J. COMBIN. MATH. COMBIN. COMPUT. 127a (2025) 7785--7802

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Research on the application of fusion of image recognition technology and sentiment analysis algorithms for psychological intervention

Zhige Lyu[1, ✉]

[1] *Guilin Normal University, Guilin, Guangxi, 541199, China*

ABSTRACT
Mental health issues have become a global concern. Aiming at the complexity of individual facial emotion expression in the task of analyzing mental health status, this study proposes a face emotion recognition method oriented to psychological intervention. The method integrates image recognition and sentiment analysis techniques, adopts Adaboost algorithm for face detection, generates an emotion region suggestion network based on face image recognition, and constructs an image sentiment classification network through feature map mapping and shared convolution. The method is then applied to the mental health recognition system. The model in this paper avoids the effects of individual and illumination differences. It has good face emotion recognition on several datasets, and the prediction accuracies are above 90%, especially for Happy emotion. In the comparison with other recognition methods, the recognition accuracy of this paper's model is improved by 12.92% to 22.95%. The experiments show that the proposed face emotion recognition method can effectively predict the emotion of facial expression data in the mental health recognition system, and promote the assessment of individual mental health status and emotion management.

Keywords: image recognition, sentiment analysis, Adaboost, face detection, neural network, psychological intervention

1. Introduction

With the rapid economic and social development and the changing environment in which people grow up, mental health problems have become more prominent. Common causes of psychological crises include acute disability or acute serious illness, broken romantic relationships, broken relationships or sudden death of relatives or friends, bankruptcy or major property or housing losses, failure in important exams, marriage and childbirth, and serious natural disasters [1-4]. At the same time, it may cause negative emotions such as stress and anxiety, escalate to depression, two-way affective disorder

✉ Corresponding author.
E-mail address: xwz20032000@163.com (Z. Lyu).
Received 20 February 2024; Revised 25 April 2024; Accepted 20 November 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-430

and other pathologies, and even self-inflicted suicide [5]. Whether a psychological crisis can be effectively intervened and dissipated is directly related to the physical and mental health of the person, the happiness and tranquility of the family, and the harmony and order of the society. Therefore, timely and effective intervention to eliminate, do a good job in psychological crisis coping strategies, for the promotion of personal physical and mental health and maintenance of social harmony and stability is of great significance [6-7]. Psychological intervention is a process of intervening and regulating the psychological problems of an individual or a group through psychological theories and methods, and preventing excessive behaviors such as self-injury, suicide, or aggressive behaviors through psychological intervention [8]. Facilitating communication, encouraging the person concerned to fully express his/her thoughts and feelings, encouraging his/her self-confidence and correct self-evaluation, and providing appropriate suggestions to facilitate problem solving [9]. Provide appropriate medical help to stabilize agitated states and assist in the treatment of diseases, improve sleep quality, and promote personal health [10-12]. Therefore, scholars have begun to seek various effective methods for psychological interventions.

With the advent of the digital age, image recognition technology has become an integral part of the computer vision field. Image recognition technology refers to the processing and analysis of images by computers in order to identify target objects or features in an image [13]. This technology plays an important role in many fields, including intelligent transportation, agricultural production, and e-commerce services [14-16]. Sentiment analysis is a method for automatically extracting and inferring the author's emotional tendencies from text by using Natural Language Processing (NLP) techniques [17]. It has important applications in many fields, such as social media analysis, brand marketing, and public opinion monitoring [18-20]. Sentiment analysis allows for a better understanding of user needs and feedback for targeted decision making and improvement. Emotions are an important part of human psychological activities, and sentiment analysis is the objective and accurate identification and analysis of emotions.

The study uses Adaboost-based face detection algorithm and facial landmark point detection based on regression tree ensemble to realize face image recognition. Then image preprocessing is carried out by image rotation adjustment, gray scale equalization and scale normalization, on the basis of which, based on the results of face detection and localization, affective regions are evaluated and candidate regions of affective regions for affective region-assisted classification are determined. Subsequently, the convolution operation of shared feature extraction for emotion region suggestion network and classification network is performed, and the classification of the original image and emotion regions is achieved by pooling the region of interest and fitting the fully connected layer. The classical expression feature extraction results under different personal characteristics and lighting conditions are analyzed to explore the effectiveness and stability of the feature extraction methods in this paper. Compare the recognition accuracy of different expression recognition methods and this paper's method for various expressions, and analyze the classification results of this paper's method for various expressions under different datasets, to evaluate the effect of the proposed method for emotion analysis of facial expressions. Finally, a mental health status recognition system is constructed, and the proposed facial emotion recognition method is used in the facial emotion recognition module, which helps in individual mental health detection and psychological intervention.

2. Overview

Traditional methods of psychological intervention are overly dependent on the experience of the psychotherapist, which is not only inefficient, but also subjective and ineffective intervention. Counseling is the main way of psychological interventions in schools, but Østergård et al [21] mentioned that the effectiveness of counseling-assisted mediated psychological interventions has a requirement on the number of times students are counseled, and the higher the recovery is likely to be under high demand. While Moreno-Peral et al [22] statistically found that psychoeducationally mediated psychological interventions had only a minor effect on preventing anxiety psychological problems.Kwok [23] integrated positive psychology with music therapy and effectively alleviated students' symptoms of anxiety while also improving subjective well-being. Music therapy counsels psychotherapy through musical activities, but the scope of treating symptoms is limited due to the fact that individuals do not perceive music in the same way, presenting individual differences.Day and Thorn [24] introduced to cognitive-behavioral therapy to alleviate the problem of psychological pain by helping individuals to identify and change adverse cognitions and establish positive behavioral habits. However, this approach is difficult to accomplish accurately, and unsuccessful changes can lead to "reinforcement" of the individual's thinking, which is not conducive to recovery. In addition, when psychological problems reach a certain level of severity, it is necessary to use medication for recovery. Depression, as a common psychological problem with complex causes, has developed a variety of therapeutic drugs on the market, but due to individual allergies, drug resistance, severity and other reasons, the drugs are not always effective. From the study of Ionescu and Papakostas [25], it is known that tricyclic antidepressants and monoamine oxidase inhibitors have no therapeutic effect on some patients, and new types of drugs are still under development. Based on this, more and more scholars have begun to explore new psychological interventions in combination with current intelligent technologies.

Lau et al [26] assessed that artificial intelligence in psychological intervention does not significantly improve psychological problems such as anxiety and ease, but mediation in addition to conventional treatment can alleviate depressive symptoms. Differently, Fulmer et al [27] integrated psychological and artificial intelligence to intervene in the treatment of anxiety with significant therapeutic effects, and the method was cost-effective. Joshi and Jadeja [28] utilized emotion computing techniques such as emotion recognition, linguistic analysis, and sentiment mining under the guidance of clinical psychology theories to identify the individual's emotions in depth, and assisted the therapist with psychological interventions. Chen et al [29] designed the emotion recognition and psychotherapy system, mainly through face recognition to obtain individual facial emotions, then the convolutional neural network to recognize emotions, feature recognition algorithms to classify the emotions, while the interaction technology in the recognition of the classified results under the results of human-computer interaction to carry out psychological interventions, to a certain extent, to avoid too much subjective participation.

