

Facial recognition and analysis: A machine learning-based pathway to corporate mental health management

Zicheng Zhang^{1,2} , Tianshu Zhang³ and Jie Yang²

Abstract

Background: In modern workplaces, emotional well-being significantly impacts productivity, interpersonal relationships, and organizational stability. This study introduced an innovative facial-based emotion recognition system aimed at the real-time monitoring and management of employee emotional states.

Methods: Utilizing the RetinaFace model for facial detection, the Dlib algorithm for feature extraction, and VGG16 for micro-expression classification, the system constructed a 10-dimensional emotion feature vector. Emotional anomalies were identified using the K-Nearest Neighbors algorithm and assessed with a 3σ -based risk evaluation method.

Results: The system achieved high accuracy in emotion classification, as demonstrated by an empirical analysis, where VGG16 outperformed MobileNet and ResNet50 in key metrics such as accuracy, precision, and recall. Data augmentation techniques were employed to enhance the performance of the micro-expression classification model.

Conclusion: These techniques improved the across diverse emotional expressions, resulting in more accurate and robust emotion recognition. When deployed in a corporate environment, the system successfully monitored employees' emotional trends, identified potential risks, and provided actionable insights into early intervention. This study contributes to advancing corporate mental health management and lays the foundation for scalable emotion-based support systems in organizational settings.

Keywords

Emotion recognition, corporate mental health, facial analysis, micro-expression classification, deep learning

Received: 27 December 2024; accepted: 1 April 2025

Introduction

In today's fast-paced and high-pressure social environment, mental health issues have gradually become a global concern.¹ Mental health not only affects an individual's quality of life but is also directly related to workplace productivity and societal stability.² While the mental health of adolescents has been a key focus of research and policy,^{3,4} that of adults, particularly middle-aged and young working populations, has not received sufficient attention. As the backbone of society, these individuals bear the dual responsibilities of economic and family obligations, while also facing high-intensity workloads, workplace competition, and the challenge of balancing career development with family life.

Under such pressures, emotional fluctuations and mental health issues are becoming increasingly prevalent in middle-aged and young working populations, affecting their job performance and interpersonal relationships.

If emotional problems are not promptly identified and managed, they can trigger a series of chain reactions including productivity decline, workplace conflict, employee turnover, disruptions to business operations, and broader societal stability.^{2,5} Thus, building an effective emotion data collection and analysis system for the real-time monitoring

¹School of Modern Post, Nanjing University of Posts and Telecommunications, Nanjing, China

²Research and Development Department, Nanjing Yunshe intelligent technology Co., LTD, Nanjing, China

³School of Information Management, Nanjing University, Nanjing, China

Corresponding author:

Zicheng Zhang, School of Modern Post, Nanjing University of Posts and Telecommunications, Nanjing, China; Nanjing Yunshe intelligent technology Co., LTD, Nanjing, China.

Email: 18551701375@163.com



and management of employees' emotional states has become a critical topic in mental health management.⁶

This study aimed to develop a facial-based emotion data collection and analysis system to monitor employees' emotional states in real time and identify potential emotional anomalies, thereby providing businesses with effective management and intervention tools. Specifically, this study employs the RetinaFace model for facial detection, the Dlib algorithm for facial feature extraction, and deep learning models for micro-expression classification, ultimately constructing a 10-dimensional emotion feature vector. Additionally, the system utilizes the K-Nearest Neighbors (KNN) algorithm to identify emotional anomalies and applies the 3σ principle for risk level assessment, helping companies detect and address employees' emotional issues promptly.

To validate the proposed method, we will conduct empirical research in a corporate environment, analyze emotional data, and identify anomalous patterns to evaluate the practical effectiveness of the system. The goal is to provide businesses with a scientific framework for emotion monitoring and intervention, enhancing employees' mental health and productivity, while mitigating potential risks arising from emotional issues.

Related studies

Emotion recognition technology has broad applications across various fields, particularly healthcare,⁷ education,⁸ transportation,⁹ and marketing.¹⁰ In the domain of mental health, emotion recognition enables real-time monitoring of individuals' emotional states, helping professionals identify and intervene in emotional disorders, thus improving patients' psychological well-being.¹¹ In education, teachers can use emotion recognition technology to evaluate students' emotional responses, adjust their teaching strategies, and enhance classroom engagement and learning outcomes.¹² In transportation, emotion recognition analyzes drivers' emotions to deliver intelligent services and improve driving experience and safety.¹³ In marketing, emotion analysis technology is increasingly adopted, including the tourism sector, where facial expression analysis (FEA) tracks viewers' emotional reactions to different segments of advertisements, enabling more precise emotional targeting and optimized marketing strategies.¹⁴ Despite significant progress in these areas, the real-time monitoring and management of employee emotions in corporate environments remain underdeveloped, underscoring the need for more effective research and solutions.

Emotion recognition methods currently include text sentiment analysis, speech emotion recognition, FEA, and physiological signal monitoring.^{15–17} Speech emotion recognition relies on acoustic features such as prosody, spectral features, and voice quality¹⁸ and employs machine learning algorithms, such as support vector machines^{19,20} and hidden

Markov models,^{21,22} and deep learning algorithms, such as convolutional neural networks (CNNs),^{23–25} recurrent neural networks,^{26,27} and capsule networks.²⁸ Facial expression recognition (FER) often employs deep learning models, such as CNNs, region-based convolutional neural networks (R-CNN), and vision transformers.²⁹ These models demonstrate exceptional performance in handling FER tasks, particularly CNNs, which are widely used in FER systems owing to their effective integration of feature extraction and classification.³⁰ FER methods are especially suitable for corporate environments because of their non-intrusive nature and the widespread availability of the required devices, offering high applicability and scalability. Additionally, physiological signal monitoring, which collects data such as heart rate and skin conductance to infer emotional changes,³¹ is challenging to use in corporate settings owing to equipment requirements and privacy concerns.

Therefore, this study aimed to develop an integrated emotion recognition system that combines real-time data collection and deep learning analysis to identify and assess employees' emotional states and provide effective intervention strategies. Such a system not only contributes to advancing corporate mental health management and improving employee job satisfaction but also mitigates potential societal risks arising from emotional issues, fostering sustainable corporate development.

Method

This study established a comprehensive emotion monitoring and evaluation framework to quantify and detect anomalies in employees' emotional states. The framework consists of four parts: emotion data collection, emotion vector construction, anomaly detection using the KNN algorithm, and final risk level determination. It employs a camera-based system for seamless data collection, deep learning models for facial image analysis and feature extraction, and emotion vectors to represent individual emotional states. Anomalies are identified using the KNN, and a risk-grading mechanism is applied to provide quantitative assessments and early intervention for employees' mental well-being.

Emotion data collection

The camera-based seamless data collection system continuously captures video data from the surrounding environment and automatically extracts facial images from the background. This enables the real-time monitoring of employees' emotional states while minimizing privacy concerns.

