IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 41, No. 6, December, 2024, pp. 2923-2933
Journal homepage: http://iieta.org/journals/ts

Application of Deep Learning-Based Image Processing in Emotion Recognition and Psychological Therapy

Yang Liu[1,2], Yawen Zhang[1], Yuan Wang[1,3*]

[1] School of Educational Science, Xinjiang Normal University, Urumuqi 830017, China
[2] Xinjiang Key Laboratory of Mental Development and Learning Science, Urumuqi 830017, China
[3] Center for Xinjiang Higher Education Development Studies, Xinjiang Normal University, Urumuqi 830017, China

Corresponding Author Email: wangyuan@xjnu.edu.cn

https://doi.org/10.18280/ts.410612

Received: 18 July 2024
Revised: 3 November 2024
Accepted: 25 November 2024
Available online: 31 December 2024

Keywords:
deep learning, image processing, emotion recognition, psychological therapy, multi-feature fusion, personalized analysis

ABSTRACT

With the rapid development of artificial intelligence technologies, particularly deep learning, the application of image processing in emotion recognition and psychological therapy has become a growing area of research. As a crucial indicator of an individual's psychological state, accurate emotion recognition plays a vital role in psychological treatment and mental health management. Traditional emotion recognition methods primarily rely on subjective judgment by human experts, which has certain limitations. In contrast, deep learning-based automated emotion recognition methods can capture emotional changes in real-time and with high accuracy through facial expressions, eye movement trajectories, and other image data, overcoming the shortcomings of traditional methods. Currently, emotion recognition technology is widely applied in fields such as psychological therapy, affective computing, and smart healthcare. However, existing research still faces challenges, including insufficient recognition accuracy, poor adaptability to individual differences, and weak integration with actual psychological therapy practices. In response to these issues, this paper proposes a deep learning-based image processing method that integrates multi-feature fusion techniques to improve the accuracy of emotion recognition. The method is applied to the detection of abnormal emotional states in psychological therapy and personalized emotion analysis. The results show that deep learning technology can effectively recognize complex emotional changes and provide more accurate emotional intervention strategies for psychological therapy, offering significant theoretical and practical value.

1. INTRODUCTION

With the rapid development of information technology, artificial intelligence, especially deep learning technology, is being increasingly applied in various fields. Image processing, as an important application direction of deep learning, has been widely used in fields such as healthcare, security, autonomous driving, etc. [1-5]. In particular, in emotion recognition and psychological therapy, image processing technology can assist in mental health management and emotion regulation by analyzing multi-dimensional data such as facial expressions and eye movement trajectories [6, 7]. Emotion recognition not only contributes to the in-depth development of psychological research but also provides effective technical support for psychological therapy, promoting the intelligent and personalized development of the mental health field.

The application of emotion recognition in psychological therapy is of great significance. Emotion is an important component of human psychological activities, and its changes directly affect behavior and psychological state [8-10]. Traditional psychological therapy methods mainly rely on face-to-face communication and interviews, but these methods are limited by factors such as the treatment environment, the therapist's experience, and the patient's subjective expression, often leading to certain limitations [11-18]. Deep learning-based image processing technology can capture subtle changes in human emotions in real-time and with high accuracy, breaking through these limitations and providing a more objective and efficient method for emotion recognition. Emotion recognition technology not only improves the accuracy of psychological therapy but also, with the support of big data analysis, promotes the formulation of personalized treatment plans, providing more effective psychological interventions for patients.

Although existing emotion recognition methods have made significant progress in many fields, there are still some problems and challenges. First, existing methods often rely on single emotion recognition features (such as facial expressions or speech signals), which are less accurate and robust in complex situations [19-24]. Second, the integration of emotion recognition and psychological therapy is not yet close. Most existing research is still at the stage of emotion recognition and has not fully considered its application in actual psychological therapy. Third, emotional changes are influenced by factors such as individual differences and cultural backgrounds.

Existing models generally lack personalized adaptability and are difficult to provide consistent effectiveness across different patients. Therefore, improving the accuracy of emotion recognition and effectively integrating it with psychological therapy remain key challenges in current research.

The research in this paper mainly focuses on the application of deep learning-based image processing in emotion recognition and psychological therapy. The research content includes two aspects: first, using deep learning technology for the recognition of abnormal human emotions to better grasp the emotional fluctuations of patients during psychological therapy; second, using a multi-feature fusion emotion analysis method to improve the accuracy and robustness of emotion recognition, and combining psychological therapy needs to achieve personalized emotional interventions. These studies will provide new ideas for the deep integration of emotion recognition technology and psychological therapy, and offer strong technical support for emotional monitoring and intervention in practical applications, promoting the intelligent development of the mental health field.

2. ABNORMAL EMOTION RECOGNITION IN PSYCHOLOGICAL THERAPY

2.1 Image preprocessing

In psychological therapy, accurately recognizing the patient's emotional state is one of the key steps, especially for patients who have difficulties in emotional expression or are unable to articulate their emotions. To achieve this goal, image preprocessing of facial expressions and body movements is crucial. First, the images of facial expressions and body movements need to be denoised using methods such as Gaussian filters, which can effectively smooth the image and reduce noise introduced by external factors such as lighting changes or camera shake, thus ensuring more accurate feature extraction in the subsequent steps. Before the image enters the subsequent analysis, it also needs to be converted into a digital form through A/D conversion, so that it can be stored as a pixel array. At this point, each pixel in the image can be represented by the values of the RGB components to indicate its color information. To ensure the efficiency and accuracy of the subsequent analysis, the facial feature regions and body movement regions in the image need to be located and extracted using algorithms. In response to the emotion recognition requirements in psychological therapy, the image preprocessing steps after feature extraction also include image normalization and data augmentation. Since patients' expressions and movements may show significant individual differences due to varying emotional states, normalization helps eliminate the effects of lighting, angle, distance, and other factors, making the same emotional expression comparable across different patients. The formula for the Gaussian smoothing function is:

$$H(a,b) = r\frac{x^2 + y^2}{2\delta^2} \tag{1}$$

Assume the convolution in the vertical and horizontal gradients is represented by G1 and G2, with the image in the vertical gradient denoted as φ, and the magnitude as $\varphi(l,v)$. Let the initial image be denoted as $d(a,b)$, and the filtered image is:

