Computers in Biology and Medicine 165 (2023) 107457

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Detecting depression based on facial cues elicited by emotional stimuli in video

Bin Hu, Yongfeng Tao, Minqiang Yang *

Gansu Provincial Key Laboratory of Wearable Computin, Lanzhou University, Lanzhou, 730000, Gansu, China

ARTICLE INFO

Keywords:
Depression detection
Local face
Emotional stimulation
Facial expressions

ABSTRACT

Recently, depression research has received considerable attention and there is an urgent need for objective and validated methods to detect depression. Depression detection based on facial expressions may be a promising adjunct to depression detection due to its non-contact nature. Stimulated facial expressions may contain more information that is useful in detecting depression than natural facial expressions. To explore facial cues in healthy controls and depressed patients in response to different emotional stimuli, facial expressions of 62 subjects were collected while watching video stimuli, and a local face reorganization method for depression detection is proposed. The method extracts the local phase pattern features, facial action unit (AU) features and head motion features of a local face reconstructed according to facial proportions, and then fed into the classifier for classification. The classification accuracy was 76.25%, with a recall of 80.44% and a specificity of 83.21%. The results demonstrated that the negative video stimuli in the single-attribute stimulus analysis were more effective in eliciting changes in facial expressions in both healthy controls and depressed patients. Fusion of facial features under both neutral and negative stimuli was found to be useful in discriminating between healthy controls and depressed individuals. The Pearson correlation coefficient (PCC) showed that changes in the emotional stimulus paradigm were more strongly correlated with changes in subjects' facial AU when exposed to negative stimuli compared to stimuli of other attributes. These results demonstrate the feasibility of our proposed method and provide a framework for future work in assisting diagnosis.

1. Introduction

Mental health concerns and psychological disorders are common and can lead to disability and suffering [1]. Depression is one of the most common mental disorders, and it is estimated by the World Health Organization (WHO) that by 2030, depression will become the first leading cause of the increasing global disease burden [2]. For a long time, only a small percentage of people with major depressive disorder (MDD) have been treated [3], and they can suffer from recurrent episodes throughout their lives [4]. Approximately 350 million people worldwide are suffering from depression [5]. The lifetime risk of depression is 15%–18%, and the worst outcome is that many people end up taking their own lives, with 850,000 suicides reported each year. People with depression and anxiety are at higher risk of suicide, and early identification and treatment of these disorders can help prevent suicide. Study [6] showed that the strongest predictors of suicidal ideation and behavior four years later were previous anxiety symptoms, depressive symptoms, etc. Despite efforts to identify and treat depression, its incidence may be on the rise, especially among young people. The current evaluation methods for the diagnosis of depression mostly rely on the severity of depressive symptoms reported by patients [7,8], and clinicians have certain subjective biases in their judgment of the severity of symptoms [9]. Therefore, there is an urgent need for an objective and efficient method to predict the clinical outcome of depression, which can greatly alleviate the problems caused by the extremely low doctor–patient ratio.

Using behavioral [10,11] and physiological [12] data to model and detect depression is an objective and valid approach. Unlike fingerprints, pupil images [13], or electroencephalograms [14], facial images can be obtained quickly and without physical contact. It is generally accepted that facial data is collected in a non-invasive or non-threatening manner by facial recognition devices [15]. Facial expressions have been studied for a long time, and there are have a large number of studies [16–18] on the recognition of human emotions through facial information. Previous studies on facial expressions have made it possible to identify depression through facial information, and there are many literature on the detection of depression through facial expressions at the present stage [19–21]. Detecting depression through facial expressions is a convenient and simple, non-invasive method.

* Corresponding author.
E-mail addresses: bh@lzu.edu.cn (B. Hu), taoyf21@lzu.edu.cn (Y. Tao), yangmq@lzu.edu.cn (M. Yang).

https://doi.org/10.1016/j.compbiomed.2023.107457
Received 14 June 2023; Received in revised form 11 August 2023; Accepted 4 September 2023
Available online 6 September 2023

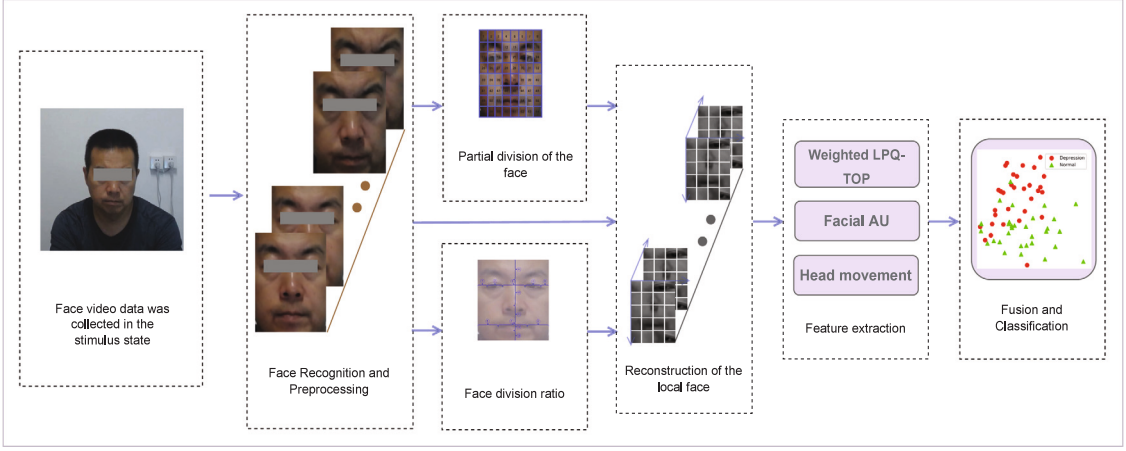


Fig. 1. Flow chart for depression detection. First, face detection and pre-processing are performed on the video stream data. Then, local regions of the face are extracted and combined to form a locally recombined face. Finally, the extracted features of the locally recombined face are fed into the classifier to detect depression. To address privacy concerns, mosaics are added to the eye region.

Previous studies have used the whole face [22,23] to extract effective features. Although different methods are used to extract recognition features from faces, they use whole faces to introduce more information that is not associated with expression recognition and even more interference information. To overcome this problem, many studies use facial landmarks [11,24–26] to reduce redundant information in faces. However, the facial landmark-based approach uses a small number of facial landmarks to represent the entire face, losing some facial details. Regions of interest (ROI) are areas or objects of interest in an image that can be used for disease detection, for example in applications such as medical image segmentation [27–30] and depression detection [31]. Extracting ROI in facial data for depression detection is important not only to reduce the effect of redundant information, but also to improve the classification accuracy of the model. Some researchers use partial facial information [20,32,33] to alleviate the above problems. However, how to make an accurate local facial division, how to use the divided facial regions, and how to solve the correlation among the divided facial regions. We found that previous studies did not consider the proportion of distance among key parts of the face and also ignored the synergistic relationship among different ROI in the face.

In view of the above limitations, this paper proposes a depression detection method based on partial partition and recombination of video streaming facial information. This method can locate the facial ROI more accurately and can obtain a partial recombination face by reconstructing the extracted facial ROI. Specifically, we first extracted and preprocessed faces from the video stream to eliminate interference factors such as illumination in the data. Secondly, inspired by previous studies [34], we accurately extracted local facial ROI according to the facial regions where depression-related facial AUs are located and the proportion among facial key parts in the average face defined by us (as shown in Fig. 1). Finally, the local phase pattern (LPQ) features with the temporal dimensions, facial AU features, and head movement features are extracted for classification. Identification model validation was performed on the depression data we collected. The contributions of this paper are as follows:

- We provide a new method to identify depressive disorder through video streaming facial information. The method provides a new idea for removing redundant information from face data in video streams.
- Comparing depression-related facial AU, a locally recombined face representation method is proposed, which effectively reduced the redundant facial information. The method showed better performance in identifying depressed patients.
- We defined a proportion of facial key points that can be used to accurately extract depression features in the face and can eliminate the effect of different proportions on different parts of the face.
- The multimodal fusion feature of facial local features extracted from different attributes of stimulation can improve the accuracy of depression recognition.