This study utilized the RetinaFace model,³² an efficient CNN-based facial detection model. RetinaFace achieves precise facial region localization through multiscale feature

fusion and a dense feature extraction network, even under complex lighting and varying angles.

After detecting a face, the system further utilizes facial key points (e.g. left and right eyes, nose tip, and mouth corners) to perform geometric corrections on the image. Assuming the coordinates of the left and right eye key points are (x_a, y_a) and (x_b, y_b) , the tilt angle θ between the eyes can be calculated as follows:

$$\theta = \arctan\left(\frac{y_b - y_a}{x_b - x_a}\right) \quad (1)$$

Then, the image is rotated by an angle of $-\theta$ to align the face into a consistent orientation, ensuring that the left and right eyes are positioned on a horizontal line. The geometrically corrected facial image can then provide a consistent input for subsequent analyses, including emotion recognition and feature extraction.

Facial feature point detection was performed using the Dlib library. Dlib is a machine learning library capable of precisely identifying important facial features (such as eyebrows, eyes, mouth, and nose). These features were then transformed into a 128-dimensional feature vector to form a data foundation for emotion analysis.

During facial identity verification, the system uses 128-dimensional facial feature vectors obtained by Dlib and compares their similarity using cosine similarity to confirm the identity of the captured subject. The formula is as follows:

$$\begin{aligned} \text{similarity} &= \cos(\theta) = \frac{A \cdot B}{\|A\| \cdot \|B\|} \\ &= \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \end{aligned} \quad (2)$$

where A_i and B_i represent the components of the 128-dimensional vectors A and B for the template and captured facial data, respectively.

Data augmentation

In this study, image alignment was performed. The main challenge anticipated in facial micro-expression classification was varying lighting conditions. To address this, we adopted a data augmentation method based on saturation adjustment, which modifies the color saturation of an image to alter its vividness. When the saturation is low, the image appears dull, whereas higher saturation makes the image more vibrant. Saturation adjustment enhances the robustness of the model to a broader range of colors, thereby improving its generalization capability.

In the HSV color space, saturation (S) is an indicator of the purity or vividness of the color. By adjusting the saturation, the colors in the image can become either more vibrant or muted. The formula for adjusting the saturation

is as follows:

$$I_{\text{new}} = I_{\text{original}} \times (1 + \Delta S) \quad (3)$$

where I_{new} is the adjusted image, I_{original} is the original image, and ΔS is the change in saturation, which controls the increase or decrease of saturation. If $\Delta S > 0$, the saturation increases, making the image colors more vivid. If $\Delta S < 0$, the saturation decreases, making the image colors more muted. If $\Delta S = 0$, the saturation remains unchanged, and the image colors remain the same.

Emotion vector construction

After emotion data collection and facial feature extraction, the next step involves constructing emotion vectors that quantify individual emotional states using micro-expression classification models and extended emotional dimensions.

Micro-expression classification and emotional dimension expansion. In the process of constructing the emotion vector, micro-expression classification was first performed on the 128-dimensional facial features extracted using a deep learning model. This model was trained using the CASME micro-expression database. The CASME micro-expression database, developed by the Institute of Psychology, Chinese Academy of Sciences, contains 195 samples from 19 participants and is labeled with eight emotional categories: amusement, sadness, tension, disgust, fear, surprise, repression, and contempt.³³ This includes onset, peak, and offset frame annotations, along with Action Unit information. However, the annotation quality for contempt was inconsistent, making the classification less reliable.³⁴ To mitigate the issues of dataset imbalance and improve classification robustness, this study excluded the contempt category and refined the categories by merging similar expressions and removing the underrepresented ones. This resulted in a final set of seven primary emotions: happy, sad, angry, disgust, fear, surprise, and neutral, which formed the core of emotion recognition in our study.

To further enhance classification accuracy and improve interpretability, this study adopted a hybrid emotion grouping strategy inspired by existing research aimed at expanding emotional dimensions. Chen et al.³⁵ proposed a three-dimensional emotional structure consisting of positive, negative, and neutral emotions, whereas Li et al.³⁶ introduced a four-category encoding approach: positive, negative, surprise, and others. Based on these frameworks, this study classified seven primary emotions into three higher level emotional dimensions: positive emotions (happiness, surprise), negative emotions (sadness, anger, fear), and complex emotions (disgust, neutral). This leads to a 10-dimensional emotion feature vector: happy, sad, angry, disgust, fear, surprise, neutral, positive, negative, and

complex emotions. Expanding emotional dimensions helps to capture more complex emotional states, as detailed categories may struggle with nuanced or mixed emotions, whereas broader dimensions provide a more effective framework for classification.

Quantifying individual emotion vectors. In the micro-expression classification task of deep learning models, the Softmax layer is a critical component for multiclass classification. It converts the output of the model into a probability distribution across the emotional categories. For example, emotions such as happiness, sadness, anger, disgust, fear, surprise, and neutrality can be quantified using a Softmax layer.

The first is the choice of deep learning models. MobileNet, introduced by Howard et al.,³⁷ is a lightweight deep neural network tailored for mobile and embedded vision applications. It employs depth-wise separable convolutions to construct a streamlined structure, significantly reducing the model size and computational demands, while maintaining commendable performance.

Another prominent model is ResNet50, introduced by He et al.,³⁸ which is known for its deep residual learning framework. The key innovation of ResNet is the use of residual connections that allow the model to bypass certain layers, thereby enabling the training of significantly deeper networks without the vanishing gradient problem. ResNet50, a variant with 50 layers, achieved state-of-the-art performance in various computer vision tasks by leveraging residual blocks. This model has shown exceptional accuracy in image classification tasks and is widely used because of its ability to generalize well even with a large number of layers.

VGG16 is a classic deep learning model that extracts high-level features of an image through stacked convolutional layers. VGG16 comprises 13 convolutional layers and three fully connected layers. Figure 1 illustrates its structure.

In deep learning models, the final output layer of the network is referred to as the logit, denoted as Z . These logits are unnormalized scores, which are typically real numbers. For a 7-class emotion classification task, the model outputs seven log components as follows:

$$Z = [z_1, z_2, z_3, z_4, z_5, z_6, z_7] \quad (4)$$

where z_1 to z_7 represent the predicted scores for each emotional category. Specifically, z_1 represents happiness, z_2 represents sadness, z_3 represents anger, z_4 represents disgust, z_5 represents fear, z_6 represents surprise, z_7 represents neutral.

The Softmax function is employed to convert logits into probabilities, ensuring that the sum of the probabilities across all emotional categories equals one. The Softmax

function is expressed as follows:

$$p_i = \frac{e^{z_i}}{\sum_{j=1}^7 e^{z_j}} \quad (5)$$

The probabilities obtained from the Softmax layer served as feature components for each emotion. For the 7-class emotion classification task, the Softmax output was

$$p = [p_1, p_2, p_3, p_4, p_5, p_6, p_7] \quad (6)$$

where each p_i corresponds to the probability of a specific emotional category: p_1 represents probability of happy, p_2 represents probability of sad, p_3 represents probability of anger, p_4 represents probability of disgust, p_5 represents probability of fear, p_6 represents probability of surprise, p_7 represents probability of neutrality.