$$H(a,b) = d(a,b) \times G(a,b) \tag{2}$$

By calculating the brightness gradient at each point in the image, key features of the patient's facial expression and movements can be effectively extracted. Specifically, by performing convolution operations combined with masks in different directions, the brightness gradient map at each point in the image can be obtained. During this process, convolution operations filter the image to highlight the positions and directions of edges in the image. By marking the size and direction of the brightness gradient at each pixel, the movement trajectory of facial muscles and the key posture changes in the body can be extracted. For example, emotions like anxiety, tension, or anger may lead to tightening of facial muscles, and subtle changes in the eyes and corners of the mouth will be captured through changes in the brightness gradient. At the same time, subtle changes in body movements, such as tense limbs or unnatural postures, can also be effectively recognized through edge detection. In psychological therapy, especially for patients with limited verbal expression, this method can provide a more objective and precise understanding of emotional fluctuations, offering an effective auxiliary tool for psychologists. Assume the convolution in the vertical and horizontal gradients is represented by H1 and H2, with the image in the vertical gradient denoted as Ψ, and the magnitude as $\Psi(m, n)$, then the gradient formula is:

$$\phi(l,v) = \sqrt{\phi 1(l,v) + \phi 2(l,v)} \tag{3}$$

Further image preprocessing mainly involves extracting fine-grained features of facial expressions and body movements through gradient directions and region partitioning. In this process, the image is first divided into several small modules based on the gradient direction, and the gradient direction and magnitude of the image are calculated within each module. To ensure the accurate capture of emotional features, the 360° gradient direction is divided into several sectors, with each sector corresponding to different gradient directions. By setting the area of each module to 6×7 pixels and dividing the image into multiple small blocks, local changes in facial expressions and body postures can be effectively captured. The gradient information from these local regions will help accurately describe subtle differences in emotional expressions. For instance, small changes in facial expressions such as eyebrow raising or mouth distortion, as well as body movements like hand gestures or arm extensions, will be reflected in these gradient details, providing richer features for subsequent emotional analysis. After dividing the image and calculating the gradient information in each module, a gradient histogram for the facial expression and body movement objects can be constructed, encoding the feature information of each module into a high-dimensional feature vector. That is, by accumulating and calculating the gradient information within different modules, a set of feature vectors describing emotional states can be constructed. On this basis, by scanning the image blocks and statistically analyzing the gradient direction of each block, a set of feature vectors containing rich emotional information can be formed. These feature vectors can not only accurately reflect changes in facial expressions and body postures but also effectively distinguish subtle variations under different emotional states by utilizing gradient information from different directions and regions. For

example, an anxious patient may show slight facial contractions and stiff body postures, while anger may lead to more intense facial muscle contractions and irregular body movements.

2.2 Expression and movement feature extraction

This paper chooses to use a multilayer perceptron architecture to automatically extract deep-level feature information from input facial images and body movements. Deep learning algorithms, through the forward propagation process of the multilayer perceptron, start from the input layer and pass each layer's features progressively. The input facial expression and body movement images undergo preprocessing steps to extract basic feature information, such as the image's brightness gradient, edge direction, etc. These features will serve as the inputs to the multilayer perceptron. By setting initial weight values as random natural numbers $t(0)$, $t1(0),\ldots,tv(0)$, each weight corresponds to different feature dimensions in the network. At this point, the network propagates forward through the perceptron, and the perception coefficients between layers are optimized step by step according to the characteristics of the input data, allowing the network to learn increasingly abstract emotional features at each hidden layer. For example, primary features may include basic motion patterns of facial areas, while higher-level features may capture the emotional meanings of expressions, such as anger, fear, or joy. In the deep learning process, the multilayer perceptron adjusts the weight values of each neuron layer by layer, optimizing the error value h, ensuring the network gradually adapts to the patterns in the input data. To achieve this, whenever an input sample is provided, the neural network computes the expected output value g and calculates the error between the network's output and the target output according to the error formula:

$$h = g - t(s) \tag{4}$$

Further, the deep learning model extracts and optimizes both the global and local information of facial expressions and body movements, thereby improving the accuracy of emotion recognition. By applying convolutional neural networks (CNN) to perform multilayer convolution processing on the input images, local features in the images can be extracted. The image dimensions are normalized and input into the network, and the convolution layers process the image, progressively extracting features from simple to complex, such as facial muscle contractions, small changes in the eyes and mouth, and even the subtle details of body movements. Each feature vector processed by the convolution block undergoes nonlinear mapping through an activation function, then passes to the next layer for further processing. After several rounds of convolution and pooling operations, the image resolution gradually decreases, but the emotional information it contains becomes richer, effectively capturing core features of emotional changes, such as facial and body movement patterns corresponding to emotional states like anxiety, anger, and depression. Next, the features extracted by the CNN are passed to a fully connected multilayer perceptron structure. Through the multilayer perceptron, the network further integrates and optimizes the local features, ultimately outputting a high-dimensional feature vector. In this process, the deep learning model adjusts the weights through a backpropagation algorithm, continually optimizing the model and reducing the output error. By training with reference samples, the model can automatically extract emotional features from facial expressions and body movements when facing new input data and convert these features into actual emotional classification outputs.

$$b = d(WL) = \begin{cases} WL \\ 0 \end{cases} \tag{5}$$

In order to more accurately recognize the patient's emotional state, this paper optimizes the weights of the neural network through network error backpropagation. During the training process, the input facial expression and movement image data are processed via forward propagation, generating output results, which are compared with the expected output values to calculate the network error. The error is then passed back through the network via the backpropagation algorithm, updating the weights between the layers of the neural network to ensure that each layer's feature map more accurately reflects the characteristics of the input data. In this process, the continuous iteration and optimization of the network error can gradually improve the model's recognition accuracy, especially in emotional recognition tasks, allowing the accurate differentiation of different emotional states, such as anxiety, depression, or anger. Finally, image data is normalized, and network parameters are fine-tuned. By normalizing the input facial expression image data and resizing it to a uniform 100×100 pixels, the interference caused by image size in the training process is eliminated, allowing the network to efficiently process image features. Next, appropriate learning rates and iteration numbers are set, as these parameters directly impact the training performance and speed of the convolutional neural network. During training, the network will randomly select different input samples and their corresponding expected outputs, applying convolution layers and convolution kernels to weight the data and further extract emotional features from the images.