The aim of this paper is to propose a depression detection model based on local reconstruction of faces under video stimuli with different attributes. Face video streaming data is used to accurately identify and reconstruct the ROI of a face and to extract spatio-temporal features for better depression detection. Section 2 discusses the relevant work of current research. In Section 3, materials and methods used in our work are described. The results and discussions are presented in Section 4. Section 5 draws the conclusion of this paper and future work.

2. Related work

People with depression tend to have facial expressions that are contemptuous [34], low in pleasure, and highly pleasurable expressions are reduced or absent. They also have difficulty accurately identifying basic emotions in facial expressions [35–37] and are less accurate at decoding emotions than healthy people [38]. Hand-crafted [39] features and deep learning models [11,40,41] to extract features are two common methods for detecting depression, but deep learning requires a large amount of data to train the model, which can be alleviated by pre-training the model. The aim of this paper is to solve the problem of local facial region delineation in depression detection by combining facial AU and accurately acquiring the ROI region, which reduces the redundant information in the face.

Facial AUs are the basic movement of muscle groups or individual muscles. It was first proposed by Ekman et al. [42] and then adopted by Cohn et al. [43] to analyze depressive states. AU has been used to assess the severity of depression and has achieved promising results [44,45]. Cohn et al. selected 17 predefined AUs from facial video data and recognized depression by using four parameters such as proportion and average duration of these AUs in the video [43]. Stratou et al. [46] recognized depressed patients by using facial AU related to basic expressions. Girard et al. [47] found that when depression was severe, participants had fewer AU12, AU15, and more AU14 in their AU. However, in reality, changes in human expression of intensity and direction were different. Depression analysis using some basic expressions or depression-related facial AU of the video streaming facial information largely misses some of the original useful information in the data.

Table 1
Six AUs are associated with depression.

Facial AU	Description
AU1	Inner Brow Raiser.
AU4	Brow Lowerer.
AU6	Cheek Raiser.
AU12	Lip Corner Puller.
AU15	Lip Corner Depressor.
AU17	Chin Raiser.

Carnegie Mellon University (CMU) is prominent in the research on expression recognition in depression [34]. From the perspective of psychology and facial expression recognition technology, they analyzed 6 kinds of AU related to depression expression, as shown in Table 1. We found that several AUs mainly concentrated in the region of the eyebrow, eye, mouth, and nose area. The above problems in the study of depression, and inspired by the location of depression-related AUs, we use the ratio between key parts of the face to extract local facial areas. This would provide more detailed information about the face than simply using the presence or absence of certain AUs associated with depression as identifying features of depression.

There are studies that show faces are partitioned directly after face alignment, and then corresponding operations are performed on each partition area. However, they ignore the differences among facial features, such as the fact that someone may have a bigger mouth or a longer nose. Liu et al. [32] divided the face into 36 regions of interest and then calculated the information about the main optical flow in each ROI. And that introduces a new problem, which is that it takes into account facial areas that are less relevant to expression. To solve the above problems, we first defined the facial proportion of distance among the key parts of the subjects (Fig. 3(a)) and divided the face into 64 regions on average (Fig. 3(b)). We then used the defined proportions to accurately extract 20 regions of interest face blocks. In this way, the key regions of the face are accurately extracted, and redundant information is also removed. To retain the interaction among regions of interest, we reassemble them according to their original relative positions to form locally reconstructed faces.

3. Materials and methods

3.1. Participants

Sixty-two subjects were recruited from the Second People's Hospital of Gansu Province (shown in Table 2), including 31 patients (11 males, 20 females; Aged 18–55) with depression and 31 normal subjects (15 males and 16 females; 18 to 55 years old). All subjects with normal or corrected vision were right-handed and had no neurological disease or cognitive impairment. The PHQ9 and PHQ15 are two commonly used questionnaires to assess patients' depressive symptoms and physical symptoms, and there is a relationship between the PHQ9 and PHQ15 scores. Specifically, the more severe the depressive symptoms, the more severe the physical symptoms. As shown in Fig. 2, as the PHQ9 score increases or decreases, the PHQ15 score increases accordingly. Before the experiment, all subjects signed informed consent and the experiment was approved by the Ethics Committee of The Second People's Hospital of Gansu Province. According to the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) [48] criteria, all patients underwent a structured Mini-International Neuropsychiatric interview (MINI) [49], which met the diagnostic criteria for major depressive disorder.

3.2. Emotional face stimuli task

Compared with healthy people, depressed patients have a weaker response to positive emotions (such as happiness) and a stronger response to negative emotions (such as sadness, fear, and anger), and

Table 2
Demographic variables for participants with depressive disorder (DP) and healthy controls (HC).

Variables	DP	HC	P
N(%female)	31(64.5)	31(51.6)	0.303
Age(years)	41(±9.58)	43(±8.75)	0.108
PHQ9	16.14(±5.05)	2.11(±1.50)	0.000**
PHQ15	12.21(±4.72)	2.71(±1.88)	0.000**

Note: Means are displayed standard deviations in parentheses. The last column shows the p-values for the χ^2-test (for gender) and the t-test.
** Indicates $p < 0.001$.

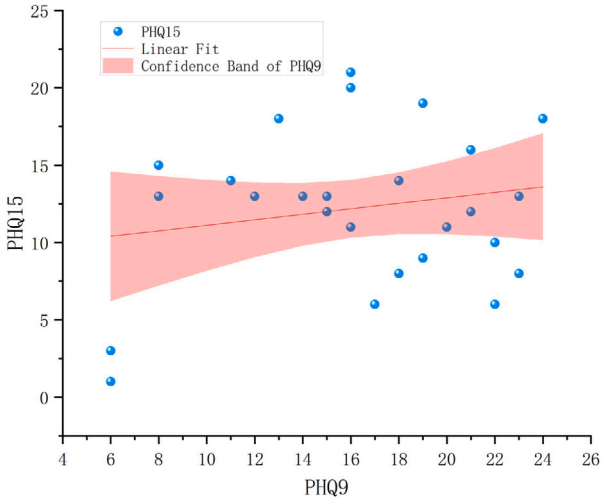


Fig. 2. A linear fit between PHQ15 scores and PHQ9 scores.

this deviation is related to the severity of depression in patients' and dysfunctional symptoms [50]. They found that people with depression rated the negative emotions they felt (fear and sadness) higher when stimulated by happy sounds. It is mentioned in DSM-IV that the existence of depression can be inferred from facial expressions and behaviors, and that the recognition of depression through facial expressions has attracted the attention of researchers [20,21]. People's psychological states can be reflected in their expressions. Therefore, emotional stimulation is needed to make people show a simple and subtle expression.

Previous studies have shown that video stimuli can be more effective than static pictures in inducing changes in subjects' emotions and facial expressions. Culture-specific differences may lead to underreporting or misreporting of psychological distress [51]. The aim of this study was to detect depression in Chinese subjects using positive, neutral and negative culturally specific (Chinese film) induced mood changes. The selection of stimulus materials in the experimental paradigm referred to the design process of emotion Analysis database (DEAP) [52] based on physiological signals and SJTU Emotion EEG data set (SEED) [53]. Three Chinese film clips with high arousal and emotional emphasis were selected from the Chinese Emotional Image Sourcebook (CEVS) [54] as positive, negative and neutral stimulus materials for the experimental design. Each stimulus video lasted about 2 min. At the beginning of the experiment, a voice prompted the subjects to prepare to watch a video. After watching one video, the subjects were given a 2-minute break (to relax and regulate their emotions) before watching the next video. In particular, the order of stimuli was randomized for each participant to avoid sorting effects.

3.3. Facial data acquisition

Our facial stimulation was composed of three Chinese film clips with different stimulus attributes. As the subjects watched the film

clips, their facial expressions data were recorded by a logitech CC2000e camera with a resolution of 1920 * 1080 and 30 fps. The facial expression data was stored separately according to different film clips they watched. The whole recording process was carried out in a quiet, clean and comfortable air-conditioned room. The subject sat on a chair and kept the distance between his head and the video playing screen at 50–60 cm. The experiment lasted for about 10 min for one subject. Facial expression data from 62 subjects were used in this study, with each subject having three facial video data stimulated by different attribute stimulus materials.

3.4. Data preprocessing

Due to various external conditions, there is a lot of noise in the face picture (such as shooting angle). In addition, the light and facial noise have an influence on subsequent extraction of facial features. Therefore, it is very important to preprocess the face images to reduce the effects of illumination, occlusions, etc. The process of image preprocessing generally includes histogram equalization, normalization, and smoothing processing.