Based on this, three additional extended emotional components were calculated as follows:

$$p_8 = \frac{p_1 + p_6}{2} \quad (7)$$

Negative emotion is as follows:

$$p_9 = \frac{p_2 + p_3 + p_5}{3} \quad (8)$$

Complex emotion is as follows:

$$p_{10} = \frac{p_4 + p_7}{2} \quad (9)$$

Thus, an enhanced 10-dimensional emotion feature vector is constructed:

$$p_{enhance} = [p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8, p_9, p_{10}] \quad (10)$$

The purpose of expanding the emotion feature vector to 10 dimensions was to capture individual emotional characteristics more comprehensively, enabling finer quantification and analysis of emotional states. Although the original seven emotional categories represented basic emotional states, real-world emotions are often more complex. This multidimensional emotion vector not only enhances emotion recognition accuracy but also provides robust data support for subsequent personality trait analysis and mental health interventions. A flowchart of the emotion data collection and emotion vector construction is shown in Figure 2.

Anomaly detection using KNN

To construct emotion vectors, further identification of anomalies in individual emotional fluctuations was performed. For this purpose, this study employs the KNN algorithm for anomaly detection. The primary aim was to identify anomalies based on the local density of data points indicating unusual emotional fluctuations. The basic principle of the KNN algorithm is as follows: if the density of a data point is significantly lower than that of its KNN,

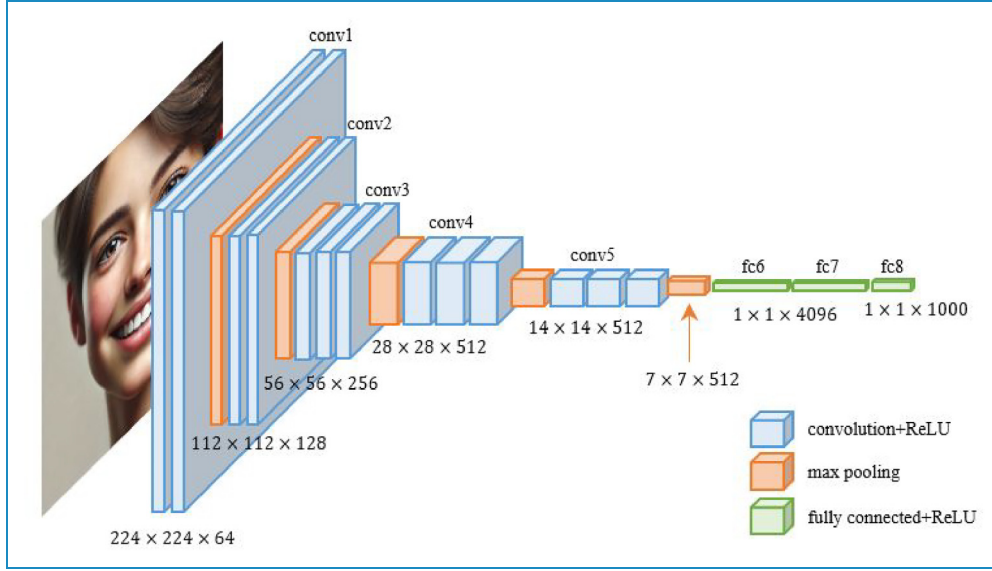


Figure 1. VGG16's model structure.

the point is likely to be an anomaly. Specifically, the distance between a data point and its neighboring points was calculated, and a local outlier factor (LOF) was used to quantify the anomaly. The anomaly detection steps were as follows:

1. Distance calculation: The Manhattan distance is used to compute the distance between each data point and the other data points and is defined as follows:

$$d(x_i, x_j) = \sum_{k=1}^n |x_{ik} - x_{jk}| \quad (11)$$

where x_i and x_j represent the two samples in the dataset, n is the number of features, and x_{ik} and x_{jk} are the k th feature values of the samples.

2. Find the KNN: For each data point, sort the distances in ascending order and select the K -nearest data points. The value of K is adjusted based on the size and characteristics of the dataset. In this study, K was set to 4.
3. Calculate local reachability density:

Determine the $K - \text{distance}$ for each data point p , representing the distance to its K th nearest neighbor:

$$K - \text{distance}(p) = d(p, p_k) \quad (12)$$

where p_k is the K th nearest neighbor of p .

Define the reachability distance of point p to point o :

$$\begin{aligned} \text{Reachability} - \text{distance}_K(p, o) \\ = \max(K - \text{distance}(o), d(p, o)) \end{aligned} \quad (13)$$

where $d(p, o)$ is the actual distance between p and o .

Compute the local reachability density of p :

$$\text{lrd}_K(p) = \frac{K}{\sum_{o \in N_K(p)} \text{reachability} - \text{distance}_K(p, o)} \quad (14)$$

where $N_K(p)$ represents the K nearest neighbors of p .

4. Calculate LOF:

The LOF of point p is calculated as follows:

$$\text{LOF}_K(p) = \frac{\sum_{o \in N_K(p)} \frac{\text{lrd}_K(o)}{\text{lrd}_K(p)}}{K} \quad (15)$$

If the LOF value of a point is significantly greater than 1, the density of the point is noticeably lower than that of its surrounding points, suggesting that it might be an anomaly. Conversely, if the LOF value is close to or less than 1, the point's density is similar to that of its neighbors, and the point is likely normal.

An anomaly threshold was set (e.g. 1.5 or 2), and points with LOF values exceeding this threshold were classified as anomalies.

In this study, K was set to four, meaning that for every point, the four nearest emotion vectors were used for the analysis. Weekly emotion vectors for individuals were considered ($n = 7$), and anomalies were detected using the KNN algorithm, followed by risk level assessments.