In sentiment analysis, NLP technology has efficient recognition effect in emotion recognition. Ahmed et al [30] used deep entropy active learning supported by NLP and attention mechanism to label the language with emotions, and input the unlabeled linguistic text into another active learning mechanism to loop until obtaining the optimal solution, and this learning procedure can effectively help psychological interventions in the Internet. Eberhardt et al [31] used NLP as a means of sentiment analysis to understand the patient's psychology through sentiment recognition of the language communicated in a psychotherapy session, and the study also mentioned that sentiment analysis is not limited to language, and may be effective for multimodal recognition of facial expressions, voices, and so on.

3. Image recognition and emotion perception techniques

Mental health is an important issue related to individual development and social stability. In recent years, with the rapid development of society, people are facing more and more various psychological pressures, and the phenomenon of psychological crisis is becoming more and more prominent. How to effectively detect the psychological crisis of individuals and provide timely interventions has become a hot spot in current research. For this reason, this paper combines image recognition technology and sentiment analysis algorithms to propose a psychological intervention-oriented facial expression emotion recognition technology.

3.1. Face Image Recognition

Face detection and expression image preprocessing is an important part of expression recognition, is the basis of feature extraction and classification recognition, accurate face detection and reasonable expression image preprocessing can improve the recognition rate, and vice versa will reduce the recognition rate.

Face detection and localization refers to determining the location and size of all faces if they exist in the input image. The input to a face detection system is an image that may contain a face, and the output is a parameterized description of whether a face exists in the image as well as information such as the number, location, scale, and position of the face.

Adaboost based face detection algorithm is able to detect frontal faces with fast detection speed, in order to improve the real time and computing speed of this algorithm, Adaboost based face detection algorithm is used to detect faces in the captured video images and dlib c++ library is used for detecting face landmark points and obtaining the position coordinates.

3.1.1. Face Detection. In this paper we are using Adaboost based face detection algorithm in face detection.Adaboost algorithm is an improved algorithm based on Boosting.This algorithm is based on eigenvalues for face detection.The eigenvalues are obtained from the features. In this algorithm, a huge number of feature values are first obtained from haar-like features and integral map. Then some of the features with better classification are extracted to produce weak classifiers. Finally, these weak classifiers are combined with the strategy of additive modeling to form a series of strong classifiers that cascade to obtain a complex but accurate face detector.

To ensure the effectiveness of the algorithm, classifiers with simple features and simple judgment conditions should be selected. The haar-like features consist of simple rectangles, including the four ABCD, where AB is a 2-rectangle feature, C is a 3-rectangle feature, and D is a 4-rectangle feature.The key to the fast running of the Adaboost algorithm is the integrality map, which is the integral of the pixel gray values from the upper left corner of the image to the point in the map, which is the value of a location in the map.

Point A is a point in a two-dimensional matrix that stores the sum of the gray values of the entire gray area. Equation (1) is the formula for calculating the data of point A:

$$A(x,y)=\sum_{x'\le x,y'\le y} I(x',y') \tag{1}$$

where $I(x',y')$ denotes the grayscale value of point (x',y') in the original map in the interval [0, 255].

The training of the classifier is based on the haar features and the integral map described above. The mathematical expression of the weak classifier is shown in equation (2):

$$h(x,f,p,\theta)=f(x)=\begin{cases}1 & pf(x)<\theta\\ 0 & \text{other}\end{cases} \tag{2}$$

Consists of corresponding threshold information and feature information, where f refers to the type of feature, θ refers to the threshold of the classifier, p refers to the sign direction of the weak classifier inequality, and x refers to one of the detection sub-windows.

3.1.2. Facial landmark point detection. Face alignment is also known as facial landmark point detection, where facial landmark points are used to localize and represent prominent regions of the face. In this paper, we use the classical 68 facial landmark points.

In order to get real-time performance with high quality predictions, this paper uses a regression tree ensemble to estimate the landmark locations of faces directly from a sparse subset of pixel intensities. A general architecture based on gradient boosting is used for learning an ensemble of regression trees that optimizes the sum of squares of error losses and naturally handles missing or partially labeled data. As shown below this method is divided into three steps:

1) A training set of facial feature points labeled on the images. These images are manually labeled to specify specific (x, y) coordinates for the region around each facial structure.

2) Probability of distances between pairs of input pixels.

3) Based on this training data, a collection of regression trees is trained to estimate facial feature point locations directly from the pixel intensities themselves.

3.2. Image Preprocessing

Since the data used for the experiment is unconstrained to the movement of the tester in front of the camera, there will be different sizes of the images as well as certain rotational distortions and changes in posture In addition, due to the different environment in which the shots are taken lighting factors can also cause significant interference, to get the desired recognition results, the pose, size and lighting must be adjusted.

3.2.1. Rotation adjustment. The frontal image of a face is often also tilted at a certain angle. Although the angle of tilt is not large, the slight change will have a certain degree of influence, so it is necessary to correct the angle of tilt of the face to ensure the consistency of the position of the face.

According to the physiology of the face, the vast majority of faces satisfy the following assumption, i.e., the two eyes on the frontal image of a face without rotation are located on the same horizontal line. So find out the angle between the two eyes and the horizontal line, and reverse the rotation according to this angle to get the corrected image.

Let the pupil coordinates of the two eyes be (x_1, y_1) and (x_2, y_2), respectively, then the formula for θ is as follows:

$$\theta = \sin^{-1}\left(\frac{y_2 - y_1}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}\right) \tag{3}$$

Let the coordinates of any point in the image be (x, y) and the coordinates of the point after rotation be (x', y'), then:

$$\begin{bmatrix} x' & y' \end{bmatrix}^T = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} [xy]^T \tag{4}$$

3.2.2. Scale Normalization. The idea of scale normalization is to transform face images with different sizes into a unified standard size image for face feature extraction. In this paper, the bilinear interpolation method with better effect and faster speed is chosen. The specific algorithms are as follows:

1) Size reduction:

$$f'(x, y) = f\left(\frac{x \cdot l}{l'}, \frac{y \cdot h}{h'}\right) \tag{5}$$

where, f' is the gray value function of the transformed image, f is the gray value function of the original image, l and l', h and h' are the width and height of the image before and after the transformation respectively.