• **Histogram equalization:** The color face images are susceptible to different illuminations, so it is necessary to grayscale the image. General color image is RGB color mode (is a color standard in the industry), this color mode can be obtained by the change and overlay of three color channels (red (R), green (G), and blue (B)). There are many ways to convert RGB images to grayscale images. The formula used in this article is as follows:

$$\text{Gray} = \text{R}^*0.299 + \text{G}^*0.587 + \text{B}^*0.114 \tag{1}$$

where R, G, and B range from 0 to 255.

The grayscale normalization of the image can reduce the interference of the information in the face due to light changes. Generally, using the histogram equalization processing method, the processed image will have higher contrast than before, and the details of the picture will be clearer. In the histogram equalization operation, the grayscale histogram of the image needs to be calculated, which mainly reflects the relationship between each gray value in an image and the frequency of that gray value in the image. The gray histogram of an image is a one-dimensional discrete function, which can be written as:

$$h(k) = n_k, \quad k = 0, 1, \ldots, L-1 \tag{2}$$

where n_k is the number of pixels with gray level k in the image $f(x, y)$. The height of each column of the histogram (called bin) corresponds to n_k. Histogram provides all kinds of grey value distribution in the original image. On the basis of the histogram, further define the normalized histogram into a grayscale relative frequency of $P_r(k)$:

$$P_r(k) = n_k / N \tag{3}$$

where N denotes the total number of $f(x, y)$ pixels in the image and n_k denotes the number of pixels in the image with gray level k. Through this equalization process, the image pixels will be evenly distributed on a certain gray scale, and then the face image will be recalculated.

• **Normalization:** Face normalization includes geometric normalization and gray scale normalization. Proper normalization prior to the feature extraction step will reduce the intra-class variation per subject and help improve the robustness of the recognition.

Geometric normalization includes normalization of both angles and scales. For angle normalization, we process data mainly based on data characteristics. OpenFace [55] is used to calculate the center coordinates of left and right eyes respectively. Then the angle θ between the connecting line (between the left and right eyes) and the horizontal line is then calculated, and the coordinates of the center of the connecting line are calculated. Finally, based on the center coordinates, the face picture $f(x, y)$ is rotated by θ, so that the line between the left and right eyes is level.

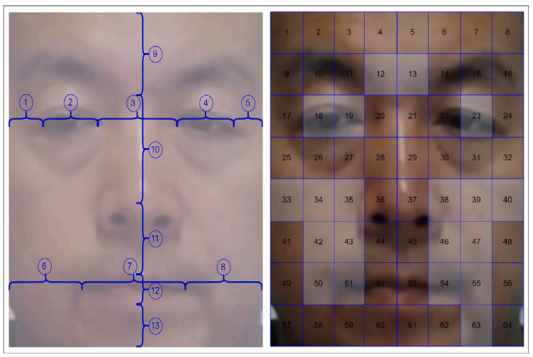


Fig. 3. The distance ratio of face key points (a) and local face area division (b).

The scale normalization adjusts the size of the image to ensure that faces of different sizes are consistent in size. Scale normalization needs to be achieved by scaling the image. Image scaling is essentially the mapping of pixel points from one image to another through a geometric space transformation of pixels. Face image scaling can be calculated according to the following matrix scaling formula:

$$\begin{bmatrix} x & y & 1 \end{bmatrix} = \begin{bmatrix} x_0 & y_0 & 1 \end{bmatrix} \begin{bmatrix} r_x & 0 & 0 \\ 0 & r_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4}$$

where r_x and r_y respectively represent the image length and width scaling ratio. The gray value interpolation operation of the input image pixels is usually used to achieve the gray value interpolation of the output pixels. The commonly used interpolation methods include nearest neighbor interpolation and bilinear interpolation.

• **Smoothing:** Face images are subjected to external factors such as interference, uneven lighting, and noise at the pixel level. The noise interferes with the effective acquisition of information in face images, so we need to smooth the initial image operations.

Smoothing technique is generally divided into two categories: the first category is the global processing of the whole image; the second type uses local operators on the image to smooth a pixel point by computing only the pixel points in the region around it, which optimizes the computational efficiency and allows smoothing multiple pixel points at the same time. We use the median filter in the local smoothing technology for smoothing. The core idea is to start with a point in the initial image, then sort all the pixels around the neighborhood centered on the point, and take their median value as the median of the response of the point.

3.5. Patch generation

When depression analysis is performed from facial regions, local plaques may have discriminatory features for estimating the severity of depression. In [56,57], the authors considered that the multi-view information of the highlighted regions is important for image retrieval. We proposed a method for local facial regions acquisition according to depression-related regions, with the following main acquisition steps:

(a) Selection of key points and face interception: face regions and feature points are detected using the Openface toolkit. We selected 13 points to cover the main regions of the face related to depression. The selected points are indexed as follows: 9, 18, 25, 27, 31, 37, 40, 43, 46, 49, 52, 55, 58. The faces in each image are intercepted according to index points 18 (5 pixels more to the left), 27 (5 pixels more to the right), 25 (15 pixels more to the top) and 9, as shown in Fig. 3(b).

(b) Definition of face proportions: To accurately extract the region of interest from the facial region, 13 face key point proportions were defined in this study, as shown in Fig. 3(a). These 13 proportions are

Table 3
Eleven features of head movement.

Extracted features	Formulas	Description
Mean value	$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$	x_i is the head motion sequence coordinate at time i;
Standard deviation	$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$	N is the length of the sequence of operations;
Root amplitude	$X_r = \left(\frac{1}{N}\sum_{i=1}^{N}\sqrt{\lvert x_i\rvert}\right)^2$	$X_p = \max\{\lvert x_i\rvert\}$.
Skewness	$\alpha = \frac{1}{N}\sum_{i=1}^{N} x_i^3$	
Root mean square (RMS)	$X_{rms} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$	
Peak-to-peak value	$X_{pp} = \max\{x_i\} - \min\{x_i\}$	
Kurtosis	$\beta = \frac{1}{N}\sum_{i=1}^{N} x_i^4$	
Crest factor	$C = \frac{X_p}{X_{rms}}$	
Clearance factor	$L = \frac{X_p}{X_r}$	
Shape factor	$S = \frac{X_{rms}}{\lvert\bar{x}\rvert}$	
Impulse factor	$I = \frac{X_p}{\bar{x}}$	

calculated by the coordinates of the selected index points (same as the previous 13 index points) on the average face.

(c) According to the scale defined in Fig. 3(a), the region associated with depression (gray area in Fig. 3(b)) is extracted and reorganized to form the reorganized face region as shown in part reconstruction of the local face of Fig. 1.

3.6. Feature extraction

Chan et al. [58] adopted a newly introduced operator, Local Phase Quantization (LPQ), in fuzzy face recognition [59] to solve the face recognition problem. LPQ operator is based on Fourier transform phase quantization of the local area and is generally considered to have fuzzy invariant properties under specific conditions.

