 2016.
[256] M. J. Nordin and A. Saadoon, "A survey of gait recognition based on skeleton model for human identification," *Research Journal of Applied Sciences, Engineering and Technology*, vol. 12, no. 7, pp. 756–763, 2016.
[257] Y. Wang, J. Sun, J. Li, and D. Zhao, "Gait recognition based on 3d skeleton joints captured by kinect," in *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016, pp. 3151–3155.
[258] T. Verlekar, "Gait analysis in unconstrained environments," Ph.D. dissertation, PhD thesis, Ph. D. dissertation, Electrical and Computer Engineering …, 2019.
[259] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 2, pp. 316–322, 2005.
[260] X. Yang, Y. Zhou, T. Zhang, G. Shu, and J. Yang, "Gait recognition based on dynamic region analysis," *Signal Processing*, vol. 88, no. 9, pp. 2350–2356, 2008.
[261] O. Tuzel, F. Porikli, and P. Meer, "Human detection via classification on riemannian manifolds," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2007, pp. 1–8.
[262] Y. Wang, J. Wang, X. Liu, and T. Zhu, "Detecting depression through gait data: examining the contribution of gait features in recognizing depression," *Frontiers in psychiatry*, vol. 12, p. 661213, 2021.
[263] W. Li, Q. Wang, X. Liu, and Y. Yu, "Simple action for depression detection: using kinect-recorded human kinematic skeletal data," *BMC psychiatry*, vol. 21, no. 1, pp. 1–11, 2021.
[264] W. Shao, Z. You, L. Liang, X. Hu, C. Li, W. Wang, and B. Hu, "A multi-modal gait analysis-based depression detection system," *IEEE journal of biomedical and health informatics*, 2021.
[265] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 07, 2020, pp. 13 001–13 008.

[266] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le, "Autoaugment: Learning augmentation policies from data," *arXiv preprint arXiv:1805.09501*, 2018.

[267] ——, "Autoaugment: Learning augmentation strategies from data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 113–123.

[268] Z. Hussain, F. Gimenez, D. Yi, and D. Rubin, "Differential data augmentation techniques for medical imaging classification tasks," in *AMIA annual symposium proceedings*, vol. 2017. American Medical Informatics Association, 2017, p. 979.

[269] J. Yang, H. Lu, C. Li, X. Hu, and B. Hu, "Data augmentation for depression detection using skeleton-based gait information," *arXiv preprint arXiv:2201.01115*, 2022.

[270] T. Wang, C. Li, C. Wu, C. Zhao, J. Sun, H. Peng, X. Hu, and B. Hu, "A gait assessment framework for depression detection using kinect sensors," *IEEE Sensors Journal*, vol. 21, no. 3, pp. 3260–3270, 2021.

[271] J. Fang, T. Wang, C. Li, X. Hu, E. Ngai, B.-C. Seet, J. Cheng, Y. Guo, and X. Jiang, "Depression prevalence in postgraduate students and its association with gait abnormality," *IEEE Access*, vol. 7, pp. 174 425–174 437, 2019.

[272] H. Lu, S. Xu, X. Hu, E. Ngai, Y. Guo, W. Wang, and B. Hu, "Postgraduate student depression assessment by multimedia gait analysis," *IEEE MultiMedia*, 2022.

[273] Y. Yuan, B. Li, N. Wang, Q. Ye, Y. Liu, and T. Zhu, "Depression identification from gait spectrum features based on hilbert-huang transform," in *International Conference on Human Centered Computing*. Springer, 2018, pp. 503–515.

[274] N. Zhao, Z. Zhang, Y. Wang, J. Wang, B. Li, T. Zhu, and Y. Xiang, "See your mental state from your walk: Recognizing anxiety and depression through kinect-recorded gait data," *PLoS one*, vol. 14, no. 5, p. e0216591, 2019.

[275] S. Alghowinem, R. Goecke, M. Wagner, J. Epps, M. Hyett, G. Parker, and M. Breakspear, "Multimodal depression detection: fusion analysis of paralinguistic, head pose and eye gaze behaviors," *IEEE Transactions on Affective Computing*, vol. 9, no. 4, pp. 478–490, 2016.

[276] C. Sobin and H. A. Sackeim, "Psychomotor symptoms of depression," *American Journal of Psychiatry*, vol. 154, no. 1, pp. 4–17, 1997.

[277] H. Ellgring, *Non-verbal communication in depression*. Cambridge University Press, 2007.

[278] J. Mackintosh, R. Kumar, and T. Kitamura, "Blink rate in psychiatric illness," *The British Journal of Psychiatry*, vol. 143, no. 1, pp. 55–57, 1983.

[279] R. Gupta, N. Malandrakis, B. Xiao, T. Guha, M. Van Segbroeck, M. Black, A. Potamianos, and S. Narayanan, "Multimodal prediction of affective dimensions and depression in human-computer interactions," in *Proceedings of the 4th international workshop on audio/visual emotion challenge*, 2014, pp. 33–40.

[280] L. Sörnmo and P. Laguna, *Bioelectrical signal processing in cardiac and neurological applications*. Academic press, 2005, vol. 8.

[281] Z. He, Z. Li, F. Yang, L. Wang, J. Li, C. Zhou, and J. Pan, "Advances in multimodal emotion recognition based on brain–computer interfaces," *Brain sciences*, vol. 10, no. 10, p. 687, 2020.