2) Size enlargement: In the process of size enlargement, some pixel points that are not present in the original image will appear, this is the interpolation operation that needs to be performed to calculate the pixel value of the point. In order to best eliminate the mosaic phenomenon that occurs when zooming in, the algorithm of bilinear interpolation is used, where the gray values of the vertices of the

rectangle in the original image are first copied to the corresponding vertices of the rectangle of the zoomed image. Then the gray values of all points in the original image are calculated using the bilinear interpolation algorithm to achieve size normalization. Assuming that points (x_0, y_0) and (x_1, y_1) are the two diagonal vertices of a rectangle and point (x, y) is contained in that rectangle and satisfies $x \in (x_0, x_1), y \in (y_0, y_1)$, the algorithm to find the gray value $f(x, y)$ of that point is as follows:

Step one:

$$f(x, y_0) = f(x_0, y_0) + (x - x_0)/(x_1 - x_0)[f(x_1, y_0) - f(x_0, y_0)] \tag{6}$$

Step two:

$$f(x, y_1) = f(x_0, y_1) + (x - x_0)/(x_1 - x_0)[f(x_1, y_1) - f(x_0, y_1)] \tag{7}$$

Step Three:

$$f(x, y) = f(x, y_0) + (y - y_0)/(y_1 - y_0)[f(x, y_1) - f(x, y_0)] \tag{8}$$

3.2.3. Gray scale equalization. Since changes in illumination in image acquisition tend to cause images to show different degrees of lightness and darkness, it is necessary to perform grayscale equalization on face images. Gray scale equalization, its role is to enhance the overall contrast of the face image and make the gray scale distribution uniform to eliminate the effect of lighting changes, in addition to eliminating the differences in skin color of different races.

The calculation steps are as follows:

1) List the gray levels $f_j, j = 0,1,\cdots,L-1$ of the original image, where L is the number of gray levels.

2) Count the number of pixels in each gray level $n_j, j = 0,1,\cdots,L-1$.

3) Calculate the frequency of each gray level of the histogram of the original image $P_f(f_j) = \frac{n_j}{n}, j = 0,1,\cdots,L-1$, where r is the total number of pixels of the original image.

4) Calculate the cumulative distribution function $C(f) = \sum_{j=0}^{k} P_f(f_j), j = 0,1,\cdots,k,\cdots,L-1$.

5) Calculate the gray level of the output image after mapping by applying the following formula $g_i, i = 0,1,\cdots,P-1$, P is the number of gray levels of the output image: $g_i = INT[(g_{\max} - g_{\min})C(f) + g_{\min} + 0.5]$, where INT is the rounding symbol.

6) Count the number of pixels in each gray level after mapping $n_i, i = 0,1,\cdots,P-1$.

7) Calculate the output image histogram $P_g(g_i) = \frac{n_i}{n}, i = 0,1,\cdots,P-1$.

8) Modify the gray levels of the original image using the mapping relations f_i and g_i to obtain an output image whose histogram is approximately uniformly distributed.

3.3. Image emotion perception technology

Firstly, a neural network-based emotion region generation method is proposed, followed by an improvement for the utilization of emotion regions for image emotion-assisted classification, and an end-to-end deep learning-based image emotion classification network, SentiR-CNN, is realized.

3.3.1. Emotional Region Generation. The emotion region is a piece of subgraphs in the original image, which is more concentrated compared to the original image, and is a concentrated representation of the emotion information in the image. If the emotion region can be used to assist emotion classification of the recognized original image, the accuracy of image emotion classification can be greatly improved. In this paper, neural networks are utilized to replace the traditional target detection method to generate candidates for emotional regions, and the Sentiment Region Proposal Network (Senti-RPN) is proposed.

Senti-RPN contains three parts in a broad sense: a neural network that generates candidate frames for emotional regions, a neural network module that evaluates the emotional information of the candidate frames, and a computational module that synthesizes and evaluates the comprehensive emotional information of the candidate frames.

Redundant bounding box regions are generated using a reduced confidence threshold for face image recognition output and are used as candidate boxes for emotion regions. The object confidence of the face image recognition output is used as the object score $Score\_Obj_p$ of the emotion region, which represents the confidence of the containing object of its p nd emotion candidate region for an original image, and $Score\_Senti_p$ is used as the measure of its containing emotion information. For the acquisition of $Score\_Senti_p$, the VGG16 classification model is used, and Softmax is used as the last output layer of VGG16. For each emotion region, after the feature extraction convolutional layer and the fully connected layer, its output is the probability corresponding to the current emotion region in the emotion category.

The training process is to maximize the probability of the correct category as much as possible through the backpropagation algorithm, defined d_i as the output of the penultimate layer of the VGG16 model. Its Softmax loss function is:

$$Loss(W) = \sum_{i=1}^{N} \sum_{j \in I} F(l_i = j) \log p(l_i = j \mid d_i, w_j) \tag{9}$$

where $W = \{w_j\}_{j \in I}$ is the set of parameters of the model. When the prediction function $F(x)$ when x is true, $F(x) = 1$. And vice versa, $F(x) = 0$. The probability $p(l_i = j \mid d_i, w_j)$ for each sentiment classification can be defined by the Softmax function:

$$p(l_i = j \mid d_i, w_j) = \frac{e^{w_j^T d_i}}{\sum_{x \in I} e^{w_x^T d_i}} \tag{10}$$

Using the model to calculate Senti-Anchor's classification probability distribution for emotion s_{pq}, then $Score\_Senti_p$ is calculated by the following equation:

$$Score\_Senti_p = \sum_{q=1}^{N} s_{pq} * \log s_{pq} + 1 \tag{11}$$

Up to this point, an affective region was measured by an object score and an affective score. Combining the two factors provides a measure of the final quality of the emotional region:

$$Score_p = \sqrt{Score\_Obj_p^2 + \alpha * Score\_Senti_p^2} \tag{12}$$

where α is the weighting coefficient of the object score and emotion score of the affective region, in this study α was taken as a value of 0.9. The final measure of the quality of the affective region is based on the final $Score_p$.

3.3.2. Auxiliary classification of emotional regions:

1) Aided classification of emotional regions based on feature map mapping. Using the convolutional feature extraction layer to extract features uniformly for the original map and the emotional regions on it, and using the fully connected layer of the classification network to classify the batch of emotional regions and the original map to realize the demand of obtaining the classification results of all the

emotional regions and the original map through only one inference calculation under the condition of ensuring that each emotional region is not affected.

Based on the translation invariance characteristics of convolutional neural networks, the mapping of emotion regions in the feature map is performed by using equal scale coordinate transformation. Suppose an emotion region in the original map represents set $\{c_x, c_y, w, h\}$, where $\{c_x, c_y\}$ is the center coordinate of the emotion region, and w and h are the width and height of the emotion region. The resolution of the original map is $M \times N$, and the corresponding resolution of the feature map is $m \times n$, then the subset of emotional region features on the feature map is:

$$\left\{c_x^{'}, c_y^{'}, w^{'}, h^{'}\right\} = \left\{\frac{m}{M}c_x, \frac{m}{M}c_y, \frac{m}{M}w, \frac{m}{M}h\right\} \tag{13}$$

Because the size of each emotion region is different, the feature maps after feature extraction by convolution are bound to be of different sizes, but the fully connected layer must have a fixed input scale, so after the convolutional layer and before the fully connected layer, the region of interest pooling layer in Fast-CNN is introduced. The feature maps extracted from the last convolutional layer are partitioned into S × S grids on the feature maps corresponding to each sentiment region feature subgraph. Pooling operations are performed for each network.