At the position z = (x,y) of each pixel of image $f(x, y)$, a two-dimensional discrete Fourier transform (2D-DFT) defined below is performed for the rectangular neighborhood Nz.

$$F(u, z) = \sum_{w \in N_z} f(z - w)e^{-i2\pi u^T w} \tag{5}$$

2D-DFT is calculated only at four frequency points, where the phase of the DFT is proved to be invariant centrosymmetric blur: u1 = [a, 0], u2 = [0, a], u3 = [a, a] and u4 = [a, a]. where a is a scalar to satisfy this condition and then extract the sign of the real and imaginary parts of each Fourier coefficient, yielding a binary coefficient q_j(x):

$$q_j(\mathrm{x}) = \begin{Bmatrix} 0 & \text{if } g_j(\mathrm{x}) \geq 0 \\ 1 & \text{otherwise} \end{Bmatrix} \tag{6}$$

where g_j(x) is the jth component of vector Gx,

$$\mathrm{Gx} = [\mathrm{Re}\{\mathrm{F(u,z)}\}, \mathrm{lm}\{\mathrm{F(u,z)}\}] \tag{7}$$

The eight binary codes obtained give the phase information. Finally, the F_{LPQ} of the image is obtained by expressing the binary codes as integer values between 0 and 255.

$$F_{LPQ}(\mathrm{x}) = \sum_{j=1}^{8} q_j(\mathrm{x})2^{j-1} \tag{8}$$

Considering the actual situation of faces in our data (face blur), we used LPQ-TOP to extract face features. LPQ-TOP takes the time (t) dimension into consideration on the basis of LPQ and can obtain the changes of facial expressions in the time dimension. This process is similar to Local Binary Patterns on Three Orthogonal Planes (LBP-TOP) after the introduction of time dimensions in Local Binary Patterns (LBP). In this work, we use the same parameter as Ahonen et al.'s [20] paper to fix the length of video blocks so that each subject can form multiple video blocks. The number of video blocks is the total number of video frames divided by the number of frames of a single video block. Studies have shown that features in different directions have different contributions to facial features. Therefore, we used the same method as [60] to weight features in different directions, and finally each video block formed 256-dimensional features. Due to the introduction of temporal information while increasing the feature contribution weights in the xy direction. Therefore, weights of 5, 2 and 1 are assigned to the xy, xt and yt directions respectively. As shown in Fig. 4, Principal Component Analysis (PCA) of different kernel functions (linear kernel, poly kernel, RBF kernel, and Sigmoid kernel) was used for dimensionality reduction and feature visualization under fused stimuli.

We used OpenFace to extract a time series information of human head movement from human face video data. Then the head motion sequence is normalized and 11 time–frequency domain features are extracted, and the detailed information on these features is shown in Table 3.

We also extracted six types of depression-related AU using OpenFace (Table 1), and then the percentage of each AU unit in the total number of frames was calculated as a classification feature of depression. The formula for solving is:

$$P_{AU_i} = \frac{1}{N}\sum_{j=1}^{n} AU_{ij}, i = 1, 2, \ldots, 6 \tag{9}$$

where N represents the length of the face sequence, and n denotes the total number of AU_i.

3.7. The proposed approach

To achieve effective depression detection, we constructed a depression detection framework based on partial face recombination, and the complete process is shown in Fig. 1. For each frame in the video data, we use the OpenCV Haar classifier [61] to detect the facial region, form a face sequence, and then preprocess these sequence through preprocessing processes such as histogram equalization, normalization, and smoothing. Head movements cause shifts and rotations of faces in video data at different time periods, and calculating the fixed proportions among landmarks of individual faces is inaccurate. We overcome the above problem by calculating the ratio among the average face landmarks (Fig. 3(a)). The formula for solving the average face is:

$$f_{\mathrm{mean}} = \frac{1}{T}\left(\sum_{t=1}^{T} f_t(x, y)\right), t = 1, 2, \ldots, T \tag{10}$$

where T represents the length of the face sequence, and $f_t(x, y)$ represents the face image at the time t.

The local face regions of interest are extracted accurately according to a fixed scale and region delimitation rule for each subject's face. In

Fig. 4. Visualization of Fusion Feature Distribution. Each fusion feature uses four different kernel PCA dimensionality reduction methods, and the kernel functions are linear kernel, poly kernel, rbf kernel and sigmoid kernel respectively. The triangle (green) represents data from normal people, and the circle (red) represents data from depressed people. (a) Positive and negative, (b) Neutral and negative, (c) Positive and neutral, (d) Positive, neutral and negative.

Table 4
Accuracy, precision, F1-score, recall, and specificity of SVM classifier for local and whole face features under different stimulus materials, where ± denotes the standard deviation of the results of multiple experiments.

Features	Stimulus	Accuracy	Precision	F1-score	Recall	Specificity
Whole face	Positive	0.5250 ± 0.1119	0.5131 ± 0.1612	0.5589 ± 0.1903	0.6137 ± 0.2576	0.4447 ± 0.2597
	Neutral	0.6125 ± 0.1077	0.5895 ± 0.1116	0.6546 ± 0.1174	0.7360 ± 0.1808	0.4929 ± 0.1737
	Negative	0.5586 ± 0.1256	0.5674 ± 0.1845	0.5358 ± 0.1338	0.5076 ± 0.1757	0.6146 ± 0.2296
Local face	Positive	0.6048 ± 0.1008	0.5763 ± 0.0994	0.6860 ± 0.0972	**0.8471 ± 0.1576**	0.3563 ± 0.1968
	Neutral	0.6384 ± 0.1187	0.6106 ± 0.1295	0.6741 ± 0.1136	0.7524 ± 0.1712	0.5301 ± 0.2098
	Negative	**0.7230 ± 0.1125**	**0.7968 ± 0.1613**	**0.6876 ± 0.1468**	0.6049 ± 0.1671	**0.8438 ± 0.1351**

order to capture as much as possible of the tendency of the subject's face to change with the stimulus paradigm, a longer set of 60 consecutive frames was used to extract facial video features. The local face blocks are formed by superimposing 60 consecutive frames (multiple local face blocks can be formed depending on the total frame length of the video). For each local face block, the LPQ-TOP method is used to extract facial features, and the features in xy, xt, and yt directions are weighted. The method at [62] is used to obtain the blocks with the maximum number of saliency in the local face blocks. Finally, the 256-dimensional local facial features formed for each person were

Table 5
Accuracy, precision, F1-score, recall, and specificity of SVM classifier for local facial features and facial AU features and head movement features under different stimulus materials, where ± denotes the standard deviation of the results of multiple experiments.

Features	Stimulus	Accuracy	Precision	F1-score	Recall	Specificity
Local face	Positive	0.6048 ± 0.1008	0.5763 ± 0.0994	0.6860 ± 0.0972	0.8471 ± 0.1576	0.3563 ± 0.1968
	Neutral	0.6384 ± 0.1187	0.6106 ± 0.1295	0.6741 ± 0.1136	0.7524 ± 0.1712	0.5301 ± 0.2098
	Negative	0.7230 ± 0.1125	0.7968 ± 0.1613	0.6876 ± 0.1468	0.6049 ± 0.1671	0.8438 ± 0.1351
Local face and AU	Positive	0.5826 ± 0.1145	0.5564 ± 0.0971	0.6692 ± 0.1084	0.8393 ± 0.1689	0.3247 ± 0.2002
	Neutral	0.6346 ± 0.1188	0.6065 ± 0.1058	0.6724 ± 0.1093	0.7543 ± 0.1592	0.5151 ± 0.1813
	Negative	0.7250 ± 0.1119	0.8126 ± 0.1681	0.6829 ± 0.1426	0.5889 ± 0.1707	0.8642 ± 0.1584
Local face andhead movement	Positive	0.5990 ± 0.1085	0.5671 ± 0.089	0.6843 ± 0.086	**0.8625 ± 0.1408**	0.3330 ± 0.1574
	Neutral	0.6375 ± 0.1276	0.6078 ± 0.1359	0.6718 ± 0.1236	0.7509 ± 0.1714	0.5276 ± 0.2376
	Negative	**0.7375 ± 0.1065**	**0.8382 ± 0.1777**	0.6949 ± 0.1547	0.5935 ± 0.1789	**0.8851 ± 0.0807**
Local face, AU, andhead movement	Positive	0.6298 ± 0.1157	0.5916 ± 0.1243	**0.7013 ± 0.1250**	0.8609 ± 0.1757	0.3979 ± 0.1895
	Neutral	0.6461 ± 0.1243	0.6119 ± 0.1268	0.6902 ± 0.1239	0.7915 ± 0.1791	0.5024 ± 0.2034
	Negative	0.7317 ± 0.1017	0.8146 ± 0.1466	0.6910 ± 0.1363	0.6005 ± 0.1670	0.8628 ± 0.1376

fused with facial AU and head motion features (simple splicing fusion) for depression detection analysis. Similarly, the trained classification model is constructed using these features, and then the test data is classified.

To ensure the validity of the experimental results, we use 10-fold cross-validation to test the accuracy of the classification. The main idea is to divide the data set into ten parts, and take turns to use nine of them as training data and one as test data. Each test will have a corresponding accuracy rate. The average accuracy rate of the results of 10 times is used as an estimate of the accuracy of the classification. To prevent data leakage, all examples for each participant are assigned to the same fold. In general, to reduce the effect of uneven distribution of data features due to individual differences on model stability, it is necessary to perform multiple 10-fold cross-validation (e.g. 20 times 10-fold cross-validation) and then calculate its average value as an estimate of the accuracy of the algorithm.