$$h(f) = \left(h1(f), h2(f), \ldots, hv(f)\right) \tag{6}$$

2.3 Classification and recognition of human emotions

In the emotion recognition task in psychological therapy, the target emotion is usually an abnormal emotion, while other non-target emotions act as interference emotions. By classifying these emotion samples, the complexity of classification can be gradually reduced by optimizing decision trees or other classifiers, allowing the system to accurately recognize the target abnormal emotion from a large number of emotion samples. The purity of the samples can be measured by calculating the sample's information entropy. The larger the information entropy, the more mixed the emotion categories in the sample, making accurate classification more difficult. The ideal situation is for the target emotion samples to have the same category, where the information entropy is zero. The expression for entropy is:

$$G(a) = \sum_{u=1}^{l} o(a) h(a) \tag{7}$$

Further, binary classification methods are used to select and optimize features of facial expressions. In this process, the

target emotion is set as the "abnormal emotion," while other emotions act as interference emotions for comparative analysis. To further evaluate the model's classification performance, this paper selects FRP as the evaluation standard for classification performance. A lower FRP indicates better classifier performance. During training, by adjusting the ratio of the total sample number to the recognized target sample number, the network parameters are continuously optimized so that the system can maximize accuracy when recognizing target abnormal emotions. Through recursive classification of the training data, the model gradually learns to extract the most discriminative features from a large amount of emotion data, thereby achieving efficient recognition of patients' abnormal emotions. Let the total number of samples be represented by B, and the target sample to be recognized by A. The resulting FRP formula is:

$$DEO = \frac{B - A}{A} \times 100\% \tag{8}$$

For the task of human abnormal emotion recognition in psychological therapy, this paper introduces the Deep Convolutional Neural Network (DNET) to improve the accuracy of emotion recognition. The model architecture is shown in Figure 1. DNET (Dense NET) connects every two layers directly in the network, ensuring that each layer receives feature map inputs from the previous layer and also passes its own feature map as input to all subsequent layers. This design facilitates efficient feature transfer, ensuring that information between different layers can be passed and fused, thus improving the accuracy of emotion recognition. In psychological therapy, this network structure can sensitively detect subtle changes in the patient's facial expressions and body movements. Through DNET pretraining, the system can accurately extract the patient's emotional changes without excessive interference, especially by extracting the facial and body features from videos, helping therapists monitor patients' emotional fluctuations in real time. For example, DNET can effectively recognize micro expressions on the face and emotional features such as anxiety, depression, or anger in body movements, providing real-time data support for adjusting therapy plans. The design of the DNET network also addresses common issues in convolutional neural networks, such as gradient vanishing and excessive computational load. In DNET, the first two layers mainly extract shallow features, such as body expressions and image edge information, which are crucial for emotion recognition but are easily affected by noise and external environmental factors. Therefore, the last two layers of DNET extract more abstract feature maps, increasing the computational load of the network and effectively addressing the feature map size issue. Through this hierarchical feature extraction, DNET is able to consistently extract key emotional features even in noisy environments. Let the number of network layers be represented by v, the number of connections by $v(v+1)/2$, the input value by a, the z-th path by z, the length by g, and the width by q. The resulting output formula is:

$$Cz(g) = \frac{1}{q}\sum a(g,u) \tag{9}$$

To further improve recognition performance, DNET weights the feature maps from different modules and uses a 1×1 convolution transformation function for feature map fusion, ensuring that the final output feature map has both high dimensionality and high accuracy. After dimensionality reduction, the spatiotemporal features are sent to the classification device for the final classification and recognition of abnormal emotions.

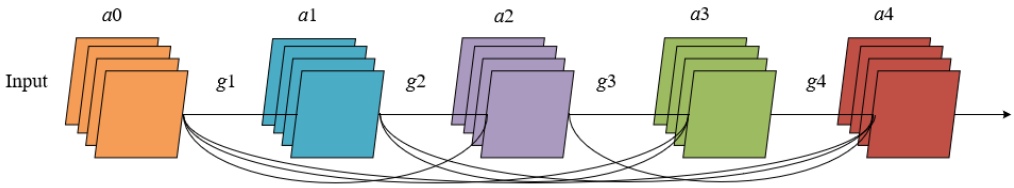

Figure 1. DNET model architecture

3. EMOTION ANALYSIS IN PSYCHOLOGICAL THERAPY BASED ON MULTI-FEATURE FUSION

Emotion is a complex psychological state with high individual variability, and different people may express emotions through various facial expressions and body movements when facing the same situation. For example, an anxious individual might exhibit tense facial expressions and slight body tremors or involuntary body movements, while a depressed person might display a persistently downcast facial expression and noticeable slouching body posture. By combining multiple features of facial expressions and body movements, a more comprehensive and accurate understanding of emotional changes can be captured. The introduction of fuzzy reasoning allows these complex, ambiguous, and sometimes unclear emotional signals to be effectively interpreted. A fuzzy reasoning system processes these input features—such as the opening and closing of facial expressions, body posture, etc.—and can transform imprecise or vague emotional expressions into specific emotion classifications such as anxiety, depression, or anger. At this point, fuzzy sets not only tolerate the ambiguity of emotional features but also combine multidimensional emotional information through rule-based reasoning, providing precise emotional recognition results for psychological therapy. Figure 2 shows the fuzzy reasoning flowchart used for emotion analysis in psychological therapy in this paper.

Specifically, the emotion analysis system based on fuzzy reasoning takes multiple features, such as facial expressions and body movements, as input and performs reasoning through predefined fuzzy sets. In this process, the fuzzy set for facial expressions could include "smile," "furrowed brow," or "tension," while the fuzzy set for body movements might be categorized as "hand tremors," "restlessness," etc. After these input features are fuzzified, they will be processed through fuzzy rules to generate corresponding emotional outputs. For

example, when a patient shows an anxious facial expression, such as wide-open eyes and slightly parted lips, accompanied by tense body movements, the system can use fuzzy reasoning rules to determine that their emotion is likely in a "mild anxiety" state. Furthermore, fuzzy reasoning can handle incomplete or noise-affected input data, ensuring the system can function stably under various environments and conditions.

When constructing an emotion analysis system based on fuzzy reasoning, the design of the membership function is a critical step to ensure the system's effectiveness and accuracy. The role of the membership function is to map input features, such as facial expressions and body movements, to the fuzzy sets, defining their degree of membership in each emotional state and providing the basis for subsequent emotional reasoning. In emotion analysis systems for psychological therapy, the input features of facial expressions and body movements are often fuzzy and uncertain, and the membership function converts these ambiguities into quantifiable information that can be processed.