After the fully connected layer and Softmax classification, for each original image, a sentiment classification result $\vec{R}_{global}$ for the original image and a result $\vec{R}_{score_p}$ for each sentiment region are obtained. Final sentiment classification result $\vec{R}$:

$$\overline{R} = (1-\beta) \times \overline{R}_{global} + \beta \times \frac{1}{K}\sum_{p=1}^{K}\overline{R}_{score_p} \tag{14}$$

Where K is the number of emotional regions, and β is the weight parameter to adjust the fusion ratio between the emotional region results and the original map prediction results.

2) Network merging and feature sharing. Figure 1 shows the network structure of SentiR-CNN. The classification probability calculation for each emotion region in the emotion region suggestion network and the emotion classification network using feature map mapping in the emotion classification network are merged into one, and both of them share the VGG16 network for feature extraction and calculating the probability distribution of each emotion region. In this way, the two networks can share the same features, avoiding repeated computations. Meanwhile, the arbitrary input feature of the emotion classification network can also be applied to the emotion region suggestion network, avoiding the problem of input deformation affecting the classification accuracy due to the different sizes and dimensions of emotion regions.

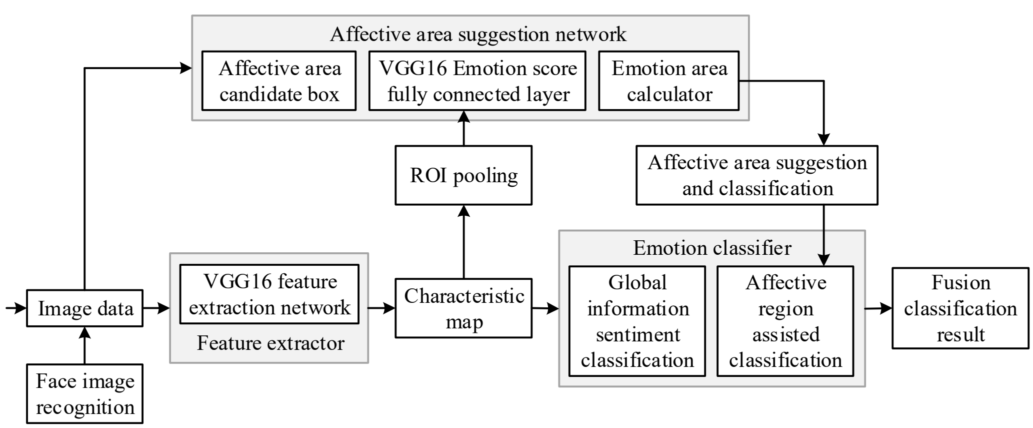

Fig. 1. Network structure of SentiR-CNN

4. Experimental results and analysis

4.1. Expression feature extraction

4.1.1. Experimental design. In order to verify the feasibility of extracting expression features using convolutional neural network, the convolutional feature extraction layer is used in the experiment to extract the expression features in the grayscale image, and the expression features corresponding to the six basic emotions, namely, happy, angry, sad, surprised, disgusted, and afraid, are extracted respectively. In order to analyze the relationship between the convolutional feature extraction layer and different expression states, and to study the connection between the expression features and factors such as lighting conditions and personal characteristics, the expression-related facial expression features are extracted in different lighting environments and under different testers' conditions, and these features are compared and analyzed intuitively.

Firstly, several typical expression images are selected to teach convolutional feature extraction, and the extraction results are used as the expression feature templates of the type, then the expression images of different testers under different lighting conditions are randomly selected as the test images, and the tested images are subjected to convolutional feature extraction, and finally, the tested expression features are compared with the feature templates of the six typical expressions, and the Euclidean distance is used as a similarity metric of the two feature vectors. The Euclidean distance can be calculated using the difference between the expression feature template and the expression feature to be recognized.

4.1.2. Experimental results:

1) Comparison of expression features when the same person makes different expressions under the same lighting conditions. Under the condition of constant illumination, the left eye feature vectors corresponding to different expressions were extracted for the same person, and the expression features of the same person under constant illumination conditions are shown in Fig. 2, in which the six feature curves correspond to the same person's six expressions of happiness, anger, sadness, surprise, disgust, and fear, respectively. The Euclidean distances of the six expression features range from 0.08 to 21.81. Even under the same lighting and the same tester conditions, the differences of the expression features corresponding to different expression images are very obvious, and the Euclidean distances between them are large.

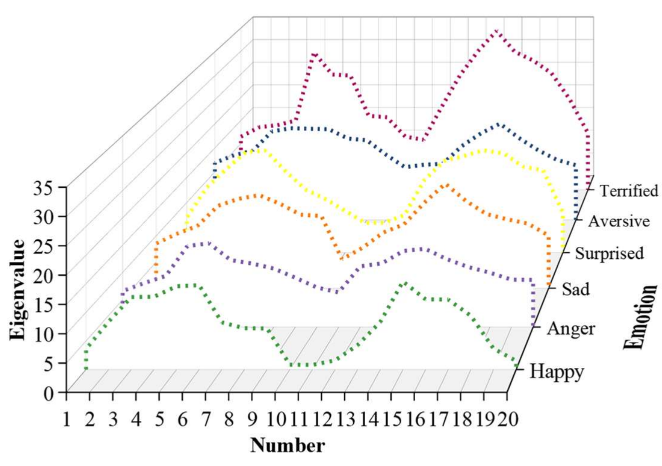

Fig. 2. The expression of the same person in the same light condition

2) Comparison of expression features when different people make the same expression under different lighting conditions. The left eye feature vectors were extracted from different testers when they made

the same expression in different lighting environments with different degrees of brightness and darkness, and the expression features of different people in different lighting conditions are shown in Fig. 3, in which the six curves represent the left eye feature vectors of the six testers when they made the happy expression in two different lighting conditions. From the figure, it can be seen that the expression features corresponding to different people making the same expression have a similar trend of change, and the Euclidean distance between their features is small, within 3.66, and is not significantly affected by changes in illumination.

