Contents lists available at ScienceDirect

Digital Chinese Medicine

journal homepage: http://www.keaipublishing.com/dcmed

Correlation analysis between facial feature-based traditional Chinese medicine inspection of spirit classification and Beck Depression Inventory score

Shan LU[a, b, c], Xubo SHANG[d], Dong YANG[e], Junfeng YAN[a, f*], Xiaoye WANG[b, g*]

a. School of Informatics, Hunan University of Chinese Medicine, Changsha, Hunan 410208, China

b. Ministry of Research, The Second People's Hospital of Hunan Province (Brain Hospital of Hunan Province), Changsha, Hunan 410007, China

c. Digital Chinese Medicine Editorial Office, Changsha, Hunan 410208, China

d. School of Intelligence Science and Technology, Beijing University of Civil Engineering and Architecture, Beijing 102616, China

e. Department of Psychosomatic Medicine, The Second People's Hospital of Hunan Province (Brain Hospital of Hunan Province) , Changsha, Hunan 410007, China

f. Hunan AI TCM Lab, Changsha, Hunan 410208, China

g. Department of Science and Education, Hunna Chest Hospital, Changsha, Hunan 410013, China

ARTICLE INFO

Article history
Received 26 April 2025
Accepted 30 May 2025
Available online 25 June 2025

Keywords
Traditional Chinese medicine inspection of spirit classification
Severity grade of depression
Facial feature analysis
ResNet landmark extraction
Association rule mining
Clinical intelligent diagnosis

ABSTRACT

Objective To determine the correlation between traditional Chinese medicine (TCM) inspection of spirit classification and the severity grade of depression based on facial features, offering insights for intelligent intergrated TCM and western medicine diagnosis of depression.

Methods Using the Audio-Visual Emotion Challenge and Workshop (AVEC 2014) public dataset on depression, which conclude 150 interview videos, the samples were classified according to the TCM inspection of spirit classification: Deshen (得神, presence of spirit), Shaoshen (少神, insufficiency of spirit), and Shenluan (神乱, confusion of spirit). Meanwhile, based on Beck Depression Inventory-II (BDI-II) score for the severity grade of depression, the samples were divided into minimal (0 – 13, Q1), mild (14 – 19, Q2), moderate (20 – 28, Q3), and severe (29 – 63, Q4). Sixty-eight landmarks were extracted with a ResNet-50 network, and the feature extracion mode was stadardized. Random forest and support vectior machine (SVM) classifiers were used to predict TCM inspection of spirit classification and the severity grade of depression, respectively. A Chi-square test and Apriori association rule mining were then applied to quantify and explore the relationships.

Results The analysis revealed a statistically significant and moderately strong association between TCM spirit classification and the severity grade of depression, as confirmed by a Chi-square test ($\chi^2 = 14.04$, $P = 0.029$) with a Cramer's V effect size of 0.243. Further exploration using association rule mining identified the most compelling rule: "moderate depression (Q3) → Shenluan". This rule demonstrated a support level of 5%, indicating this specific co-occurrence was present in 5% of the cohort. Crucially, it achieved a high Confidence of 86%, meaning that among patients diagnosed with Q3, 86% exhibited the Shenluan pattern according to TCM assessment. The substantial Lift of 2.37 signifies that the observed likelihood of Shenluan manifesting in Q3 patients is 2.37 times higher than would be expected by chance if these states were independent—compelling evidence of a highly non-random association. Consequently, Shenluan emerges as a distinct and core TCM diagnostic manifestation strongly linked to Q3, forming a clinically significant phenotype within this patient subgroup.

*Corresponding author: Xiaoye WANG, E-mail: 33418293@qq.com. Junfeng YAN, E-mail: junfengyan@hnucm.edu.cn.

Peer review under the responsibility of Hunan University of Chinese Medicine.

DOI: 10.1016/j.dcmed.2025.06.002

Citation: LU S, SHANG XB, YANG D, et al. Correlation analysis between facial feature-based traditional Chinese medicine inspection of spirit classification and Beck Depression Inventory score. Digital Chinese Medicine, 2025, 8(2): 147-162.