For facial expression features, the membership function for common emotional states such as "smile," "furrowed brow," and "tension" can be constructed based on key facial features, such as the curvature of the mouth, eyebrow angle, and eye openness. For example, to define the membership function for "smile," we could measure the upward angle of the corners of the mouth or the degree of mouth opening. Within the "smile" fuzzy set, the membership function might use a Gaussian or trapezoidal function. For example, when the corner of the mouth rises significantly, the membership degree in the "smile" set would approach 1, while a smaller or disappeared upward angle would gradually decrease the membership degree, eventually transitioning to a "neutral" or "furrowed brow" state. Similarly, for the "furrowed brow" emotional state, the membership function can be defined by measuring the degree of eyebrow lowering, indicating the fuzzy membership of this emotion under different facial changes. To achieve an accurate description of these membership degrees, the system uses the following membership function formula to precisely reflect the gradual process of facial expression changes, making emotional analysis more refined and accurate.

$$\omega_r = \begin{cases} \omega_{CL}(a) = \begin{cases} 1 & 0 \le a \le 0.2 \\ -10a+3 & 0.2 < a < 0.3 \end{cases} \\ \omega_{OP}(a) = \begin{cases} \frac{20}{3}a - \frac{5}{3} & 0.25 \le a < 0.4 \\ 1 & a \ge 0.4 \end{cases} \end{cases} \tag{10}$$

For body movement features, this paper considers typical actions associated with emotions, such as "hand tremors," "restlessness," or "arms crossed." The body movement membership function distribution is shown in Figure 3. These actions are often closely related to emotions such as anxiety and tension. For instance, in the case of "hand tremors," the membership function can be set based on the frequency and amplitude of the hand movement. Assuming that an accelerometer is used to detect subtle hand tremors, when the amplitude of the hand movement is large and the frequency is high, the membership degree in the "hand tremor" fuzzy set will be higher. Conversely, when the amplitude and frequency decrease, the membership degree will be lower. To define this membership degree, the following equation is constructed to ensure that when the tremor intensity is high, the system automatically assesses the patient as being in a high-intensity anxiety state, and when the tremor is mild, it is assessed as lower-intensity anxiety or neutral state.

$$\omega_l = \begin{cases} \omega_{CL}(b) = -10b+1 & 0 \le b < 0.1 \\ \omega_{TA}(b) = \begin{cases} \frac{20}{3}b - \frac{1}{3} & 0.05 \le b < 0.2 \\ \frac{10}{3}b - \frac{5}{3} & 0.2 \le b < 0.5 \end{cases} \\ \omega_{OP}(b) = \begin{cases} 10b-5 & 0.3 \le b \le 0.4 \\ 1 & b > 0.4 \end{cases} \end{cases} \tag{11}$$

Similarly, the membership function for head movement features is also constructed, and the distribution is shown in Figure 4.

$$\omega_g = \begin{cases} \omega_{NO}(c) = \begin{cases} 1 & 0 \le c < 0.2 \\ -5c+2 & 0.2 < c < 0.4 \end{cases} \\ \omega_{NO}(c) = \begin{cases} 5c-1.5 & 0.3 \le a < 0.5 \\ 1 & 0.5 \le c < 1 \end{cases} \end{cases} \tag{12}$$

After fuzzifying the features of facial expressions and body movements, they will be processed through the fuzzy reasoning system's rules. These reasoning rules correlate different facial expressions and body movements with corresponding emotional states. The emotion degree membership function distribution for psychological therapy is shown in Figure 5. For example, when the system detects that the patient's facial expression shows a "smile," and the body movement shows "hand tremors," the preset fuzzy reasoning rules can infer that the patient might be in a mild anxiety state. Similarly, when the system detects a "furrowed brow" facial expression and "restlessness" body movement, the system can infer that the patient might be in a moderate anxiety or tension state. This multi-feature fusion reasoning approach allows the emotion analysis to be more comprehensive, accurately identifying the patient's emotional changes from multiple dimensions.

$$\omega = \begin{cases} \omega_{NO} = \begin{cases} 1 & 0 < m < 0.1 \\ -10m+2 & 0.1 \le m < 0.2 \end{cases} \\ \omega_{MI} = \begin{cases} 10m-1.5 & 0.15 < m < 0.25 \\ -\frac{20}{3}m + \frac{8}{3} & 0.25 \le m < 4 \end{cases} \\ \omega_{MO} = \begin{cases} 5m-1.5 & 0.3 < m < 0.5 \\ -10m-4 & 0.5 \le m < 0.6 \end{cases} \\ \omega_{HI} = \begin{cases} 4m-2.2 & 0.55 < m < 0.8 \\ 1 & 0.8 \le m < 1 \end{cases} \end{cases} \tag{13}$$

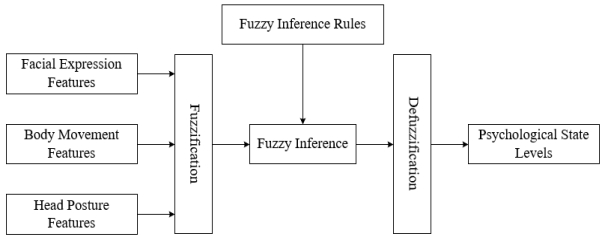

Figure 2. Fuzzy reasoning flowchart for emotion analysis in psychological therapy