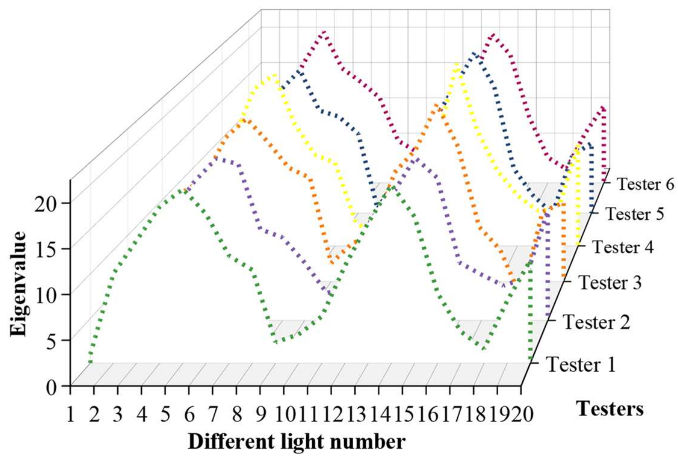

Fig. 3. The expression of different persons in the different light condition

The experiment obtains the average Euclidean distance comparison results of the expression features of 12 testers by extracting and comparing the expression features of more testers and further analyzing and comparing the feature vectors of several other basic expressions. A typical expression image is extracted for one kind of expression of each tester, then a total of 72 images are obtained, and these 72 images are used as expression templates and tested expressions at the same time, and the Euclidean distances between them are calculated in two by two, and finally the average value of Euclidean distances between the two kinds of expressions is taken as the test result. The average Euclidean distances of the six basic expression features are shown in Fig. 4, the Euclidean distances between different expression features are larger, ranging from 32 to 49, while the Euclidean distances of the same expressions from different testers are smaller, all of them are less than 15, and there is a good similarity. Therefore, it can be concluded that the convolutional neural network can effectively extract features related to expression changes, which are insensitive to light changes and can shield the influence of individual feature differences to achieve person-independent expression feature extraction.

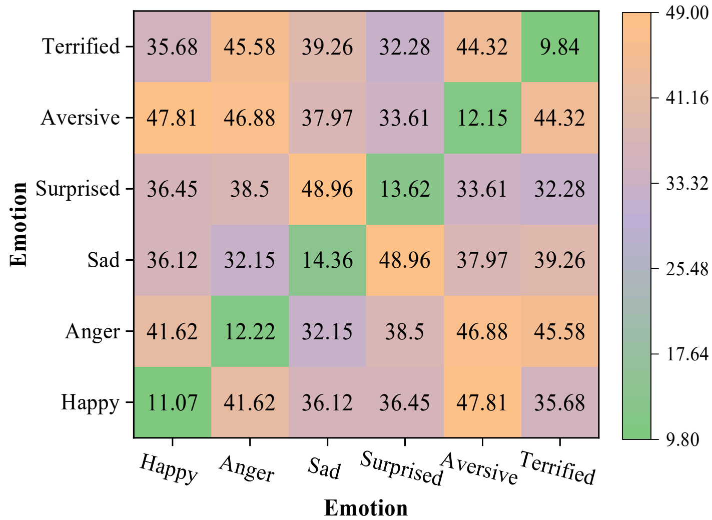

Fig. 4. The average euclosic distance of the six basic emojis

4.2. Expression Emotion Recognition

4.2.1. Experimental data. The JAFFE face database and CK+ database from Kyushu University, Japan were used for the experiments.The JAFFE database divides the training data and test data 1:1.The CK+ data experiments used static image data. Another 20 students were selected to take 4 sets of experimental data, while 1260 were collected on the web, 180 emojis for each category, collecting data totaling 1340 images. The experimental environment is Win10+visual studio 2012+OpenCV 2.4.9.

4.2.2. Comparative Experimental Analysis. The experimental results of the SentiR-CNN model, OVR-SVM, OVO-SVM, SVM classification based on LBP features, and decision tree based classification methods in this paper are compared. The results obtained from the five different classification methods are shown in Figure 5, and Table 1 shows the specific sentiment recognition accuracy of the different methods. From the experimental results, it is concluded that the average recognition accuracy of this paper using the SentiR-CNN method for the seven categories of basic expressions is 90.78%, and compared with the other methods, the method proposed in this paper improves the average experimental accuracy by 12.92-22.95 percentage points, especially in the categories of Happy, Anger, Fear, and Surprise, where the experimental accuracies are all greater than 90%. Compared with the results of other methods for emotion classification and recognition, the SentiR-CNN model experimental accuracies are improved overall.

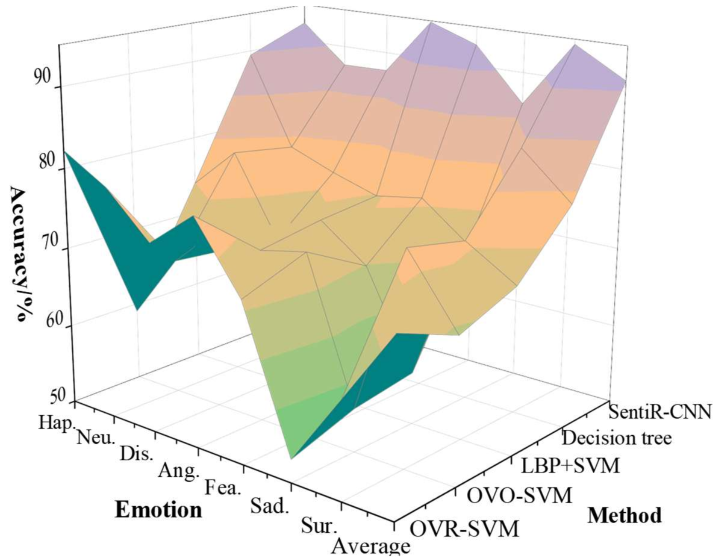

Fig. 5. The results of five different classification methods(JAFFE dataset)

Table 1. Specific emotional recognition accuracy of different methods

Method	Hap.	Neu.	Dis.	Ang.	Fea.	Sad.	Sur.	Average
OVR-SVM	82.25	78.96	73.53	78.15	69.87	52.83	61.94	71.08
OVO-SVM	59.02	67.2	72.6	71.56	72.81	55.37	76.23	67.83
LBP+SVM	73.87	79.13	69.05	72.91	68.51	56.54	74.52	70.65
Decision tree	89.43	77.92	75.64	73.76	74.77	70.09	83.44	77.86
SentiR-CNN	92.47	87.48	87.59	94.82	92.67	86.04	94.37	90.78

4.2.3. Emotion recognition results. The CK+ dataset is trained using the SentiR-CNN model, and in order to further analyze the experimental accuracies, the experimental accuracies of each class of expressions are analyzed, and the distribution of the expression emotion recognition accuracies under the CK+ dataset is shown in Figure 6. Hap. has the highest recognition accuracy of 96.76% while Sad. has a lower recognition accuracy of 89.63% in comparison to other categories recognition accuracy in the experimental results recognition accuracy.SentiR-CNN model trained on CK+ dataset has a higher average sentiment recognition accuracy of 93.64%.