Risk level determination

To comprehensively evaluate an individual's emotional state, this study expands the seven basic emotional components to include 10 dimensions: encompassing positive, negative, and complex emotions. Each emotional

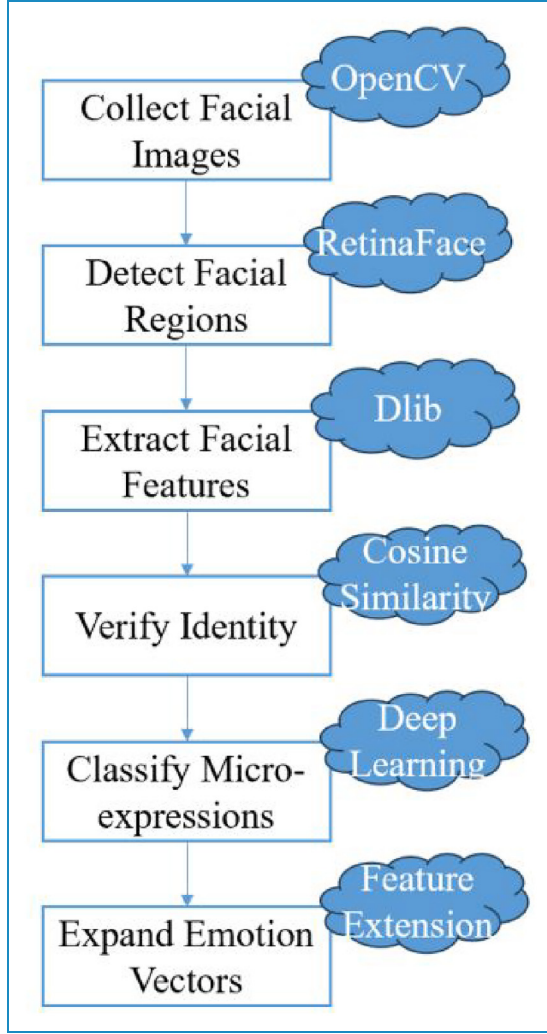


Figure 2. Emotion data collection and emotion vector construction process.

component and combined dimension was evaluated using a 3σ principle for risk detection. Specifically, if a negative emotional component exceeds the average value plus three standard deviations, the risk level increases by 1; if a positive emotional component falls below the average value minus three standard deviations, the risk level increases by 1; and if a complex emotional component exceeds the average value plus three standard deviations, the risk level increases by 1.

The series of past weekly emotional components for the i th emotional component is $E_i = \{e_{i1}, e_{i2}, \dots, e_{in}\}$, where n represents the total number of emotional components over the past week. The calculation methods are as follows:

Mean of weekly emotion component is as follows:

$$u_i = \frac{1}{n} \sum_{j=1}^n e_{ij} \quad (16)$$

Standard deviation of weekly emotion component is as follows:

$$\delta_i = \sqrt{\frac{1}{n} \sum_{j=1}^n (e_{ij} - u_i)^2} \quad (17)$$

For different emotion components, the specific risk level calculation guidelines are as follows.

First, for negative emotion components, if the i th negative emotion component e_i satisfies $e_i > u_i + 3\delta_i$, the risk level is increased by 1. Second, for positive emotion components, if the i th positive emotion component e_i satisfies $e_i < u_i - 3\delta_i$, the risk level is increased by 1. Third, for complex emotion components, if the i th complex emotion component e_i satisfies $e_i > u_i + 3\delta_i$, the risk level is increased by 1.

Thus, for any emotion component e_i , the risk level R is expressed as follows:

$$R = \begin{cases} R + 1, & \text{if } e_i > u_i + 3\delta_i \text{ for negative or complex - contempt emotion components} \\ R + 1, & \text{if } e_i < u_i - 3\delta_i \text{ for positive emotion components} \end{cases} \quad (18)$$

This method identified individuals with significant emotional fluctuations during the week, determined their risk levels, and provided strategic support for scientific interventions. Figure 3 illustrates the process of emotion data collection, vector construction, and risk level assessments.

From Figure 3, it is evident that the process of emotion data collection, vector construction, and risk level determination follows a structured sequence. It begins by classifying emotional features from facial images to extract the probability values for various emotions. These probabilities were then averaged hourly to capture temporal trends and forming a seven-day sequence of data. Using this sequence, feature distances were calculated and the nearest neighbors were identified to evaluate potential anomalies in emotional patterns. The detected anomalies were categorized as normal or abnormal, and emotional states were analyzed to assess individual risk levels. This systematic workflow ensures accurate emotional monitoring and timely intervention for managing significant emotional deviations.

Results

Experimental parameter setting

Table 1 lists the hardware and GPU with 24GB of VRAM. The training code was implemented in Python using the PyTorch deep learning framework.

The micro-expression classification dataset used in this study consists of facial images collected from the Internet and categorized into seven distinct emotions: happy, sad, anger, disgust, fear, surprise, and neutral. As shown in

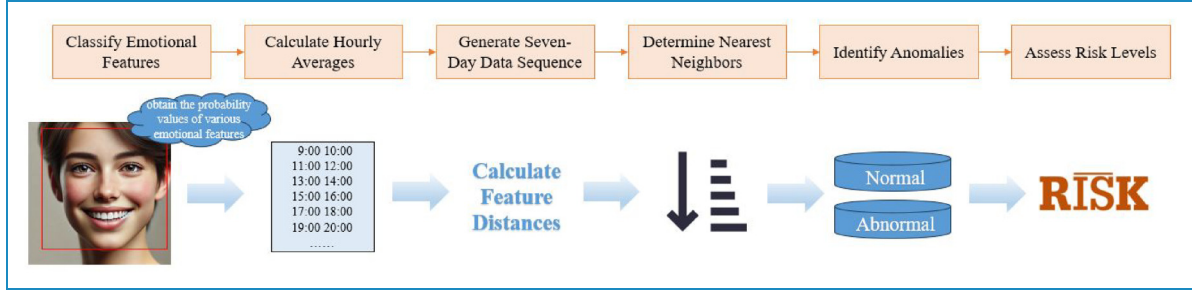


Figure 3. Flowchart of emotion data collection, vector construction, and risk level determination.

Table 1. Hardware and software configuration.

Name	Model number	Memory
CPU	Intel(R) Core(TM) i9-10850K CPU @ 3.60GHz 3.60 GHz	64.0 GB
GPU	NVIDIA GeForce RTX 3090	24.0 GB
Software	Python Torch	

Table 2, the dataset included 39,675 training and 4977 testing images for happy, 19,074 training and 2688 testing images for sad, 11,863 training and 1887 testing images for anger, 8189 training and 1479 testing images for disgust, 9690 training and 1645 testing images for fear, 14,057 training and 2131 testing images for surprise, and 41,074 training and 5133 testing images for neutral. This dataset serves as the foundation for training and evaluating the deep learning model, enabling the precise recognition and classification of micro-expressions.

Evaluation of micro-expression classification accuracy

A comparative analysis of the training results (Figure 4) showed that the VGG16 model significantly outperformed the MobileNet and ResNet50 models on the micro-expression training dataset. VGG16 demonstrated superior performance in terms of loss convergence speed and final value across several evaluation metrics. Specifically, after 170 iterations, VGG16's loss converged to 0.6069, whereas MobileNet's final loss was higher at 1.3382, and ResNet50's loss was 0.8538, indicating that VGG16 converged faster and more stably.

In terms of the accuracy, precision, and recall, VGG16 exhibited the strongest classification capabilities. Compared to MobileNet's accuracy, precision, and recall of 0.5016, 0.6869, and 0.2923, respectively, and ResNet50 values of 0.685, 0.7775, and 0.5861, VGG16 achieved accuracy, precision, and recall rates of 0.7791, 0.8329, and 0.7288, respectively. This indicated that VGG16 was better at correctly classifying micro-expression

Table 2. Data set description.