[282] M. Haghighat, M. Abdel-Mottaleb, and W. Alhalabi, "Discriminant correlation analysis: Real-time feature level fusion for multimodal biometric recognition," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 9, pp. 1984–1996, 2016.

[283] J. Yang, J.-y. Yang, D. Zhang, and J.-f. Lu, "Feature fusion: parallel strategy vs. serial strategy," *Pattern recognition*, vol. 36, no. 6, pp. 1369–1381, 2003.

[284] J. Zhu, Y. Wang, R. La, J. Zhan, J. Niu, S. Zeng, and X. Hu, "Multimodal mild depression recognition based on eeg-em synchronization acquisition network," pp. 28 196–28 210, 2019.

[285] J. Zhu, S. Wei, X. Xie, C. Yang, Y. Li, X. Li, and B. Hu, "Content-based multiple evidence fusion on eeg and eye movements for mild depression recognition," *Computer Methods and Programs in Biomedicine*, p. 107100, 2022.

[286] G. E. Bruder, R. Bansal, C. E. Tenke, J. Liu, X. Hao, V. Warner, B. S. Peterson, and M. M. Weissman, "Relationship of resting eeg with anatomical mri measures in individuals at high and low risk for depression," *Human Brain Mapping*, vol. 33, no. 6, pp. 1325–1333, 2012.

[287] X. Zhang, J. Shen, Z. ud Din, J. Liu, G. Wang, and B. Hu, "Multimodal depression detection: Fusion of electroencephalography and paralinguistic behaviors using a novel strategy for classifier ensemble," *IEEE journal of biomedical and health informatics*, vol. 23, no. 6, pp. 2265–2275, 2019.

[288] C. Jing, X. Liu, N. Zhao, and T. Zhu, "Different performances of speech and natural gait in identifying anxiety and depression," in *International Conference on Human Centered Computing*. Springer, 2019, pp. 200–210.

[289] M. Mousavian, J. Chen, Z. Traylor, and S. Greening, "Depression detection from smri and rs-fmri images using machine learning," *Journal of Intelligent Information Systems*, vol. 57, no. 2, pp. 395–418, 2021.

[290] L. He, M. Niu, P. Tiwari, P. Marttinen, R. Su, J. Jiang, C. Guo, H. Wang, S. Ding, and Z. Wang, "Deep learning for depression recognition with audiovisual cues: A review," 2021.

[291] L. He, D. Jiang, and H. Sahli, "Automatic depression analysis using dynamic facial appearance descriptor and dirichlet process fisher encoding," *IEEE Transactions on Multimedia*, vol. 21, no. 6, pp. 1476–1486, 2019.

[292] S. Jayawardena, J. Epps, and E. Ambikairajah, "Ordinal logistic regression with partial proportional odds for depression prediction," *IEEE Transactions on Affective Computing*, vol. PP, no. 99, pp. 1–1, 2020.

[293] Y. Le, D. Jiang, W. Han, and H. Sahli, "Dcnn and dnn based multimodal depression recognition," in *International Conference on Affective Computing & Intelligent Interaction*, 2017.

[294] Y. Le, D. Jiang, and S. Hichem, "Integrating deep and shallow models for multi-modal depression analysis — hybrid architectures," *IEEE Transactions on Affective Computing*, pp. 1–1, 2018.

[295] L. He, C. Guo, P. Tiwari, H. M. Pandey, and W. Dang, "Intelligent system for depression scale estimation with facial expressions and case study in industrial intelligence," *International Journal of Intelligent Systems*, no. 4, 2021.

[296] L. Yang, Y. Li, H. Chen, D. Jiang, M. C. Oveneke, and H. Sahli, "Bipolar disorder recognition with histogram features of arousal and body gestures," 2018, pp. 15–21.

[297] W. Fan, Z. He, X. Xing, B. Cai, and W. Lu, "Multi-modality depression detection via multi-scale temporal dilated cnns," in *the 9th International*, 2019.

[298] L. Zhang, J. Driscol, X. Chen, and R. H. Ghomi, "Evaluating acoustic and linguistic features of detecting depression sub-challenge dataset," in *the 9th International*, 2019.

[299] S. Yin, C. Liang, H. Ding, and S. Wang, "A multi-modal hierarchical recurrent neural network for depression detection," in *the 9th International*, 2019.

[300] M. R. Makiuchi, T. Warnita, K. Uto, and K. Shinoda, "Multimodal fusion of bert-cnn and gated cnn representations for depression detection," 2019.

[301] L. Yang, D. Jiang, X. Xia, E. Pei, M. C. Oveneke, and H. Sahli, "Multimodal measurement of depression using deep learning models," in *Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge*, 2017, pp. 53–59.

[302] Z. Zhao, Z. Bao, Z. Zhang, J. Deng, and B. Schuller, "Automatic assessment of depression from speech via a hierarchical attention transfer network and attention autoencoders," *IEEE Journal of Selected Topics in Signal Processing*, vol. PP, no. 99, pp. 1–1, 2019.

[303] M. Niu, J. Tao, B. Liu, J. Huang, and Z. Lian, "Multimodal spatiotemporal representation for automatic depression level detection," *IEEE Transactions on Affective Computing*, vol. PP, no. 99, pp. 1–1, 2020.

[304] A. Jan, H. Meng, Y. F. B. A. Gaus, and F. Zhang, "Artificial intelligent system for automatic depression level analysis through visual and vocal expressions," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 10, no. 3, pp. 668–680, 2017.