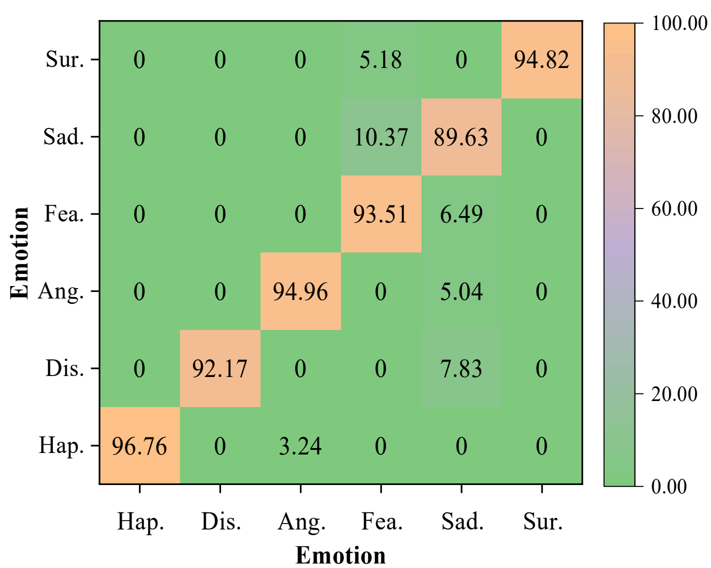

Fig. 6. The expression emotion recognition accuracy distribution under CK+ data set

Experiments are also conducted on the homemade dataset, and the distribution of expression emotion recognition accuracy under the homemade dataset is shown in Fig. 7. In the homemade dataset, the emotion recognition accuracies of the SentiR-CNN model for Hap., Dis., Ang., Fea., Sad. and Sur. are 95.14%, 93.21%, 93.97%, 94.15%, 87.77%, and 93.13%, respectively, with an average recognition accuracy of 92.90%, indicating that the SentiR-CNN model has high accuracy for face emotion recognition.

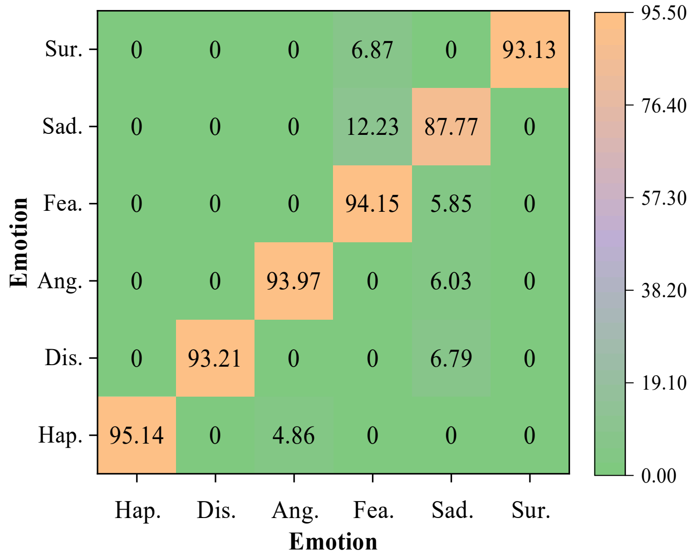

Fig. 7. The expression emotion recognition accuracy distribution under self-restraint data set

4.2.4. Comparison of Dimension Reduction Methods. In this paper, the SentiR-CNN model reduces the repetitive operation of feature extraction and realizes the dimensionality reduction of feature picking by means of feature map mapping and convolutional feature sharing. Three methods of PCA dimensionality reduction, sparse method and LPP dimensionality reduction are selected for comparison, the comparison results of dimensionality reduction methods are shown in Fig. 8, and the recognition accuracies of different dimensionality reduction methods are shown in Table 2. This paper's method is significantly better than PCA dimensionality reduction, sparse method, LPP dimensionality reduction methods, the average accuracy of all types of emotions reaches 92.51%, in addition to sadness emotion, the recognition accuracy of all other types of emotions are more than 92%. The classifier is trained by feature extraction on the samples, and if the features are not filtered, it takes 74.52 seconds to train the classifier. And after using PCA dimensionality reduction, sparse method and LPP dimensionality reduction, it takes 36.25, 33.64 and 31.71 seconds to extract features to train the classifier. It takes only 27.15 seconds to train the classifier using this method for feature extraction. Therefore, the feature map mapping and convolutional feature sharing in this paper can reduce the algorithm computation and improve the model operation efficiency.

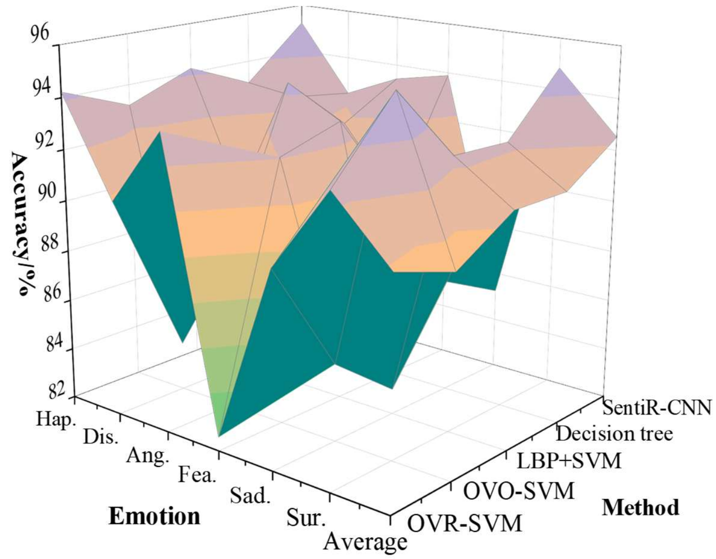

Fig. 8. Contrast results of the method of descending dimension

Table 2. Identification accuracy of different vis-reduction methods

Method	Hap.	Dis.	Ang.	Fea.	Sad.	Sur.	Average
Nondescending dimension	94.22	90.45	93.5	82.53	89.51	92.73	90.49
PCA	93.18	83.91	87.94	92.35	85.02	95.53	89.66
Sparse method	94.17	89.36	94.25	93.12	82.99	92.63	91.09
LPP	93.08	92.83	92.08	89.42	86.48	92.52	91.07
Our method	95.23	92.46	93.42	93.89	85.24	94.84	92.51

5. Mental Health Status Recognition System

The facial expression emotion recognition method oriented to psychological intervention is put into practical application to design the mental health state recognition system. The framework of the mental health state recognition system is shown in Fig. 9, which adopts the idea of layered design, dividing the system into modules such as data acquisition module, data preprocessing module, emotion analysis module, facial emotion recognition module, behavior analysis module, result output module and user-assisted suggestion module, etc., and each module provides different functions, and transfers and interacts with each other's data through interfaces.