4. Results and discussion

We explored and compared methods for detecting depression based on the reorganization of facial regions with the whole face, head movements and facial AU under different stimuli. Firstly, to verify that the effect of partial faces is better than the data input of the whole face, the experiment compares the difference between partial faces and whole faces. Secondly, we introduce the facial AU features and the head movement features. The different effects of these two features on local facial features are explored. Finally, to explore the effect of feature fusion under different stimulus materials on the accuracy of depression detection, we designed exploratory experiments under different stimuli. We used the SVM algorithm for the final classification of the above three experiments. We chose the Radial Basis Function (RBF) kernel and subjected the two parameters (C and γ) to a parameter search in the ranges {0.125, 0.25, 0.5, 1, 2, 4, 8} and {0.003, 0.004, 0.005, 0.006, 0.007}. To verify the performance of the features proposed by us under different classifiers, we selected K-nearest Neighbor (KNN), Random Forest (RF), and SVM classifiers. All algorithms were implemented in Python3.6. The classification evaluation indicators we use include accuracy, precision, F1-score, recall(sensitive), and specificity.

4.1. Comparison between local face and whole face

To verify the validity of the proposed local face recombination model, we designed experiments to compare the performance of local faces and whole faces using SVM as a classifier. As shown in Table 4, the accuracy of local faces under positive, neutral, and negative stimulus materials is higher than that of whole faces, and the classification accuracy of negative stimulus materials (72.30%) was significantly higher (where p values [63] for accuracy, precision, F1-score, recall, and specificity are all $p < 0.001$.) than that of other stimulus materials. The accuracy was low and there was little difference for the whole face

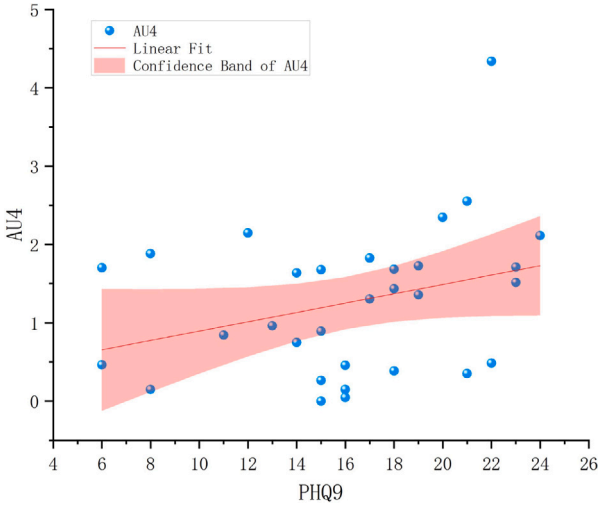


Fig. 5. A linear fit between AU4 intensity and PHQ9 scores under negative stimuli, where AU4 intensity values range from 0 to 5.

with all the stimulus materials. This shows that local facial features can reflect the difference in facial emotion between normal controls and depressed patients under different stimulus materials.

4.2. Ablative analysis of facial features

Some studies have extracted facial AU features for the recognition of depression, which indicates that facial AU features can also reflect people's depression. To verify the effectiveness of facial AU region extraction of local facial features, we set up an exploration experiment for the comparison between facial AU features and local facial features. As shown in Table 5, with positive stimulus materials, the accuracy of local facial features was higher than that of their combined features (using simple feature splicing combinations). Under neutral and negative stimuli, the classification accuracy of combined features is higher than that of local facial features. This indicates that the local facial features extracted by us contain facial AU information. Studies have shown that head movement features can reflect the differences between normal people and depressed people. Therefore, the combined features of head movement features and local facial features are explored, and the results are basically the same as the above experimental results (Table 5). Finally, we combined these three features into new features for classification and found that the results are basically the same as in the above two experiments. This suggests that local facial features can be used to distinguish between normal and depressed people. In addition, we calculated P-values for participants with six facial AUs under positive, neutral and negative emotional stimuli using t-tests. As

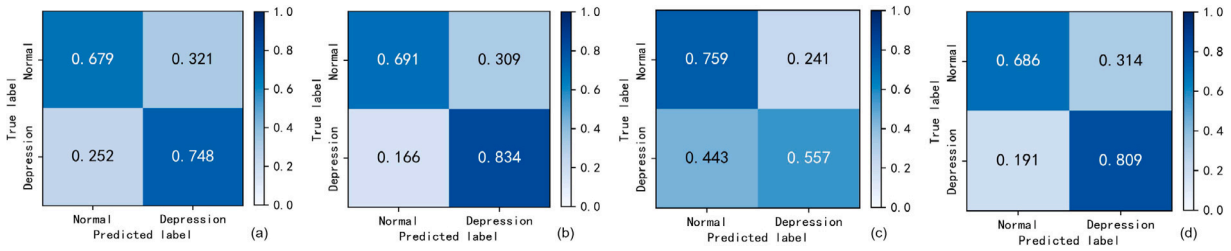


Fig. 6. Confusion matrix for the fusion method (using the same normalized confusion matrix method as in [64]). Each row of the confusion matrices represents the true label and each column represents the predicted label. The element (i, j) is the percentage of samples in class i that is classified as class j. (a) Positive and negative, (b) Neutral and negative, (c) Positive and neutral, (d) Positive, neutral, and negative.

Table 6
P-values for participants with six facial AUs under positive, neutral and negative emotional stimuli.

Facial AU	P-positive	P-neutral	P-negative
AU1	0.3875	0.0128*	0.1168
AU4	0.0002**	0.0001**	0.0011*
AU6	0.2610	0.1221	0.2924
AU12	0.0037*	0.0109*	0.0073*
AU15	0.3602	0.1171	0.0180*
AU17	0.0184*	0.0454*	0.2401

Note:
* Indicates $p < 0.05$.
** Indicates $p < 0.001$.

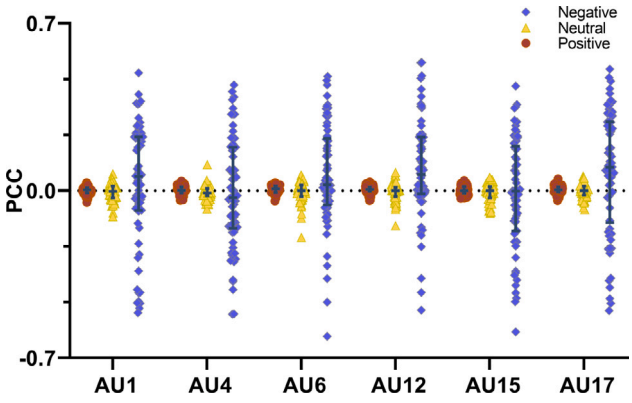


Fig. 7. Pearson correlation coefficients (PCC) between different depression-related facial AUs and changes in stimulus paradigm. Circles, triangles and diamonds indicate the correlation between changes in several depression-related facial AUs under positive, neutral and negative stimulus paradigms, respectively.

shown in Table 6, AU4 (brow lowerer) and AU12 (lip corner puller) are significantly different between normal and depressed patients, which is in line with previous studies. To further explore the relationship between facial AU intensity and scores on the PHQ9 scale, we analyzed the intensity of participants' AU4 appearance in response to negative stimuli. As shown in Fig. 5, there was a tendency for AU4 intensity to increase as PHQ9 scores increased.

4.3. Exploration among different classifiers

We also used different classifiers (SVM, KNN, and RF) to classify local face recombination features, as shown in Table 7. Between the negative and neutral stimuli, SVM had the highest classification accuracy (73.17% and 64.61% respectively). For the negative stimuli, KNN had the highest classification accuracy (64.32%). The results of the study showed that for these three classifiers, the classification accuracy under negative stimulus was the highest, and there was no difference between positive stimulus and neutral stimulus, but the classification accuracy of SVM and RF under neutral stimulus was higher than that under positive stimulus.