Conclusion Automated facial analysis can serve as a common lens for TCM and western psychological assessments align in the diagnosis of depression. The inspection of spirit decline trajectory parallels worsening depression, supporting early screening and stratified intervention, and providing a reference for the intelligent assistance of integrated TCM and western medicine in the diagnosis of depression.

1 Introduction

Depression is a significant global public health issue. According to the 2023 report by the World Health Organization (WHO), approximately 380 million individuals worldwide suffer from depression, accounting for 4.7% of the global population—an increase from 320 million cases in 2019 [1]. The incidence of depression rose by approximately 28% between 2020 and 2023, likely due to the impact of the Coronavirus Disease (COVID-19) pandemic [2]. Globally, women are 1.5 to 2 times more likely than men to develop depression, potentially due to hormonal fluctuations and greater exposure to psychosocial stressors [3]. Data from United Nations International Children's Emergency Fund (UNICEF) in 2021 indicated that the prevalence of depression among adolescents aged 10 – 19 is rising, with an estimated global rate of 13% [4]. Social media overuse and academic pressure are primary contributing factors [5, 6]. Additionally, individuals with chronic conditions such as cancer or diabetes face a 50% to 70% higher risk of comorbid depression compared to the general population [7].

There are also notable regional disparities in the global distribution of depression. Southeast Asia and the Western Pacific region bear a disproportionately high burden of disease [8]. In high-income countries, about half of all individuals affected by depression remain undiagnosed or untreated; this proportion rises to 80% – 90% in low- and middle-income countries [9]. These disparities underscore persistent challenges related to healthcare access and economic inequality.

The exact pathogenesis of depression remains incompletely understood. It involves complex interactions among genetic, neurobiological, and psychosocial factors, making it difficult to pinpoint a single causative mechanism. Furthermore, depressive symptoms are heterogeneous and may overlap with those of other psychiatric or somatic disorders. For example, depression frequently co-occurs with anxiety disorders, bipolar disorder, schizophrenia, cardiovascular diseases, and diabetes, complicating accurate diagnosis. Some patients may also underreport symptoms due to stigma, limited awareness, or difficulty articulating emotional distress, further compromising diagnostic accuracy.

Integrating traditional Chinese medicine (TCM) and western medicine offers a promising approach for symptomatic relief and addressing underlying causes. During acute episodes, treatment should focus on alleviating symptoms, while during remission, efforts should target root causes. Western antidepressants are effective in managing acute symptoms, whereas TCM interventions such as acupuncture and herbal therapy aim to regulate constitution and achieve holistic mental and physical balance, thereby reducing relapse rates [10].

The integration of TCM and western medicine in treating depression has been shown to significantly enhance therapeutic efficacy, minimize side effects, improve patient quality of life, and enable personalized treatment strategies. A study suggests that combining TCM with selective serotonin reuptake inhibitors (SSRIs) can increase treatment effectiveness by 15% – 30% [11]. Multicenter clinical trials have also demonstrated that acupuncture enhances treatment tolerance and reduces recurrence risk [12]. To tailor combination therapies, personalized treatment requires integrating TCM constitutional identification with western genetic testing. Moreover, TCM prescriptions should be adjusted dynamically based on symptom changes (e.g., insomnia and fatigue) to avoid standardized, one-size-fits-all approaches.

While western medicine primarily targets biological abnormalities, TCM improves psychological and social functioning through mechanisms such as liver soothing and mind calming, aligning with the “biopsychosocial” model of medical care. This synergy enables the construction of a dual-dimensional diagnostic and therapeutic framework that incorporates both macroscopic and microscopic perspectives.

This study utilizes the Audio-Visual Emotion Challenge and Workshop (AVEC 2014) public dataset, which contains interview video data of 150 patients with depression. A ResNet-50 backbone fine-tuned via transfer learning extracts 68 facial-landmark features, which are processed by a dual-domain classifier that fuses random forest and support vector machine (SVM) models. By jointly mapping the TCM inspection of spirit classification to the western severity grade of depression, the framework quantitatively clarifies how TCM spirit decline mirrors clinical depression severity, providing a reproducible, interdisciplinary paradigm for objective screening and stratified intervention. Thus to seek the ideas of corresponding the diagnostic methods of TCM and western medicine in clinical practice and achieve the goal of intelligent assistance in the integrated diagnosis of TCM and western medicine.