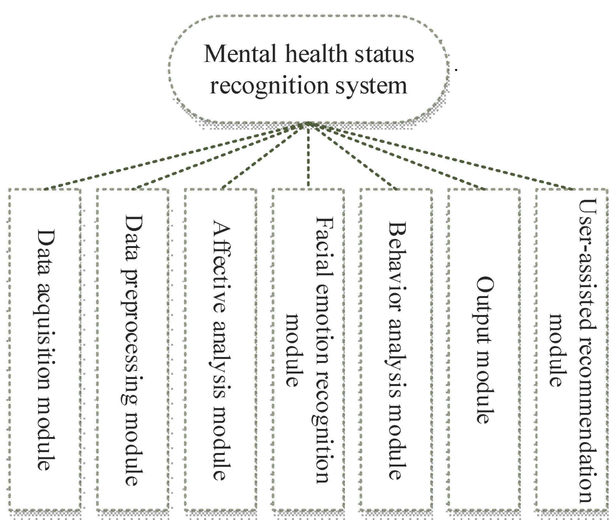

Fig. 9. Framework of mental health status recognition system

5.1. Data acquisition module

The data acquisition module is responsible for collecting the user's mental health data from different data sources, including speech data, text data, facial expression data and behavioral data. Speech data collection module collects user's voice information through high quality microphone to record features such as tone, speed and volume. The text data acquisition module collects the user's written expression through the text input box to record information such as emotion and way of thinking. The facial expression data acquisition module utilizes a high-resolution camera to capture images of the user's facial expression and extract facial features and emotional information. The behavioral data acquisition module, on the other hand, records the user's behavior and action information through motion sensors and other devices to reflect the user's behavioral patterns and emotional states. By integrating these multimodal data, the system is able to comprehensively and accurately assess the user's mental health status.

5.2. Data preprocessing module

The data preprocessing module cleans and preprocesses the collected raw data, including denoising, feature extraction and data classification and organization. In the application of the facial expression emotion recognition method, the data preprocessing is mainly to carry out face detection and facial ground point detection on the collected facial expression images, and to carry out rotational adjustment, scale normalization and grayscale equalization to generate emotion regions. At the same time, these feature data are classified and organized to provide high-quality input data for subsequent emotional and behavioral analysis.

5.3. Sentiment analysis module

The sentiment analysis module uses a variety of techniques and algorithms, including speech sentiment analysis, text sentiment analysis, facial expression sentiment analysis, and multimodal sentiment analysis. Specifically, the data is sentiment analyzed through deep learning models to identify the user's emotional state so that the system can comprehensively and accurately assess the user's emotional state.

5.4. Facial Emotion Recognition Module

The facial emotion recognition module is based on the generated emotion region candidate area, and trains the neural network to calculate the relevant emotion classification probability distribution for the emotion candidate area, as well as the comprehensive evaluation score of the emotion region. Based on the comprehensive evaluation score of the emotion region, a certain number of emotion region classification results are selected to be fused with the emotion classification results of the original image for decision-making, so as to realize the recognition and analysis of the user's emotional state.

5.5. Behavioral analysis module

The behavioral analysis module analyzes and identifies the user's behavioral data, which mainly includes two steps: behavioral feature extraction and behavioral pattern recognition. First, features such as movement speed, posture change, and activity range are extracted from the behavioral data to reflect the user's behavioral habits and emotional state. Then, machine learning algorithms or pattern recognition techniques are used to classify and identify these behavioral features, analyze the user's behavioral patterns and changes, and identify abnormal behaviors or emotional features. Through these steps, the system can more accurately understand and assess the user's behavior and emotional state.

5.6. *Results Output Module*

The result output module is responsible for organizing and presenting the analysis results, including both report generation and result presentation. First, it automatically generates a user mental health report based on the analysis results, covering the results of emotion analysis, facial expression recognition and behavior analysis, providing reference for users and professionals. Then, the analysis results are displayed to the user in the form of charts and statistics, which visually presents the user's mental health status and trends and helps the user to understand his/her own situation.

5.7. *User-assisted advice module*

The user-assisted advice module generates personalized mental health advice based on the results of the analysis, including psychological counseling and adjustment advice and community interaction and resource recommendations. Firstly, it provides psychological counseling services and psychological adjustment suggestions to guide users to manage their emotions and improve their mental health. Second, it recommends relevant mental health resources and community interaction activities to promote users' participation in community activities and resource sharing, so as to help users manage and improve their mental health in a more comprehensive way.

6. Conclusion

Based on the recognition and intervention of mental health status, this paper uses Adaboost face detection algorithm and image sentiment analysis network based on deep learning to propose a face sentiment recognition method oriented to psychological intervention. The method can be used in the mental health status recognition system to assist the user's mental health status recognition.

1) The convolutional layer feature extraction method used in the constructed SentiR-CNN model can effectively extract the feature vectors related to expression changes, and the average Euclidean distance of the same expression feature of different testers is less than 15, which effectively shields the influence of lighting changes and individual feature differences.

2) The average recognition accuracies of the SentiR-CNN model in the JAFFE, CK+ and homemade datasets are 90.78%, 93.64% and 92.90% respectively, which have superior face emotion recognition performance, and the recognition accuracies in the JAFFE dataset are improved by 12.92%-22.95% compared with the other comparative methods, among which the recognition accuracies of Happy emotion are the recognition accuracy is the highest.

3) The mental health state recognition system contains seven modules, in which the proposed face emotion recognition method can play the role of facial emotion analysis to help the system understand and evaluate user emotions more accurately. The system can not only accurately identify emotions, but also provide personalized mental health advice to help users better manage their emotions and mental health.

Funding

This work was supported by Key Research Project of Guilin Normal College in 2023 (Project No. KYA202307): "Construction of a Mental Health Monitoring Framework System for Primary Schools".

References

[1] Papadakaki, M., Ferraro, O. E., Orsi, C., Otte, D., Tzamalouka, G., Von-der-Geest, M., ... & Chliaoutakis, J. (2017). Psychological distress and physical disability in patients sustaining severe injuries in road traffic crashes: Results from a one-year cohort study from three European countries. Injury, 48(2), 297-306.

[2] Harisma, B. F. (2022). KONSELING SEBAYA: SEBUAH PENANGANAN MASALAH PSIKOLOGIS REMAJA PUTUS CINTA [PEER COUNSELING: A HANDLING OF POST-BREAKUP ADOLESCENT PSYCHOLOGICAL PROBLEM]. Al-Ihtiram: Multidisciplinary Journal of Counseling and Social Research, 1(2), 81-92.

[3] Kennedy, C., Deane, F. P., & Chan, A. Y. (2019). In limbo: A systematic review of psychological responses and coping among people with a missing loved one. Journal of clinical psychology, 75(9), 1544-1571.