4.4. Fusion research among different stimulation modalities

Humans have different emotional responses when receiving different video stimuli. We believe that perhaps this difference can better identify depressed patients. Therefore, we explore the effect of fusion features (using simple feature splicing fusion) of extracted features under different stimulus materials on depression detection. As shown in Table 8, the features classification accuracy of fusion under neutral and negative stimuli is the highest (76.25%). The accuracy under positive, neutral and negative stimuli (74.71%) is also higher than that of previous single-mode features, while the fusion accuracy of the positive (negative) and positive (neutral) features does not perform too well. Fig. 6 is the confusion matrices for the SVM algorithms. Regarding the metrics of the confusion matrix, the figures reveal that the fusion of neutral and negative stimuli in multimodal fusion has higher accuracy, precision and F1-score.

Based on the above experimental results, it can be concluded that the classification effect of partially recombined facial features is much better than that of whole face features (Table 4). The facial features extracted under negative stimuli are better for distinguishing between normal people and depressed people. Although head movement and facial AU features were introduced, the improvement in accuracy was not obvious (Table 5).

To investigate the relationship between changes in the stimulus paradigm and various depression-related facial AUs, we calculate the Pearson correlation coefficients as shown in Fig. 7. The video-based emotional stimuli used in our study contained both sound and image information. To simplify the analysis, we extracted a total of four features for sound and image. For the sound information, we extracted loudness features and Mel Frequency Cepstral Coefficient (MFCC) features. For the image information, histogram features and Haralick texture features were extracted. The normalized four features were reduced to the final video-based paradigm features of the emotional stimuli using PCA. The PCC between the facial AU changes and the paradigm features was then calculated. In Fig. 7, we observed a correlation between changes in stimulus modality and facial AUs associated with depression in response to negative stimuli. However, the correlation was not significant for all subjects, which may be influenced by individual differences. Furthermore, the correlation was even weaker for positive and negative stimuli, indicating that these paradigms did not elicit noticeable changes in facial AUs.

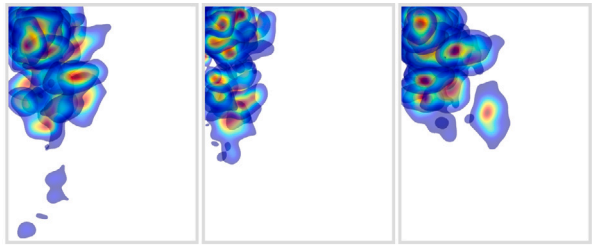
In order to explore the detection of depression by feature fusion of different stimulus modes, we conducted a fusion analysis experiment in Table 8, among which neutral (negative) feature fusion results were the highest. The results show that feature fusion in multimodal stimulus patterns can improve the recognition accuracy of depression compared to features in unimodal stimulus patterns. Compared to healthy individuals, a characteristic of clinical depression is a moderate increase in attention maintenance during dysphoric and a moderate decrease in attention maintenance during positive stimulation [65]. In [66],

Table 7
Accuracy, precision, F1-score, recall, and specificity of classification using different classifiers under different stimulus materials, where ± denotes the standard deviation of the results of multiple experiments.

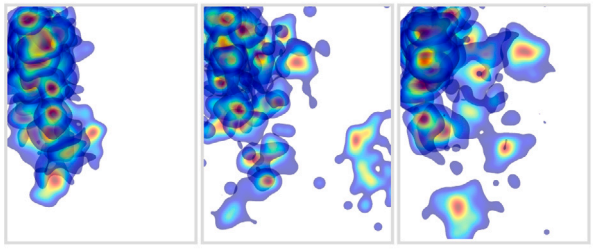
Methods	Stimulus	Accuracy	Precision	F1-score	Recall	Specificity
SVM	Positive	0.6298 ± 0.1157	0.5916 ± 0.1243	0.7013 ± 0.1250	0.8609 ± 0.1757	0.3979 ± 0.1895
	Neutral	0.6461 ± 0.1243	0.6119 ± 0.1268	0.6902 ± 0.1239	0.7915 ± 0.1791	0.5024 ± 0.2034
	Negative	**0.7317 ± 0.1017**	**0.8146 ± 0.1466**	0.6910 ± 0.1363	0.6005 ± 0.1670	**0.8628 ± 0.1376**
KNN	Positive	0.6432 ± 0.1186	0.5986 ± 0.1007	**0.7117 ± 0.0900**	**0.8773 ± 0.1235**	0.4092 ± 0.2104
	Neutral	0.6134 ± 0.1094	0.5950 ± 0.1097	0.6587 ± 0.1100	0.7376 ± 0.1843	0.4881 ± 0.1987
	Negative	0.7192 ± 0.1074	0.6022 ± 0.1492	0.6832 ± 0.1393	0.7894 ± 0.1840	0.8381 ± 0.1617
RF	Positive	0.5961 ± 0.1536	0.5769 ± 0.1641	0.6315 ± 0.1808	0.6976 ± 0.2550	0.5012 ± 0.2186
	Neutral	0.5653 ± 0.1011	0.5746 ± 0.1735	0.5767 ± 0.1562	0.5789 ± 0.2513	0.5512 ± 0.2907
	Negative	0.6692 ± 0.0970	0.6769 ± 0.0900	0.6717 ± 0.1181	0.6666 ± 0.1730	0.6619 ± 0.0944

Table 8
Accuracy, precision, F1-score, recall, and specificity of feature fusion and SVM classifier under different stimuli, where ± denotes the standard deviation of the results of multiple experiments.

Features	Accuracy	Precision	F1-score	Recall	Specificity
Positive and negative	0.7134 ± 0.1121	0.6789 ± 0.1472	0.7020 ± 0.1299	0.7267 ± 0.1749	0.7494 ± 0.1642
Neutral and negative	**0.7625 ± 0.1122**	0.6911 ± 0.1424	**0.7435 ± 0.1478**	**0.8044 ± 0.1857**	**0.8321 ± 0.1105**
Positive and neutral	0.6586 ± 0.1364	**0.7595 ± 0.1413**	0.6915 ± 0.1337	0.6347 ± 0.1766	0.5574 ± 0.1793
Positive, neutral and negative	0.7471 ± 0.1026	0.6858 ± 0.1430	0.7313 ± 0.1340	0.7833 ± 0.1962	0.8113 ± 0.1539



(a) Gaze area of all DPs



(b) Gaze area of all HCs

Fig. 8. Accumulation of gaze areas for all experimental participants with different attribute stimuli. The gaze point feature is normalized and mapped to two dimensions, with the top left corner as the origin. From left to right in sub-figure a are superimposed areas of gaze for patients with depressive disorder (DP) with positive, neutral and negative stimuli. In sub-figure b, from left to right, are the superimposed gaze areas of normal controls (HC) under positive, neutral and negative stimuli.

depressed patients exhibit slower eye movements when processing negative stimuli and tend to maintain their gaze for longer periods of time. Research has shown that people with depression may be biased towards negative emotional stimuli and find it difficult to disengage [67,68], and that depressed patients may have a reduced attentional bias towards positive stimuli. To further explore eye-movement patterns in the video stimuli, we superimposed and visualized participants' gaze points. Specifically, we normalize the gaze features extracted from Openface and map them onto a two-dimensional plane to obtain the participant's gaze patterns under different attribute stimuli, thus obtaining the participant's gaze position. As shown in Fig. 8, the range of gaze areas for different attribute stimuli was smaller in depressed patients than in healthy individuals, and was particularly pronounced for negative stimuli.

5. Conclusion and future work

In this paper, we proposed a depression detection method based on partial face recombination to analyze the accuracy of facial video data in different video stimulus states for depression detection. We found that the experimental detection results were acceptable, reflecting that this method is an effective method for recognizing depression. Our study found that under the single attribute of video stimuli, negative stimuli can effectively stimulate facial expression changes, and positive and neutral stimuli can also play a certain role in distinguishing normal people from depressed patients. We found that the results of multimodal fusion feature classification under neutral and negative stimuli were better than the optimal feature classification accuracy under single stimuli, indicating that the combination of neutral and negative stimuli is more conducive to the identification of depression. The method proposed in this paper is an objective and effective adjunct to depression detection. In addition, the method of using fusion in multi-stimulus states to improve the recognition rate may provide some research ideas for future researchers.