Train		Test	
Happy	39,675	Happy	4977
Sad	19,074	Sad	2688
Anger	11,863	Anger	1887
Disgust	8189	Disgust	1479
Fear	9690	Fear	1645
Surprise	14,057	Surprise	2131
Neutral	41,074	Neutral	5133

samples, reducing misclassifications, and improving recall rates. Notably, the performances of all three models improved with the inclusion of data augmentation, highlighting the effectiveness of color and brightness enhancement techniques. The accuracies of MobileNet, ResNet50, and VGG16 were improved by 0.0168, 0.0343, and 0.0056, respectively.

The experimental results for the validation dataset (Figure 5) demonstrated that VGG16 outperformed MobileNet and ResNet50 in terms of generalization capability and accuracy. VGG16 achieved a loss of 0.9871, which was significantly lower than that of MobileNet (1.1868) and ResNet50 (1.0028), indicating that VGG16 exhibited faster and more stable loss reduction during training.

The key performance metrics on the validation dataset further favored VGG16. Compared with MobileNet's accuracy, precision, and recall of 0.5645, 0.7369, and 0.3647, and ResNet50 values of 0.6267, 0.6915, and 0.5441, respectively, VGG16 achieved accuracy, precision, and recall rates of 0.7023, 0.7459, and 0.6627, respectively. These results highlight VGG16's superior ability to reduce both false negatives and false positives, thereby ensuring better classification performance.

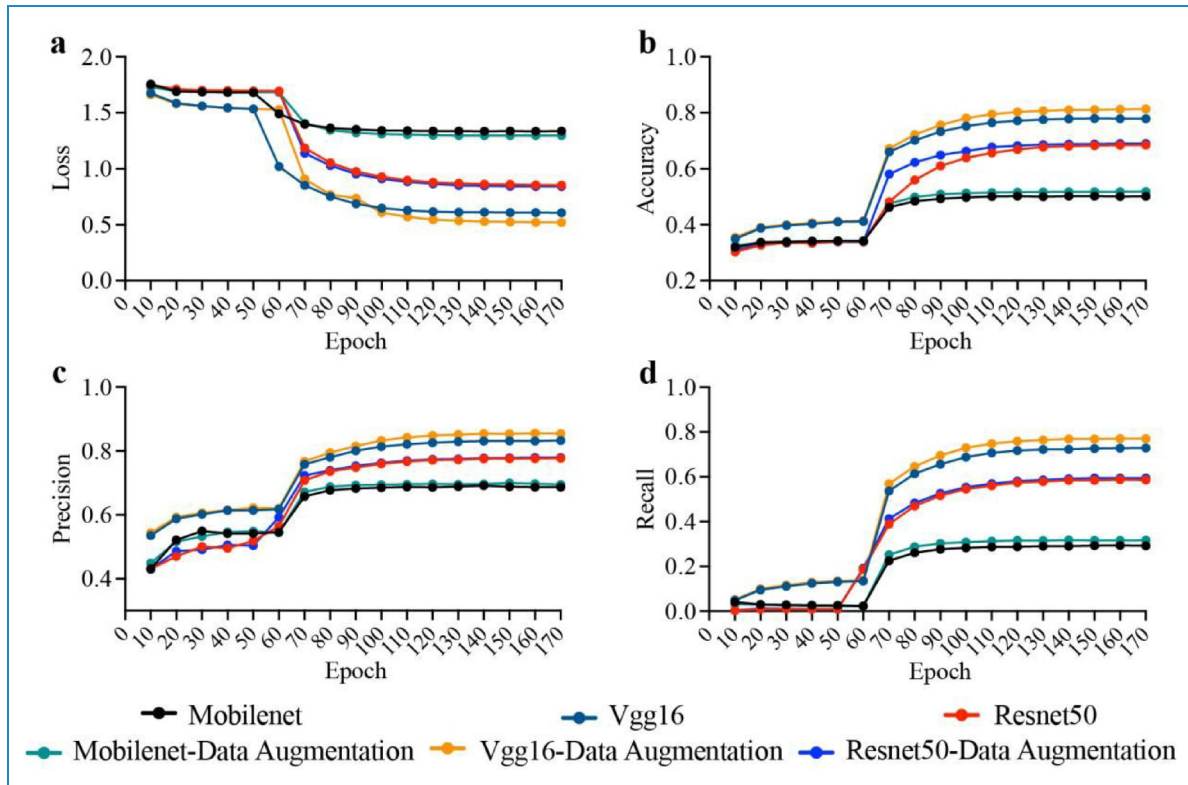


Figure 4. Comparison of training performance between MobileNet, VGG16, and Resnet50 data augmentation.

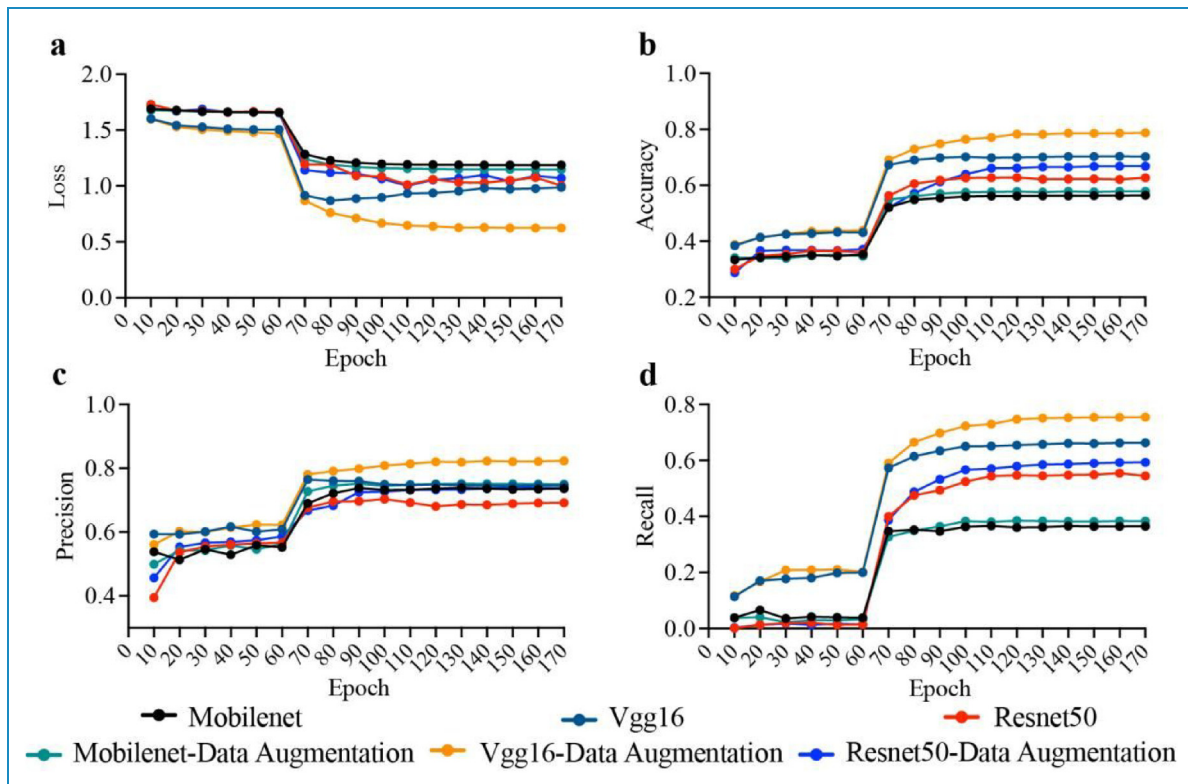


Figure 5. Comparison of validation performance between VGG16 and Resnet50 data augmentation.