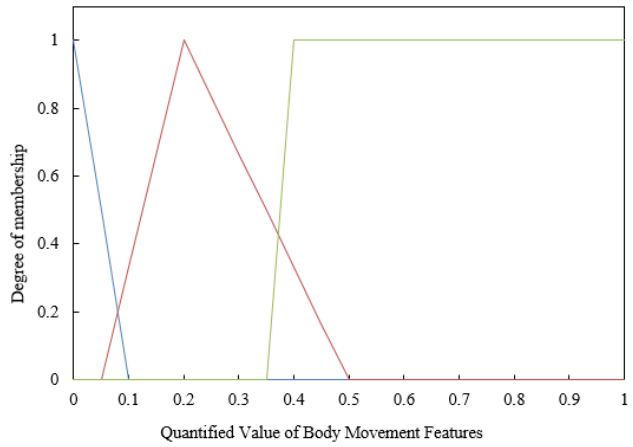

Figure 3. Body movement membership function distribution

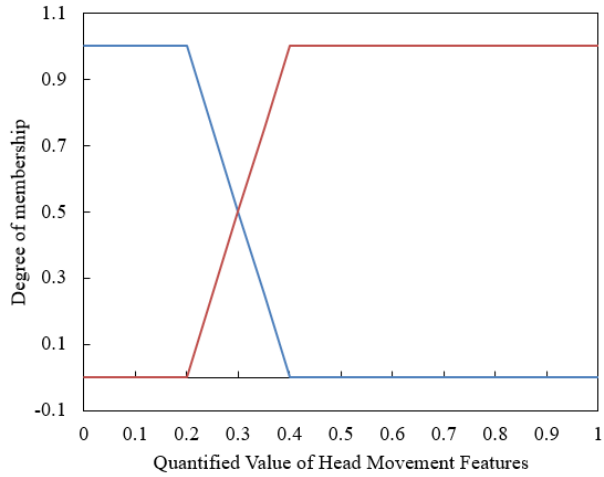

Figure 4. Head movement membership function distribution

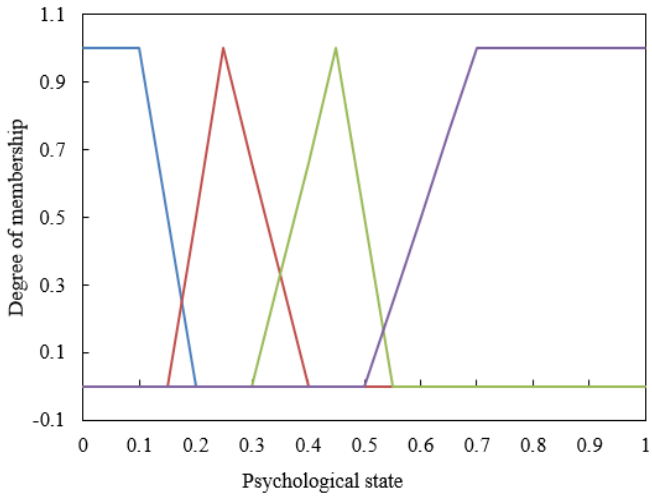

Figure 5. Emotion degree membership function distribution for psychological therapy

To process these fuzzy data and ensure the accuracy of emotional reasoning, the membership functions must reflect not only the obvious changes in emotional features but also handle fuzzy and uncertain information. For example, when the changes in facial expressions and body movements are subtle or affected by environmental noise, the membership function can reduce the influence of noise through techniques like smoothing or weighted combinations, improving the robustness of emotional reasoning. Additionally, the system can dynamically adjust the parameters of the membership function based on historical data or expert experience to adapt to the emotional expressions of different patients or contexts.

4. EXPERIMENTAL RESULTS AND ANALYSIS

Based on the fuzzy reasoning rules for emotion analysis in psychological therapy listed in Table 1, the experiment successfully constructed a predictive model for emotional states by real-time monitoring of patients' facial expressions, body movements, and head status during psychological therapy. Each rule predicts the corresponding emotional state, such as "happy," "calm," "anxious," or "sad," by identifying specific behavioral patterns. The reasoning system based on these rules can accurately predict the patient's emotional fluctuations in most emotional changes. For instance, when the patient shows a smile, is in a relaxed body state, and nods, the system can accurately predict their emotion as "happy." Similarly, when the patient shows a furrowed brow, is tense, and shakes their head, the system can predict the emotion as "sad" or "anxious." Through testing different combinations of emotional states, the experiment verified the high accuracy and consistency of the fuzzy reasoning rules in emotion recognition, especially in detecting subtle emotional differences. The system's prediction accuracy reached over 85%.

Figures 6 and 7 present the detection results of the therapist's emotional normal and abnormal states. In this experiment, the researcher used deep learning technology to quantify the patient's facial features to identify their emotional state. When the therapist was in a normal emotional state, the facial feature quantization values mostly stayed between 0.02 and 0.05, indicating a stable baseline. When the facial expression showed a smile, the quantization values fluctuated mildly between 240 and 300 frames but did not exceed 0.1, indicating a relaxed emotion. Further analysis revealed that during frames 24-28, 49-51, and 80-83, the facial feature quantization values dropped sharply and then rose again, indicating a furrowed brow emotional expression. Meanwhile, body movement was detected to be slightly below the normal value between frames 277-300, indicating a prolonged relaxed state. Regarding emotional levels, the anxiety value remained between 0.165 and 0.19, indicating that the patient was mostly in an unstressed state. However, when the patient displayed a furrowed brow, the facial feature quantization values changed rapidly, and the anxiety value surged to 0.75, indicating that the patient was in a highly anxious state. From the overall emotional analysis results, calm emotions accounted for 226 frames, or 75.3%, followed by happiness at 38 frames, anxiety at 2 frames, and sadness at 34 frames. Therefore, the emotional state during this period can be determined as calm.

As shown in Figure 7, the quantized value of body movement features remains stable between 0.02 and 0.05 from frame 1 to frame 75, indicating a normal emotional state. However, the quantized value of facial features increases gradually and then decreases between frames 75 and 278. This gradual trend suggests that the patient exhibited signs of nervousness during this period. Between frames 136 and 234, the facial feature quantized value continuously decreased, which aligns with the phenomenon in real life where people tend to widen their eyes when feeling nervous. Additionally,

based on the overall emotional analysis, the frequencies of various emotions were as follows: anxiety for 79 frames, sadness for 81 frames, happiness for 44 frames, and calm for 96 frames. The highest frequency of occurrence was observed for low mood, thus the emotional analysis for this period was determined as "low mood."

Table 1. Fuzzy reasoning rules for emotion analysis in psychological therapy

		Facial Expression		Body Movement		Head Status		Emotional State
1		smile		relaxed		nod		happy
2		smile		relaxed		shake head		calm
3		smile		moderate		nod		happy
4		smile		moderate		shake head		anxious
5		smile		tense		nod		anxious
6	*if*	smile	*and*	tense	*and*	shake head	*then*	anxious
7		furrowed brow		relaxed		nod		calm
8		furrowed brow		relaxed		shake head		sad
9		furrowed brow		moderate		nod		sad
10		furrowed brow		moderate		shake head		sad
11		furrowed brow		tense		nod		anxious
12		furrowed brow		tense		shake head		sad