Our method provides a good idea for the use of facial video data to identify depression, and the following studies will be carried out in the future: (1) Due to facial regions to identify the contribution of different depression, therefore, to explore local face weighted to improve recognition accuracy. (2) Due to the limited amount of data collected at present, with the increase of data, we will use deep learning to complete the efficient fusion of different stimulus modes and further improve the accuracy of depression identification. (3) This study's sample had a 50% prevalence of depression whereas in real-world applied settings, we would expect the prevalence to be far lower. Therefore, in the future we will try to develop algorithms that can be used to screen for depression in realistic scenarios.

Declaration of competing interest

We declare that there are no conflicts of interest regarding the publication of this paper, and the manuscript is approved by all authors for publication.

Acknowledgments

This work was supported in part by the National Key Research and Development Program of China (Grant No. 2019YFA0706200), in part by the National Natural Science Foundation of China (Grant No. 62227807), in part by the Natural Science Foundation of Gansu

Province, China (Grant No. 22JR5RA488), in part by the Fundamental Research Funds for the Central Universities (Grant No. lzujbky-2023-16). Supported by Supercomputing Center of Lanzhou University.

References

[1] Z. Steel, C. Marnane, C. Iranpour, T. Chey, J.W. Jackson, V. Patel, D. Silove, The global prevalence of common mental disorders: a systematic review and meta-analysis 1980–2013, Int. J. Epidemiol. 43 (2) (2014) 476–493.

[2] G.S. Malhi, J.J. Mann, Depression, Lancet 392 (10161) (2018) 2299–2312.

[3] J. Lu, X. Xu, Y. Huang, T. Li, C. Ma, G. Xu, H. Yin, X. Xu, Y. Ma, L. Wang, et al., Prevalence of depressive disorders and treatment in China: a cross-sectional epidemiological study, Lancet Psychiatry 8 (11) (2021) 981–990.

[4] S.M. Monroe, K.L. Harkness, Why recurrent depression should be reconceptualized and redefined, Curr. Direct. Psychol. Sci. (2023) 09637214221143045.

[5] M. Marcus, M.T. Yasamy, M.v. van Ommeren, D. Chisholm, S. Saxena, Depression: A global public health concern, 2012.

[6] P.J. Batterham, H. Christensen, Longitudinal risk profiling for suicidal thoughts and behaviours in a community cohort using decision trees, J. Affect. Disorders 142 (1–3) (2012) 306–314.

[7] K. Kroenke, R.L. Spitzer, J.B. Williams, The PHQ-9: validity of a brief depression severity measure, J. Gener. Internal Med. 16 (9) (2001) 606–613.

[8] A.T. Drysdale, L. Grosenick, J. Downar, K. Dunlop, F. Mansouri, Y. Meng, R.N. Fetcho, B. Zebley, D.J. Oathes, A. Etkin, et al., Resting-state connectivity biomarkers define neurophysiological subtypes of depression, Nat. Med. 23 (1) (2017) 28–38.

[9] V. Lingiardi, L. Muzi, A. Tanzilli, N. Carone, Do therapists' subjective variables impact on psychodynamic psychotherapy outcomes? A systematic literature review, Clin. Psychol. Psychother. 25 (1) (2018) 85–101.

[10] H. Lu, S. Xu, X. Hu, E. Ngai, Y. Guo, W. Wang, B. Hu, Postgraduate student depression assessment by multimedia gait analysis, IEEE MultiMed. 29 (2) (2022) 56–65.

[11] Y. Tao, M. Yang, Y. Wu, K. Lee, A. Kline, B. Hu, Depressive semantic awareness from vlog facial and vocal streams via spatio-temporal transformer, Digit. Commun. Netw. (2023).

[12] J. Chao, S. Zheng, H. Wu, D. Wang, X. Zhang, H. Peng, B. Hu, fNIRS evidence for distinguishing patients with major depression and healthy controls, IEEE Trans. Neural Syst. Rehabil. Eng. 29 (2021) 2211–2221.

[13] M. Yang, X. Feng, R. Ma, X. Li, C. Mao, Orthogonal-moment-based attraction measurement with ocular hints in video-watching task, IEEE Trans. Comput. Soc. Syst. (2023).

[14] S. Soni, A. Seal, A. Yazidi, O. Krejcar, Graphical representation learning-based approach for automatic classification of electroencephalogram signals in depression, Comput. Biol. Med. 145 (2022) 105420.

[15] A. Ouamane, A. Benakcha, M. Belahcene, A. Taleb-Ahmed, Multimodal depth and intensity face verification approach using LBP, SLF, BSIF, and LPQ local features fusion, Pattern Recognit. Image Anal. 25 (4) (2015) 603–620.

[16] A.T. Lopes, E. De Aguiar, A.F. De Souza, T. Oliveira-Santos, Facial expression recognition with convolutional neural networks: coping with few data and the training sample order, Pattern Recognit. 61 (2017) 610–628.

[17] S.-J. Wang, H.-L. Chen, W.-J. Yan, Y.-H. Chen, X. Fu, Face recognition and micro-expression recognition based on discriminant tensor subspace analysis plus extreme learning machine, Neural Process. Lett. 39 (2014) 25–43.

[18] N. Zeng, H. Zhang, B. Song, W. Liu, Y. Li, A.M. Dobaie, Facial expression recognition via learning deep sparse autoencoders, Neurocomputing 273 (2018) 643–649.

[19] H. Dibeklioğlu, Z. Hammal, J.F. Cohn, Dynamic multimodal measurement of depression severity using deep autoencoding, IEEE J. Biomed. Health Inform. 22 (2) (2017) 525–536.

[20] L. Wen, X. Li, G. Guo, Y. Zhu, Automated depression diagnosis based on facial dynamic analysis and sparse coding, IEEE Trans. Inf. Forensics Secur. 10 (7) (2015) 1432–1441.

[21] X. Zhou, K. Jin, Y. Shang, G. Guo, Visually interpretable representation learning for depression recognition from facial images, IEEE Trans. Affect. Comput. 11 (3) (2018) 542–552.

[22] S.-T. Liong, J. See, K. Wong, R.C.-W. Phan, Less is more: Micro-expression recognition from video using apex frame, Signal Process., Image Commun. 62 (2018) 82–92.

[23] K. Patel, H. Han, A.K. Jain, Cross-database face antispoofing with robust feature representation, in: Chinese Conference on Biometric Recognition, Springer, 2016, pp. 611–619.

[24] A. Pampouchidou, O. Simantiraki, A. Fazlollahi, M. Pediaditis, D. Manousos, A. Roniotis, G. Giannakakis, F. Meriaudeau, P. Simos, K. Marias, et al., Depression assessment by fusing high and low level features from audio, video, and text, in: Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge, 2016, pp. 27–34.

[25] L. Yang, D. Jiang, W. Han, H. Sahli, DCNN and DNN based multi-modal depression recognition, in: 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), IEEE, 2017, pp. 484–489.

[26] B. Sumali, Y. Mitsukura, Y. Tazawa, T. Kishimoto, Facial landmark activity features for depression screening, in: 2019 58th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), IEEE, 2019, pp. 1376–1381.

[27] E.H. Houssein, D.A. Abdelkareem, M.M. Emam, M.A. Hameed, M. Younan, An efficient image segmentation method for skin cancer imaging using improved golden jackal optimization algorithm, Comput. Biol. Med. 149 (2022) 106075.

[28] L. Ren, D. Zhao, X. Zhao, W. Chen, L. Li, T. Wu, G. Liang, Z. Cai, S. Xu, Multi-level thresholding segmentation for pathological images: Optimal performance design of a new modified differential evolution, Comput. Biol. Med. 148 (2022) 105910.

[29] M.M. Emam, E.H. Houssein, R.M. Ghoniem, A modified reptile search algorithm for global optimization and image segmentation: Case study brain MRI images, Comput. Biol. Med. 152 (2023) 106404.

[30] L. Liu, F. Kuang, L. Li, S. Xu, Y. Liang, et al., An efficient multi-threshold image segmentation for skin cancer using boosting whale optimizer, Comput. Biol. Med. 151 (2022) 106227.

[31] M. Yang, C. Cai, B. Hu, Clustering based on eye tracking data for depression recognition, IEEE Trans. Cogn. Dev. Syst. (2022).

[32] Y.-J. Liu, J.-K. Zhang, W.-J. Yan, S.-J. Wang, G. Zhao, X. Fu, A main directional mean optical flow feature for spontaneous micro-expression recognition, IEEE Trans. Affect. Comput. 7 (4) (2015) 299–310.