Figure 6. Non-intrusive facial data collection.

Furthermore, the generalization capability of VGG16 significantly improved with the application of data augmentation, as evidenced by a notable increase in its accuracy to 0.7876, representing an enhancement of 0.853 compared to the baseline without data augmentation. In contrast, MobileNet and ResNet50 showed improvements of 0.136 and 0.418, respectively, suggesting that data augmentation was particularly effective in optimizing the VGG16 model.

Thus, VGG16 exhibited outstanding optimization capabilities on both the training and validation datasets. Compared to MobileNet, VGG16 offered higher accuracy and better generalization for capturing micro-expression features, making it particularly suitable for high-precision tasks. Despite the advantages of MobileNet in terms of computational efficiency, VGG16's superior performance makes it the preferred choice for this study.











预测时间	编号	姓名	单位	风险等级	操作
2024081909	77		某某科技	未见异常	风险报告
2024082109	77		某某科技	未见异常	风险报告
2024082113	77		某某科技	未见异常	风险报告
2024082213	77		某某科技	未见异常	风险报告
2024082309	77		某某科技	未见异常	风险报告
2024082310	77		某某科技	低风险	风险报告
2024082313	77		某某科技	未见异常	风险报告
2024081914	88		某某科技	未见异常	风险报告
2024082013	88		某某科技	未见异常	风险报告
2024082211	88		某某科技	低风险	风险报告

Figure 7. Real-time emotional risk classification interface.

Data collection and experimental results

This study implemented a robust facial data collection system using the OpenCV, RetinaFace, and Dlib modules. The system operated stably under various environmental conditions and achieved a capture rate of 24 fps. To optimize the efficiency, it skips every 12 frames, ensuring that two facial images are collected per second. Using RetinaFace for facial detection, the system accurately identifies and tracks facial features even under low light or crowded conditions. Then, the system utilizes Dlib to align images captured by RetinaFace, effectively correcting angle deviations. After alignment, `shape_predictor_68_face_landmarks` were applied to extract 68-dimensional facial feature vectors from the collected images. These feature vectors were then compared with prestored 68-dimensional facial templates and processed using `shape_predictor_68_face_landmarks` by calculating the cosine similarity between them. The system identifies the most similar facial template, and if the similarity score exceeds 0.95, the captured face is confirmed, allowing it to proceed to the subsequent micro-expression and anomaly detection analysis. The 0.95 threshold effectively filters out low-quality or misaligned images, ensuring that only accurately identified individuals enter the analysis phase. As shown in Figure 6, over a 1-month deployment in a corporate environment, the system maintained a stable capture rate above 70%, meeting daily attendance and identity recognition needs for a 10-person team, capturing facial data for at least seven individuals each day. Experimental results demonstrate that the system remains effective even in challenging scenarios.

To further enhance the practical applicability of the system, a real-time risk prediction feature was implemented. This feature enables hourly monitoring and classification of emotional states into different risk levels based on

collected facial data. For instance, the system evaluates individuals' emotional states in real-time and categorizes them as "No Risk," "Low Risk," or higher risk levels based on detected anomalies. Over a 1-month deployment, the system demonstrated its ability to provide timely alerts and actionable insights, helping organizations proactively identify potential emotional health issues. Figure 7 illustrates the interface that displays the hourly predictions and corresponding risk classifications.

Building on the system's ability to provide real-time risk predictions, further analyses were conducted to evaluate its performance in recognizing emotional anomalies across various scenarios. An analysis of the experimental results in the four scenarios (Figure 8(a)–(d)) revealed significant differences in the models' recognition of normal and abnormal emotions. The results indicate the high sensitivity and accuracy of the model in capturing anomalous emotional features. For instance, in abnormal emotional states, the probability distributions for categories such as anger, disgust, fear, and sad are notably concentrated and significantly higher than those in normal states. This demonstrates the precise detection of emotional anomalies by the model.

Individuals with depression exhibit a significantly higher probability of experiencing negative emotions, particularly sadness. Figure 9(a) shows elevated probabilities for fear and sadness in depressed individuals compared with non-depressed individuals, with sadness showing the most pronounced difference (indicated by stars). Figure 9(b) illustrates the daily variation in sadness probabilities for individuals with depression, peaking at approximately 10:00 AM.

These findings highlight pronounced negative emotions, especially sadness, in individuals with depression, providing strong evidence for the application of emotion detection technology to identify depressive states.