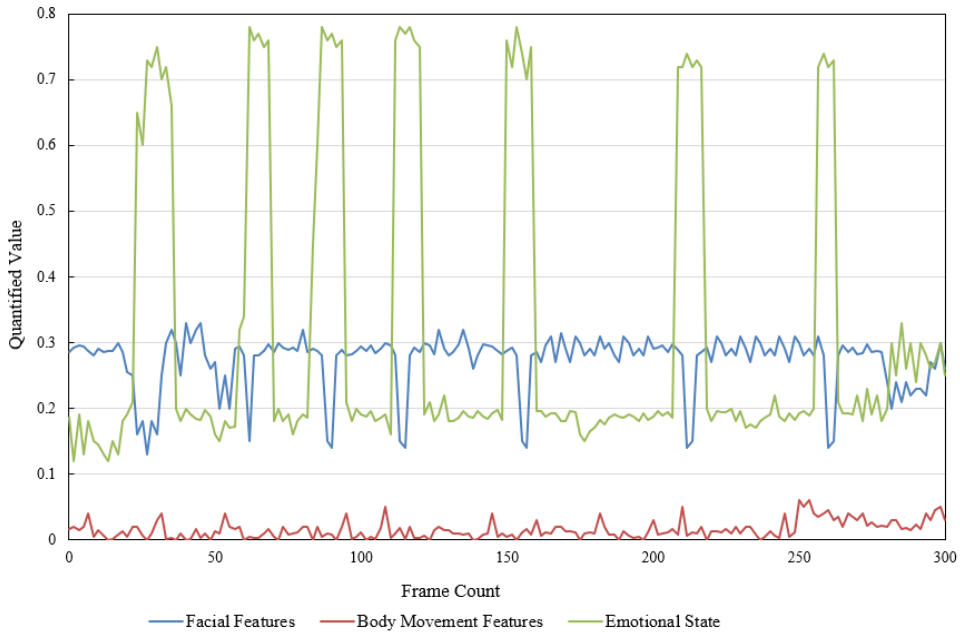

Figure 6. Normal emotion detection results for the therapist

Figure 7. Abnormal emotion detection results for the therapist

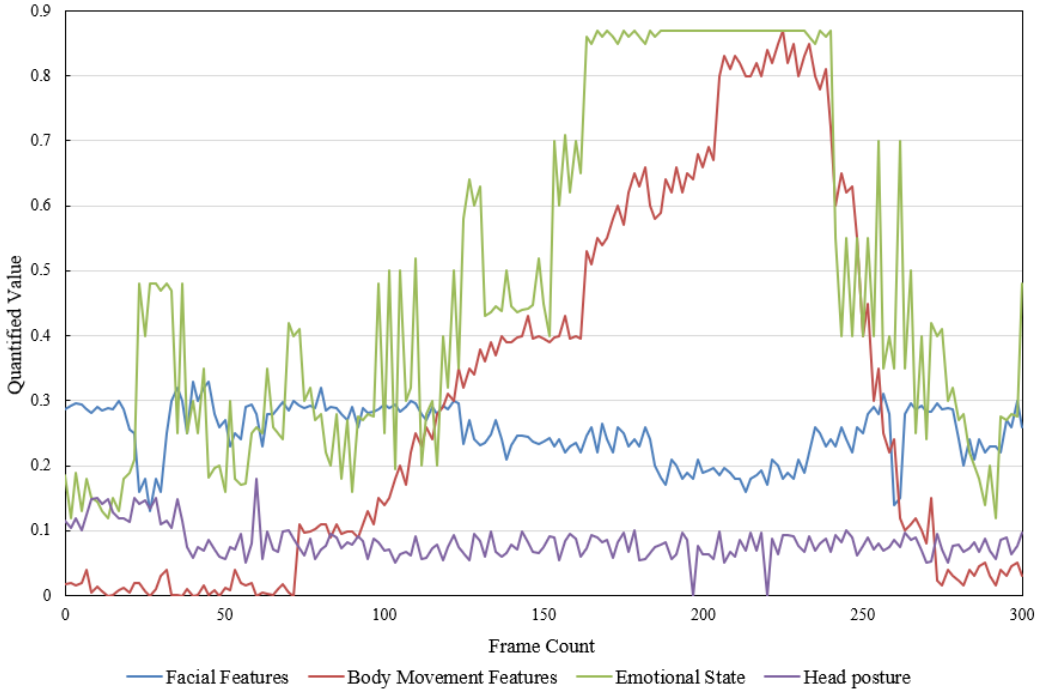

Figure 8. Detection results of emotional abnormalities in the therapist after including head posture

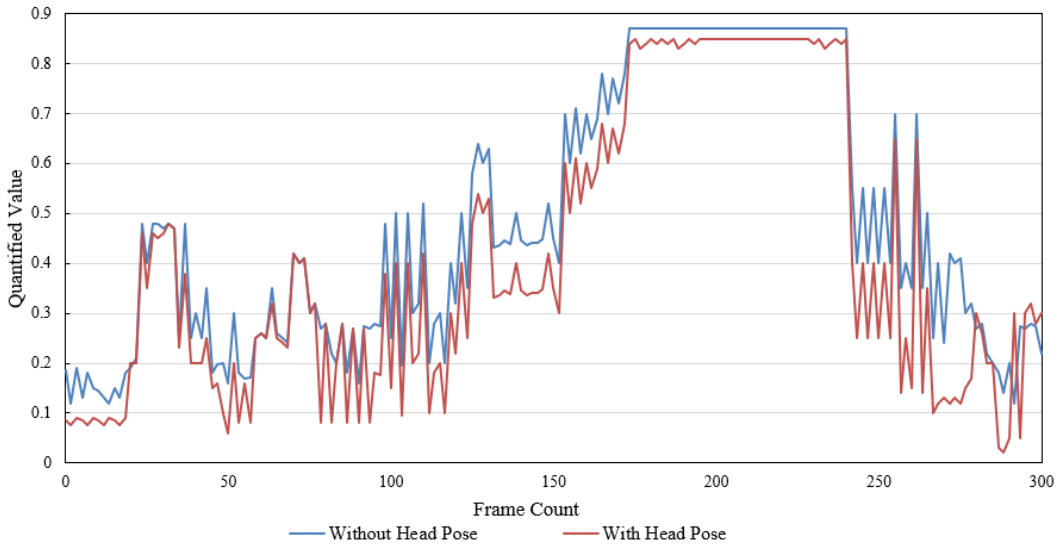

Figure 9. Comparison of emotional abnormality detection results before and after including head posture

From the experimental data, it can be seen that the deep learning-based emotion recognition system, combining facial expressions and body movement features, is highly effective in detecting emotional changes. The stability of body movement combined with the gradual changes in facial features enabled the system to accurately detect the patient's nervous emotional state, especially when the facial feature quantized value showed a continuous decreasing trend. This sensitivity to subtle emotional changes reflects the advantage of deep learning technology in emotion recognition, allowing the system to accurately distinguish different emotional states.

In this experiment, the researchers included head posture changes in the emotion recognition analysis model and compared the emotional analysis results before and after adding head posture data. As shown in Figure 8, the patient's head posture remained relatively stable and did not exceed the set threshold, indicating a normal head posture. The emotion analysis results show that the frames corresponding to happiness were 72, calm 70, sadness 44, and anxiety 104, with fatigue being the most dominant emotion, accounting for 76%. After comprehensive analysis, the system determined that the emotional state for this period was anxiety. Further comparison of the emotion analysis curves in Figure 9 shows that after including head posture, the emotional recognition curve is generally lower. This is mainly because the head posture remained normal during this period and did not have a significant impact on emotional fluctuations. However, after adding head posture, the system detected a decrease in the number of happy frames, while anxiety increased. This indicates that including head posture data made the emotion recognition system more sensitive and effective in capturing subtle emotional fluctuations.