[33] A. Uçar, Y. Demir, C. Güzeliş, A new facial expression recognition based on curvelet transform and online sequential extreme learning machine initialized with spherical clustering, Neural Comput. Appl. 27 (1) (2016) 131–142.

[34] J.M. Girard, J.F. Cohn, M.H. Mahoor, S. Mavadati, D.P. Rosenwald, Social risk and depression: Evidence from manual and automatic facial expression analysis, in: 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), IEEE, 2013, pp. 1–8.

[35] D.R. Rubinow, R.M. Post, Impaired recognition of affect in facial expression in depressed patients, Biol. Psychiatry 31 (9) (1992) 947–953.

[36] L. Branco, C. Cotrena, A. Ponsoni, R. Salvador-Silva, S. Vasconcellos, R. Fonseca, Identification and perceived intensity of facial expressions of emotion in bipolar disorder and major depression, Arch. Clin. Neuropsychol. 33 (4) (2018) 491–501.

[37] M.J. Weightman, M.J. Knight, B.T. Baune, A systematic review of the impact of social cognitive deficits on psychosocial functioning in major depressive disorder and opportunities for therapeutic intervention, Psychiatry Res. 274 (2019) 195–212.

[38] S.M. Persad, J. Polivy, Differences between depressed and nondepressed individuals in the recognition of and response to facial emotional cues, J. Abnorm. Psychol. 102 (3) (1993) 358.

[39] M. Tasnim, M. Ehghaghi, B. Diep, J. Novikova, Depac: a corpus for depression and anxiety detection from speech, 2023, arXiv preprint arXiv:2306.12443.

[40] L. He, M. Niu, P. Tiwari, P. Marttinen, R. Su, J. Jiang, C. Guo, H. Wang, S. Ding, Z. Wang, et al., Deep learning for depression recognition with audiovisual cues: A review, Inf. Fusion 80 (2022) 56–86.

[41] W. Guo, H. Yang, Z. Liu, Y. Xu, B. Hu, Deep neural networks for depression recognition based on 2d and 3d facial expressions under emotional stimulus tasks, Front. Neurosci. 15 (2021) 609760.

[42] P. Ekman, W. Friesen, J. Hager, Facial Action Coding System: The Manual. Research Nexus, Div, Network Information Research Corp., Salt Lake City, UT, 2002, 1 (8).

[43] J.F. Cohn, T.S. Kruez, I. Matthews, Y. Yang, M.H. Nguyen, M.T. Padilla, F. Zhou, F. De la Torre, Detecting depression from facial actions and vocal prosody, in: 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, IEEE, 2009, pp. 1–7.

[44] M. Gavrilescu, N. Vizireanu, Predicting depression, anxiety, and stress levels from videos using the facial action coding system, Sensors 19 (17) (2019) 3693.

[45] M.-H. Su, C.-H. Wu, K.-Y. Huang, T.-H. Yang, Cell-coupled long short-term memory with *L*-skip fusion mechanism for mood disorder detection through elicited audiovisual features, IEEE Trans. Neural Netw. Learn. Syst. 31 (1) (2019) 124–135.

[46] G. Stratou, S. Scherer, J. Gratch, L.-P. Morency, Automatic nonverbal behavior indicators of depression and PTSD: Exploring gender differences, in: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, IEEE, 2013, pp. 147–152.

[47] J.M. Girard, J.F. Cohn, M.H. Mahoor, S.M. Mavadati, Z. Hammal, D.P. Rosenwald, Nonverbal social withdrawal in depression: Evidence from manual and automatic analyses, Image Vis. Comput. 32 (10) (2014) 641–647.

[48] D. American Psychiatric Association, A.P. Association, et al., Diagnostic and Statistical Manual of Mental Disorders: DSM-5, Vol. 5, American Psychiatric Association, Washington, DC, 2013.

[49] D.V. Sheehan, Y. Lecrubier, K.H. Sheehan, P. Amorim, J. Janavs, E. Weiller, T. Hergueta, R. Baker, G.C. Dunbar, et al., The mini-international neuropsychiatric interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10, J. Clin. Psychiatry 59 (20) (1998) 22–33.

[50] J. Péron, S. El Tamer, D. Grandjean, E. Leray, D. Travers, D. Drapier, M. Vérin, B. Millet, Major depressive disorder skews the recognition of emotional prosody, Prog. Neuro-Psychopharmacol. Biol. Psychiatry 35 (4) (2011) 987–996.

[51] L.J. Kirmayer, et al., Cultural variations in the clinical presentation of depression and anxiety: implications for diagnosis and treatment, J. Clin. Psychiatry 62 (2001) 22–30.
[52] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, I. Patras, Deap: A database for emotion analysis; using physiological signals, IEEE Trans. Affect. Comput. 3 (1) (2011) 18–31.
[53] W.-L. Zheng, W. Liu, Y. Lu, B.-L. Lu, A. Cichocki, Emotionmeter: A multimodal framework for recognizing human emotions, IEEE Trans. Cybern. 49 (3) (2018) 1110–1122.
[54] P. Xu, Y. Huang, Y. Luo, Preliminary compilation and evaluation of Chinese emotional image library, Chin. J. Ment. Health 24 (2010) 551–554.
[55] T. Baltrušaitis, P. Robinson, L.-P. Morency, Openface: an open source facial behavior analysis toolkit, in: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2016, pp. 1–10.
[56] C. Yan, B. Gong, Y. Wei, Y. Gao, Deep multi-view enhancement hashing for image retrieval, IEEE Trans. Pattern Anal. Mach. Intell. 43 (4) (2020) 1445–1451.
[57] C. Yan, B. Shao, H. Zhao, R. Ning, Y. Zhang, F. Xu, 3D room layout estimation from a single RGB image, IEEE Trans. Multimed. 22 (11) (2020) 3014–3024.
[58] C.H. Chan, M.A. Tahir, J. Kittler, M. Pietikainen, Multiscale local phase quantization for robust component-based face recognition using kernel fusion of multiple descriptors, IEEE Trans. Pattern Anal. Mach. Intell. 35 (5) (2012) 1164–1177.
[59] V. Ojansivu, J. Heikkilä, Blur insensitive texture classification using local phase quantization, in: International Conference on Image and Signal Processing, Springer, 2008, pp. 236–243.
[60] M. Taini, G. Zhao, M. Pietikäinen, Weight-based facial expression recognition from near-infrared video sequences, in: Scandinavian Conference on Image Analysis, Springer, 2009, pp. 239–248.
[61] G. Bradski, The opencv library., Dr. Dobb's J.: Softw. Tools Prof. Program. 25 (11) (2000) 120–123.
[62] J.-D. Lee, C.-Y. Lin, C.-H. Huang, Novel features selection for gender classification, in: 2013 IEEE International Conference on Mechatronics and Automation, IEEE, 2013, pp. 785–790.
[63] A. Benavoli, G. Corani, J. Demšar, M. Zaffalon, Time for a change: a tutorial for comparing multiple classifiers through Bayesian analysis, J. Mach. Learn. Res. 18 (1) (2017) 2653–2688.
[64] X. Zhang, J. Shen, Z. ud Din, J. Liu, G. Wang, B. Hu, Multimodal depression detection: fusion of electroencephalography and paralinguistic behaviors using a novel strategy for classifier ensemble, IEEE J. Biomed. Health Inform. 23 (6) (2019) 2265–2275.
[65] T. Suslow, A. Husslack, A. Kersting, C.M. Bodenschatz, Attentional biases to emotional information in clinical depression: a systematic and meta-analytic review of eye tracking findings, J. Affect. Disord. 274 (2020) 632–642.
[66] T. Armstrong, B.O. Olatunji, Eye tracking of attention in the affective disorders: A meta-analytic review and synthesis, Clin. Psychol. Rev. 32 (8) (2012) 704–723.
[67] L. von Koch, N. Kathmann, B. Reuter, Lack of speeded disengagement from facial expressions of disgust in remitted major depressive disorder: Evidence from an eye-movement study, Behav. Res. Therapy 160 (2023) 104231.
[68] J. Joormann, M.E. Quinn, Cognitive processes and emotion regulation in depression, Depress. Anxiety 31 (4) (2014) 308–315.