[4] Khan, M. A., Zummer, A., & Malik, A. (2024). The Psychological Impact of Debt: Stress and its Connection to Mental Health Disorders. Review of Education, Administration & Law, 7(4), 319-333.

[5] Al-Amryeen, M. H., Baioumy, N. A., & Salieh, A. W. B. (2020). The Psychological Problems Faced by Students during General Secondary Education Certificate Examinations in Jordan. Journal of Academic Research in Psychology, 7(1), 1-16.

[6] Sharma, A., Shrivastava, D., & Sharma IV, A. (2022). Psychological problems related to infertility. Cureus, 14(10).

[7] Beaglehole, B., Mulder, R. T., Frampton, C. M., Boden, J. M., Newton-Howes, G., & Bell, C. J. (2018). Psychological distress and psychiatric disorder after natural disasters: systematic review and meta-analysis. The British Journal of Psychiatry, 213(6), 716-722.

[8] O'Connor, R. C., Ferguson, E., Scott, F., Smyth, R., McDaid, D., Park, A. L., ... & Armitage, C. J. (2017). A brief psychological intervention to reduce repetition of self-harm in patients admitted to hospital following a suicide attempt: a randomised controlled trial. The Lancet Psychiatry, 4(6), 451-460.

[9] Abu Omar, D., Kirkman, A., Scott, C., Babicova, I., & Irons, Y. (2024). Positive psychology interventions to increase self-esteem, self-efficacy, and confidence and decrease anxiety among students with dyslexia: A narrative review. Youth, 4(2), 835-853.

[10] Inchausti, F., MacBeth, A., Hasson-Ohayon, I., & Dimaggio, G. (2020). Psychological intervention and COVID-19: what we know so far and what we can do. Journal of contemporary psychotherapy, 50(4), 243-250.

[11] Celano, C. M., Albanese, A. M., Millstein, R. A., Mastromauro, C. A., Chung, W. J., Campbell, K. A., ... & Huffman, J. C. (2018). Optimizing a positive psychology intervention to promote health behaviors after an acute coronary syndrome: the Positive Emotions After Acute Coronary Events III (PEACE-III) randomized factorial trial. Biopsychosocial Science and Medicine, 80(6), 526-534.

[12] Saruhanjan, K., Zarski, A. C., Bauer, T., Baumeister, H., Cuijpers, P., Spiegelhalder, K., ... & Ebert, D. D. (2021). Psychological interventions to improve sleep in college students: A meta-analysis of randomized controlled trials. Journal of sleep research, 30(1), e13097.

[13] Wang, J., Li, X., Jin, Y., Zhong, Y., Zhang, K., & Zhou, C. (2024, June). Research on image recognition technology based on multimodal deep learning. In 2024 IEEE 2nd International Conference on Image Processing and Computer Applications (ICIPCA) (pp. 1363-1367). IEEE.

[14] Chen, Z., Su, Y., Liu, Y., Huang, J., & Cao, W. (2020). Key technologies of intelligent transportation based on image recognition. International Journal of Advanced Robotic Systems, 17(3), 1729881420917277.

[15] Yuan, Y., Chen, L., Wu, H., & Li, L. (2022). Advanced agricultural disease image recognition technologies: A review. Information Processing in Agriculture, 9(1), 48-59.

[16] Rui, C. (2020). Research on classification of cross-border E-commerce products based on image recognition and deep learning. IEEE Access, 9, 108083-108090.

[17] Stine, R. A. (2019). Sentiment analysis. Annual review of statistics and its application, 6(1), 287-308.

[18] Drus, Z., & Khalid, H. (2019). Sentiment analysis in social media and its application: Systematic literature review. Procedia Computer Science, 161, 707-714.

[19] Rambocas, M., & Pacheco, B. G. (2018). Online sentiment analysis in marketing research: a review. Journal of Research in Interactive Marketing, 12(2), 146-163.

[20] El Barachi, M., AlKhatib, M., Mathew, S., & Oroumchian, F. (2021). A novel sentiment analysis framework for monitoring the evolving public opinion in real-time: Case study on climate change. Journal of Cleaner Production, 312, 127820.

[21] Østergård, O. K., Fenger, M., & Hougaard, E. (2019). Symptomatic distress and effectiveness of psychological treatments delivered at a nationwide student counseling service. Counselling Psychology Quarterly, 32(2), 150-168.

[22] Moreno-Peral, P., Conejo-Ceron, S., Rubio-Valera, M., Fernandez, A., Navas-Campaña, D., Rodriguez-Morejon, A., ... & Bellón, J. Á. (2017). Effectiveness of psychological and/or educational interventions in the prevention of anxiety: a systematic review, meta-analysis, and meta-regression. JAMA psychiatry, 74(10), 1021-1029.

[23] Kwok, S. Y. (2019). Integrating positive psychology and elements of music therapy to alleviate adolescent anxiety. Research on Social Work Practice, 29(6), 663-676.

[24] Day, M. A., & Thorn, B. E. (2022). Psychological interventions: a focus on cognitive-behavioral therapy. Clinical Pain Management: A Practical Guide, 272-281.

[25] Ionescu, D. F., & Papakostas, G. I. (2017). Experimental medication treatment approaches for depression. Translational psychiatry, 7(3), e1068-e1068.

[26] Lau, Y., Ang, W. H. D., Ang, W. W., Pang, P. C. I., Wong, S. H., & Chan, K. S. (2025). Artificial Intelligence–Based Psychotherapeutic Intervention on Psychological Outcomes: A Meta-Analysis and Meta-Regression. Depression and Anxiety, 2025(1), 8930012.

[27] Fulmer, R., Joerin, A., Gentile, B., Lakerink, L., & Rauws, M. (2018). Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. JMIR mental health, 5(4), e9782.

[28] Joshi, R., & Jadeja, M. (2024). The Synergy of Clinical Psychology and Affective Computing: Advancements in Emotion Recognition and Therapy. In Affective Computing for Social Good: Enhancing Well-being, Empathy, and Equity (pp. 21-45). Cham: Springer Nature Switzerland.

[29] Chen, M., Liang, X., & Xu, Y. (2022). Construction and analysis of emotion recognition and psychotherapy system of college students under convolutional neural network and interactive technology. Computational Intelligence and Neuroscience, 2022(1), 5993839.

[30] Ahmed, U., Mukhiya, S. K., Srivastava, G., Lamo, Y., & Lin, J. C. W. (2021). Attention-based deep entropy active learning using lexical algorithm for mental health treatment. Frontiers in Psychology, 12, 642347.

[31] Eberhardt, S. T., Schaffrath, J., Moggia, D., Schwartz, B., Jaehde, M., Rubel, J. A., ... & Lutz, W. (2024). Decoding emotions: Exploring the validity of sentiment analysis in psychotherapy. Psychotherapy Research, 1-16.