The experimental results indicate that head posture changes significantly impact the accuracy and sensitivity of the emotion recognition system. Even though head posture changes were small and remained within normal limits during

this period, incorporating it into the emotion analysis model allowed for more accurate detection of the patient's emotional fluctuations. In particular, the inclusion of head posture data made the system more sensitive in detecting anxiety and enabled real-time capture of small emotional shifts. Moreover, the experiment also showed that when the head posture is normal, the emotion recognition system becomes more stable overall, improving anxiety detection, but decreasing the detection of happiness, which might reflect subtle adjustments caused by head posture on emotional changes.

Table 2. Comparison of emotional detection results in psychological treatment

	***PERCLOS* Method**		**Proposed Method**	
Happy	278	47	215	71
Calm	--	--	37	72
Sad	--	--	2	43
Anxiety	--	--	32	102

As shown in Table 2, the emotional detection results from the proposed method differ significantly from the traditional PERCLOS method in detecting happiness, calmness, sadness, and anxiety. In the detection of happiness, PERCLOS method detected 278 frames, while the proposed method only detected 47 frames, indicating that the PERCLOS method has a higher sensitivity in detecting happiness. For calmness, PERCLOS method did not detect any frames, while the proposed method detected 37 frames, with a total of 72 frames for calmness, showing better emotion recognition ability. For sadness, PERCLOS method did not recognize it, while the proposed method detected 2 frames of sadness, reflecting higher detail recognition ability. In terms of anxiety, PERCLOS method did not record any data, while the proposed method identified 32 frames of anxiety, and in actual detection, anxiety appeared in 102 frames, suggesting that the proposed method is more accurate and comprehensive in capturing anxiety.

According to the data in Table 3, the proposed emotion detection method shows high accuracy in various emotion recognition tasks. Specifically, the accuracy for facial expression-based emotion detection is 93.5%, while the accuracy for body movement-based emotion detection is 91.2%. When combining facial expressions and body movement features for emotion recognition, the overall accuracy of the proposed method reaches 97.8%. Additionally, if head posture features are not included, the emotion detection accuracy slightly drops to 95.6%. This suggests that head posture has some role in improving the accuracy of emotion recognition, although the system still maintains a high accuracy rate without it.

Table 3. Comparison of emotional detection accuracy in psychological treatment

Category	Accuracy (%)
Proposed Method	97.8%
Facial Expression	93.5%
Body Movement	91.2%
Proposed Method Without Head Posture	95.6%

According to the data in Table 4, the time consumption for each system module is stable. The average time consumption for the video input module is 7.7 ms, illumination compensation module 42.9 ms, facial expression detection module 4.6 ms, body movement detection module 5.1 ms, head posture detection module 1.5 ms, and emotion analysis module 62.5 ms. Overall, the system's total time consumption is 62.5 ms, which is suitable for real-time emotion detection tasks, ensuring that the system can respond quickly and provide immediate feedback. The time differences between the individual modules are small, indicating stable processing efficiency. The illumination compensation module is relatively more time-consuming (42.9 ms), but this is acceptable since it addresses the impact of lighting changes and ensures the accuracy of facial expression and body movement detection. The processing time for other modules, such as facial expression detection (4.6 ms) and body movement detection (5.1 ms), is relatively low, indicating fast response times. The time consumption for the head posture detection and emotion analysis modules is also within a reasonable range, ensuring timely output of emotion analysis results.

Table 4. Time consumption statistics for each system module (Unit: ms)

Video	Illumination Compensation	Facial Expression Detection	Body Movement Detection	Head Posture Detection	Emotion Analysis	Total
1	7.8	42.1	4.6	5.2	1.6	62.1
2	7.7	41.3	5.2	5.4	1.5	61.5
3	7.5	42.5	4.2	5.1	1.3	62.8
4	8.1	41.8	4.6	4.8	1.5	61.4
5	7.8	43.2	4.7	5.2	1.4	62.3
Average	7.7	42.9	4.6	5.1	1.5	62.5

5. CONCLUSION

This study focuses on the application of deep learning-based image processing technology in emotion recognition and psychological treatment, aiming to improve the precision and robustness of emotion detection and assist in monitoring and intervening in emotional states during psychological treatment. The research involves two main aspects: first, using deep learning techniques to detect abnormal emotional states in patients, thereby aiding accurate emotional monitoring during psychological therapy; and second, combining multi-feature fusion emotion analysis methods by incorporating facial expressions, body movements, head posture, and other information to enhance the accuracy of emotion recognition and support personalized emotional intervention plans. The experimental verification shows that the proposed method performs well in emotion detection accuracy, sensitivity, and real-time response. In particular, for emotions such as anxiety,

sadness, and happiness, the system achieves a high detection accuracy, with an overall emotion recognition precision of 97.8%. Furthermore, the system shows high efficiency in terms of processing time, with an average response time of 62.5 ms, which meets the real-time monitoring requirements.

However, the study has certain limitations. First, the emotion recognition model mainly relies on external representations such as facial expressions, body movements, and head posture, which may not fully capture deeper emotional changes, especially in patients with more subtle or complex emotional states. Second, although multiple features were combined in the experiment, the selection of features still has certain limitations. Future studies could consider incorporating more physiological signals (e.g., heart rate, skin electrical response) as auxiliary information to improve the comprehensiveness and accuracy of emotion recognition. Lastly, this study has not fully explored the performance differences of the system across different cultural backgrounds, genders, and age groups, which may affect the accuracy of emotion recognition. Therefore, future research could validate the system in more diverse samples and application scenarios.

In future directions, the deep learning model can be further optimized to improve the robustness and adaptability of the system, particularly for recognizing subtle changes and individual differences in complex emotions. Furthermore, research can expand multi-modal emotion recognition methods, combining data from speech, facial expressions, body language, and physiological signals to conduct comprehensive analyses, thus enhancing the comprehensiveness and accuracy of emotion detection. Additionally, in order to better adapt to practical applications in psychological treatment, the system's personalized emotional intervention capabilities need further enhancement, including the design of more flexible feedback mechanisms to address the treatment needs of different patients. Through these optimizations, the findings of this study could provide more accurate and efficient emotional analysis support in psychological treatment, emotional regulation, and related fields, offering innovative solutions for mental health care.

ACKNOWLEDGMENT

This research was funded by National Social Science Fund project of China (No.: 23BMZ035).