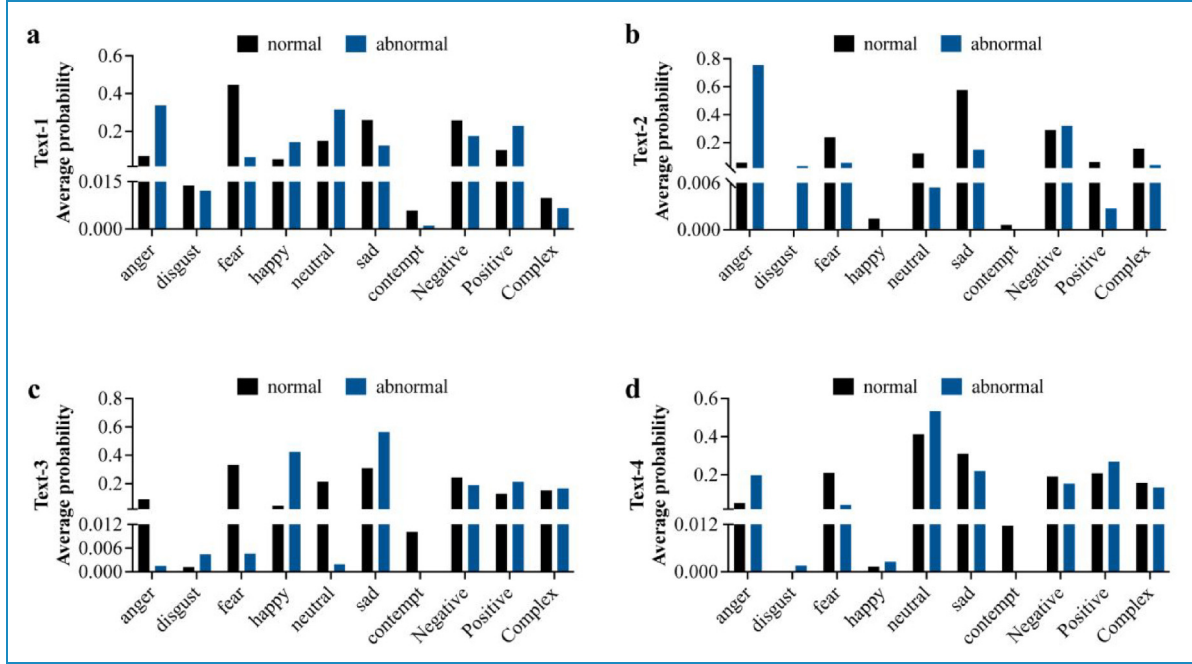


Figure 8. Experimental results of emotion recognition.

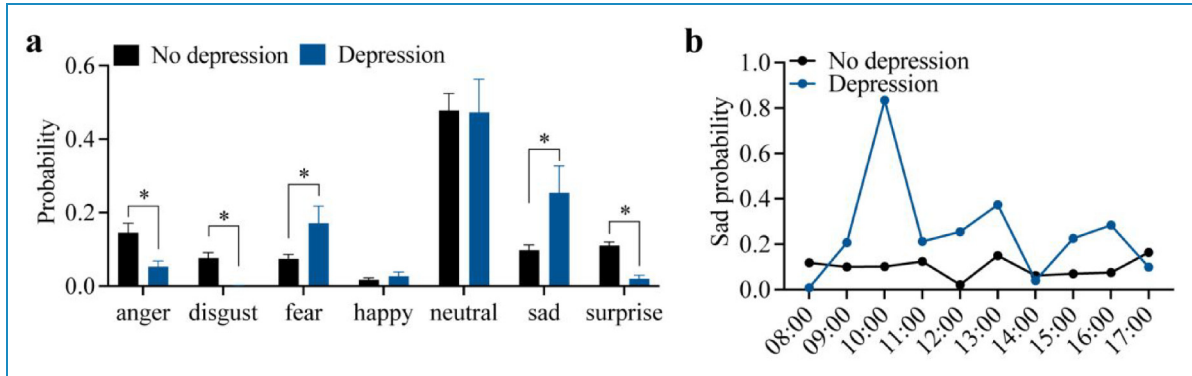


Figure 9. Comparison of emotional states between depressed and non-depressed individuals.

Discussion

This study introduces a novel approach to corporate mental health management that utilizes facial emotion recognition technologies. By integrating advanced deep learning models, including RetinaFace for face detection, Dlib for facial feature extraction, and VGG16 for micro-expression classification, we successfully constructed a 10-dimensional emotion feature vector to enhance emotional analysis. The results demonstrate that the system effectively captures employee emotional states and identifies anomalies in diverse environments, highlighting its suitability for handling complex emotion recognition tasks in corporate settings.

To address some of the challenges inherent to emotion recognition, data augmentation techniques have been

employed to mitigate the impact of lighting variations and other environmental factors, thereby improving the robustness of the system. However, this approach primarily improves general tolerance and does not fully address extreme or directional lighting conditions, which may still distort key facial features and affect classification accuracy.

This study has some other limitations, particularly regarding facial angles and cultural differences. Variations in facial angles and head poses may lead to partial occlusion or geometric distortion of micro-expressions. Although basic alignment was performed using Dlib to standardize face orientation, no further processing, such as 3D face modeling, was conducted. This may limit the system's effectiveness when users are not facing the camera directly, which is a common occurrence in real-world corporate

settings. Furthermore, cultural differences in emotional expression were not explicitly considered. The training dataset primarily comprises samples from a limited demographic, which may not capture variations in emotional display rules, expression intensity, or interpretation across cultures. As a result, the system's generalizability in multi-cultural workplaces may be affected.

Future research should explore targeted solutions to these issues. These may include incorporating illumination normalization, multi-angle training data, cross-cultural emotion datasets, and domain adaptation techniques. In addition, while the current system relies solely on facial emotion recognition, incorporating multimodal approaches, such as text-based sentiment analysis and voice-based emotion recognition, can enhance its robustness and overall performance. Expanding the dataset and addressing environmental and cultural variabilities would increase the applicability of the system across different work environments.

Conclusion

This study offers a promising solution for corporate mental health management by integrating facial emotion recognition technologies and leveraging advanced deep learning models, such as RetinaFace, Dlib, and VGG16. Our experimental results demonstrate that the proposed system effectively captures the emotional states of employees. This system also enables real-time monitoring and early detection of emotional anomalies, facilitating timely intervention and management in workplace settings. The use of the KNN algorithm and a 3σ -based risk evaluation method further enhances the system's ability to assign risk levels based on emotional data.

Despite its effectiveness, further research is necessary to improve the robustness of the system by addressing confounding factors such as facial angle variations and exploring the integration of multimodal emotion recognition approaches. Expanding the dataset and refining the system to handle diverse work environments and cultural differences will improve its practical applicability. Thus, this study lays the foundation for future innovations in emotion-based workplace support systems that contribute to enhanced employee well-being and productivity.


Acknowledgements

We would like to express our heartfelt gratitude to all individuals and organizations that contributed to the successful completion of this study. Special thanks go to Nanjing Yunshe Intelligent Technology Co., LTD for providing the necessary resources and technical support. We appreciate valuable feedback and insightful discussions from our colleagues, which significantly enhanced the quality of this study.

Research involving human participants and/or animals

Not applicable.

ORCID iD

Zicheng Zhang  <https://orcid.org/0000-0001-6081-4039>

Statements and declarations

Informed consent

The facial image collection (Figure 6) involved two authors, Zhang Zicheng and Yang Jie, both of whom agreed to the collection and publication of the images.

Author contributions/CRedit

ZZ designed the algorithms, proposed core ideas, and developed the project outline. TZ drafted and revised the manuscript. JY led software development and ensured the implementation and functionality of the proposed methodologies. All authors reviewed and approved the final version of the manuscript.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Natural Science Research Start-up Foundation of Recruiting Talents of Nanjing University of Posts and Telecommunications, (grant number NY223210).

Conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

1. Wainberg ML, Scorza P, Shultz JM, et al. Challenges and opportunities in global mental health: a research-to-practice perspective. *Curr Psychiatry Rep* 2017; 19: 28.
2. de Oliveira C, Saka M, Bone L, et al. The role of mental health on workplace productivity: a critical review of the literature. *Appl Health Econ Health Policy* 2023; 21: 167–193.
3. Neill RD, Lloyd K, Best P, et al. Understanding adolescent mental health and well-being: a qualitative study of school stakeholders' perspectives to inform intervention development. *SN Soc Sci* 2022; 2: 161.
4. Rothenberg WA, Bizzego A, Esposito G, et al. Predicting adolescent mental health outcomes across cultures: a machine learning approach. *J Youth Adolesc* 2023; 52: 1595–1619.
5. Reb J, Narayanan J, Chaturvedi S, et al. The mediating role of emotional exhaustion in the relationship of mindfulness with turnover intentions and job performance. *Mindfulness (N Y)* 2017; 8: 707–716.
6. Yang N and Qiu J. Text analysis and policy guidance of emotional intonation of enterprise management based on deep learning. *Comp Intell Neurosci* 2022; 2022: 3428078.
7. Suzuki K, Iguchi T, Nakagawa Y, et al. A multi-modal interaction robot based on emotion estimation method using physiological signals applied for elderly*. In: Proceedings of the 2023 32nd IEEE International Conference on Robot and Human Interactive Communication, 2023, pp. 2051–2057. Piscataway: IEEE.

8. Tanko D, Dogan S, Demir FB, et al. Shoelace pattern-based speech emotion recognition of the lecturers in distance education: shoePat23. *Appl Acoust* 2022; 190: 108637.
9. Li WB, Li GF, Tan RC, et al. Review and perspectives on human emotion for connected automated vehicles. *Automot Innov* 2024; 7: 4–44.
10. Caruelle D, Shams P, Gustafsson A, et al. Affective computing in marketing: practical implications and research opportunities afforded by emotionally intelligent machines. *Mark Lett* 2022; 33: 163–169.
11. Li HC, Pan T, Lee MH, et al. Make patient consultation warmer: a clinical application for speech emotion recognition. *Appl Sci* 2021; 11: 4782.
12. Li W, Zhang Y and Fu Y. Speech emotion recognition in e-learning system based on affective computing. 2007, pp. 809–813.
13. Tan L, Yu K, Lin L, et al. Speech emotion recognition enhanced traffic efficiency solution for autonomous vehicles in a 5G-enabled space-air-ground integrated intelligent transportation system. *IEEE Trans Intell Transp Syst* 2021; 23: 2830–2842.
14. Weismayer C and Pezenka I. Tracing emotional responses to nature-based tourism commercials and tagging individual sequences. *Tourism Recreat Res* 2024: 1–9.
15. Yun HI and Park JS. End-to-end emotional speech recognition using acoustic model adaptation based on knowledge distillation. *Multimedia Tool Appl* 2023; 82: 22759–22776.
16. Li L, Chen T, Ren F, et al. Bimodal emotion recognition method based on graph neural network and attention. *J Comput Appl* 2023; 43: 700–705.
17. Younis EMG, Mohsen S, Houssein EH, et al. Machine learning for human emotion recognition: a comprehensive review. *Neural Comput Appl* 2024; 36: 8901–8947.
18. Song P, Zheng W, Yu Y, et al. Speech emotion recognition based on robust discriminative sparse regression. *IEEE Trans Cogn Dev Syst* 2021; 13: 343–353.
19. Seehapoch T and Wongthanavas S. Speech emotion recognition using support vector machines. In: 5th International Conference on Knowledge and Smart Technology (KST), 2013, pp.86–91.
20. Aouani H, Ben Ayed Y. Deep support vector machines for speech emotion recognition. In: Abraham A, Siary P, Ma K, et al. (eds) *Intelligent systems design and applications. Adv Intell Syst Comput. ISDA 2019, vol 1181*. Cham: Springer, 2021, pp.406–415. DOI: 10.1007/978-3-030-49342-4_39.
21. Lin Y and Wei G. Speech emotion recognition based on HMM and SVM. In: International Conference on Machine Learning and Cybernetics, vols. 4898–4901, 2005.
22. Ntalampiras S. Toward language-agnostic speech emotion recognition. *J Aud Eng Soc* 2020; 68: 7–13.
23. Nam Y and Lee C. Cascaded convolutional neural network architecture for speech emotion recognition in noisy conditions. *Sensors (Basel)* 2021; 21: 4399.
24. Zhao J, Mao X and Chen L. Speech emotion recognition using deep 1D & 2D CNN LSTM networks. *Biomed Signal Process Control* 2019; 47: 312–323.
25. Atila O and Şengür A. Attention guided 3D CNN-LSTM model for accurate speech based emotion recognition. *Appl Acoust* 2021; 182: 108260. DOI: 10.1016/j.apacoust.2021.108260.
26. Zhang T and Wu J. Speech emotion recognition with i-vector feature and RNN model. In: IEEE China Summit and International Conference on Signal and Information Processing(ChinaSIP), 2015, pp.524–528.
27. Han J, Zhang A, Schmitt M, et al. From hard to soft: towards more humanlike emotion recognition by modelling the perception uncertainty. In: Proceedings of the 25th ACM international conference on multimedia, 2017, pp. 890–897.
28. Shahin I, Hindawi N, Nassif AB, et al. Novel dual-channel long short-term memory compressed capsule networks for emotion recognition. *Expert Syst Appl* 2022; 188: 116080.
29. Pereira R, Mendes C, Ribeiro J, et al. Systematic review of emotion detection with computer vision and deep learning. *Sensors (Basel)* 2024; 24: 3484.
30. Mohana M and Subashini P. Facial expression recognition using machine learning and deep learning techniques: a systematic review. *SN Comput Sci* 2024; 5: 432.
31. Taley SG and Pund MA. Physiological signals for emotion recognition. In: Yadav A, Nanda SJ and Lim MH (eds) *Proceedings of the international conference on paradigms of communication, computing and data analytics. PCCDA 2023. Algorithms for Intelligent Systems*. Singapore: Springer, 2023, pp.221–231. DOI: 10.1007/978-981-99-4626-6_18.
32. Deng J, Guo J, Ververas E, et al. Retinaface: single-shot multi-level face localisation in the wild. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 5202–5211. New York: IEEE.
33. Yan WJ, Wu Q, Liu YJ, et al. CASME Database: a dataset of spontaneous micro-expressions collected from neutralized faces. In: 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2013, pp. 1–7. Shanghai, China: IEEE.
34. Yu Y, Wang XM, Cen SX, et al. Spatiotemporal association and graph-attention guided micro-expression recognition network. *J Hebei Univ Technol* 2024; 53: 29–40.
35. Chen H, Gu Y, Wang F, et al. Facial expression recognition and positive emotion incentive system for human-robot interaction. In: Proceedings of the 13th world congress on intelligent control and automation, 2019, pp. 407–412. Changsha: IEEE.
36. Li JT, Wang T and Wang SJ. Facial micro-expression recognition based on deep local-holistic network. *Appl Sci* 2022; 12: 4643.
37. Howard AG, Zhu M, Chen B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861, 2017.
38. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778. New York: IEEE.