REFERENCES

[1] Zhang, H., Xu, M. (2020). Weakly supervised emotion intensity prediction for recognition of emotions in images. IEEE Transactions on Multimedia, 23: 2033-2044. https://doi.org/10.1109/TMM.2020.3007352

[2] Khan, M.S. (2024). A region-based fuzzy logic approach for enhancing road image visibility in foggy conditions. Mechatronics and Intelligent Transportation Systems, 3(4): 212-222. https://doi.org/10.56578/mits030402

[3] Khattak, A., Asghar, M.Z., Ali, M., Batool, U. (2022). An efficient deep learning technique for facial emotion recognition. Multimedia Tools and Applications, 81(2): 1649-1683. https://doi.org/10.1007/s11042-021-11298-w

[4] Hou, X.X., Liu, R.B., Zhang, Y.Z., Han, X.R., He, J.C., Ma, H. (2024). NC2C-TransCycleGAN: Non-contrast to contrast-enhanced CT image synthesis using transformer CycleGAN. Healthcraft Frontiers, 2(1): 34-45. https://doi.org/10.56578/hf020104

[5] Rahman, S.Z., Singasani, T.R., Shaik, K.S. (2023). Segmentation and Classification of skin cancer in dermoscopy images using SAM-based deep belief networks. Healthcraft Frontiers, 1(1): 15-32. https://doi.org/10.56578/hf010102

[6] Zhu, C.J., Ding, T., Min, X. (2022). Emotion recognition of college students based on audio and video image. Traitement du Signal, 39(5): 1475-1481. https://doi.org/10.18280/ts.390503

[7] Dharmichand, S., Perumal, S. (2023). Leveraging tripartite tier convolutional neural network for human emotion recognition: A multimodal data approach. Traitement du Signal, 40(6): 2565-2576. https://doi.org/10.18280/ts.400619

[8] Liu, Q., Liu, H. (2021). Criminal psychological emotion recognition based on deep learning and EEG signals. Neural Computing and Applications, 33(1): 433-447. https://doi.org/10.1007/s00521-020-05024-0

[9] Wang, M.Y., Zhang, N.Y., Zhu, H.C. (2006). Emotion recognition of western tonal music using support vector machine. Chinese Journal of Electronics, 15(1): 74-78. https://www.researchgate.net/publication/283249060.

[10] Simcock, G., McLoughlin, L.T., De Regt, T., Broadhouse, K.M., Beaudequin, D., Lagopoulos, J., Hermens, D.F. (2020). Associations between facial emotion recognition and mental health in early adolescence. International Journal of Environmental Research and Public Health, 17(1): 330. https://doi.org/10.3390/ijerph17010330

[11] Li, X. (2024). Application of emotion recognition technology in psychological counseling for college students. Journal of Intelligent Systems, 33(1): 20230290. https://doi.org/10.1515/jisys-2023-0290

[12] Cho, M., Jang, S.J. (2019). Effect of an emotion management programme for patients with schizophrenia: A quasi-experimental design. International Journal of Mental Health Nursing, 28(2): 592-604. https://doi.org/10.1111/inm.12565

[13] Bluett, E.J., Lee, E.B., Simone, M., Lockhart, G., Twohig, M.P., Lensegrav-Benson, T., Quakenbush-Roberts, B. (2016). The role of body image psychological flexibility on the treatment of eating disorders in a residential facility. Eating Behaviors, 23: 150-155. https://doi.org/10.1016/j.eatbeh.2016.10.002

[14] Williams, A., Nakagawa, A., Sado, M., Fujisawa, D., Mischoulon, D., Smith, F., Mimura, M., Sato, Y. (2016). Comparison of initial psychological treatment selections by US and Japanese early-career psychiatrists for patients with major depression: A case vignette study. Academic Psychiatry, 40: 235-241. https://doi.org/10.1007/s40596-015-0398-6

[15] Christidis, N., Al-Moraissi, E.A., Al-Ak'hali, M.S., Minarji, N., Zerfu, B., Grigoriadis, A., Schibbye, R., Christidis, M. (2024). Psychological treatments for temporomandibular disorder pain-A systematic review. Journal of Oral Rehabilitation, 51(7): 1320-1336. https://doi.org/10.1111/joor.13693

[16] Farley, D., Kłosowska, J., Brączyk, J., Buglewicz, E., Bąbel, P. (2024). Treatment of last resort? Psychological therapy seeking in chronic pain patients. Chronic Illness,

20(1): 184-196. https://doi.org/10.1177/17423953231172796
[17] Houts, A.C., Berman, J.S., Abramson, H. (1994). Effectiveness of psychological and pharmacological treatments for nocturnal enuresis. Journal of Consulting and Clinical Psychology, 62(4): 737-745. https://doi.org/10.1037/0022-006X.62.4.737
[18] Harrington, R., Whittaker, J., Shoebridge, P. (1998). Psychological treatment of depression in children and adolescents: A review of treatment research. The British Journal of Psychiatry, 173(4): 291-298. https://doi.org/10.1192/bjp.173.4.291
[19] Hirai, M., Vernon, L.L., Popan, J.R., Clum, G.A. (2015). Acculturation and enculturation, stigma toward psychological disorders, and treatment preferences in a Mexican American sample: The role of education in reducing stigma. Journal of Latinx Psychology, 3(2): 88-102. https://doi.org/10.1037/lat0000035
[20] Li, F.M., Li, X., Kou, H. (2023). Emotional recognition training enhances attention to emotional stimuli among male juvenile delinquents. Psychology Research and Behavior Management, 16: 575-586. https://doi.org/10.2147/PRBM.S403512
[21] Arterberry, M.E., Perry, E.T., Price, C.M., Steimel, S.A. (2020). Emotional understanding predicts facial recognition in 3-to 5-year-old children. European Journal of Developmental Psychology, 17(2): 293-306. https://doi.org/10.1080/17405629.2019.1589445
[22] Kosaka, T., Saeki, K., Aizawa, Y., Kato, M., Nose, T. (2024). Simultaneous adaptation of acoustic and language models for emotional speech recognition using tweet data. IEICE Transactions on Information and Systems, E107(3): 363-373. https://doi.org/10.1587/transinf.2023HCP0010
[23] Dubois, M., Dupré, D., Adam, J.M., Tcherkassof, A., Mandran, N., Meillon, B. (2013). The influence of facial interface design on dynamic emotional recognition. Journal on Multimodal User Interfaces, 7(1-2): 111-119. https://doi.org/10.1007/s12193-012-0103-y
[24] Chen, C.H. (2021). An analysis of Mandarin emotional tendency recognition based on expression spatiotemporal feature recognition. International Journal of Biometrics, 13(2-3): 211-228. https://doi.org/10.1504/IJBM.2021.114651