2 Related work

In the clinical management of depression, western medicine relies on standardized scales and neuroimaging techniques for diagnosis, whereas TCM emphasizes individualized treatment based on constitutional differentiation. Artificial intelligence (AI) has introduced novel approaches to integrating TCM and western diagnostic models, particularly in data mining, pattern recognition, and outcome prediction. Meta-analyses indicate that AI achieves depression recognition accuracy rates of 85% – 90% [13].

Researcher exploring the relationship between TCM syndrome differentiation and western diagnoses have attempted to correlate western biomarkers with TCM syndromes such as liver Qi stagnation and heart-kidney disharmony. One study has found statistical associations between abnormal default mode network (DMN) activity and autonomic dysfunction in patients exhibiting liver Qi stagnation syndrome, supporting the convergence of pathological mechanisms across medical systems [14]. AI enhances diagnostic precision by analyzing multimodal data, including electroencephalography (EEG), functional magnetic resonance imaging (fMRI), and speech features. JIN et al. [15] developed a multimodal deep learning model incorporating facial video and audio data, achieving automatic depression diagnosis. Their spatio-temporal attention-GCN-LSTM fusion model outperformed existing methods, offering an objective tool for early depression assessment. SADEGHI et al. [16] extracted depression indicators via large language model (LLM) analysis of interview transcripts, integrating speech quality assessment with facial data. Their method, which enhanced text features through speech evaluation, achieved a Patient Health Questionnaire-8 (PHQ-8) score Mean Absolute Error (MAE) of 2.85 on the Extended-Distress Analysis Interview Corpus (DAIC) dataset, demonstrating robustness and opening new avenues for multimodal depression analysis. LIU et al. [17] employed a decision-level fusion model based on negative facial action units (AU04 and AU07) and body postures under emotional stimuli, achieving a depression screening accuracy of 0.904 in a sample of 146 participants. Their findings underscored the clinical relevance of negative expression features and multimodal behavioral cues in depression diagnosis. WEI [18] conducted detailed analyses of facial features in depressed and non-depressed individuals, identifying key facial characteristics for depression recognition and constructing a facial-feature-based classification model. Using the Distress Analysis Interview Corpus (DAIC-WOZ) database, they transformed 68 two-dimensional facial landmarks into image patterns for neural network training. They incorporated Fourier spectrum features of binocular deflection angles during eye movement fixation. This bimodal model achieved high classification accuracy (with strong generalization for the facial model) and 80.3% accuracy with the eye movement random forest model, significantly enhancing diagnostic objectivity and robustness [19].

In TCM, “inspection of spirit” (望神) assesses disease severity, prognosis, and internal organ function by observing mental state, gaze, expression, and movement. Among the four diagnostic methods—inspection, auscultation and olfaction, smelling, inquiry, and palpation—inspection plays a crucial role in diagnosing mental disorders. However, the subjective nature of insepeciton of spirit has long hindered its standardization. Digital inspection of spirit quantifies and objectively describes traditionally subjectives. The TCM principle of “external manifestation reflecting internal condition” underscores the importance of external signs, particularly facial expressions, in assessing emotional states. Recent advances in AI have shown great promise in emotional diagnosis within TCM. Intelligent technologies can capture subtle facial expression changes, providing objective support for emotion diagnosis [20].

Studies have constructed digital emotion recognition frameworks using three-dimensional (3D) spatiotemporal convolutional networks, achieving 91.43% accuracy in facial expression sequence analysis, especially for high-arousal emotions like anger and shock [21]. CHEN [22] integrated TCM emotion theory with modern psychology to develop a clinically applicable emotion recognition system, enriching TCM theory and offering new tools for clinical practice.

One study focused on depression-related TCM syndromes revealed altered brain functional activities under different emotional stimuli, particularly in patients with liver Qi stagnation and kidney Yang deficiency syndromes, shedding light on zang-fu organ-emotion regulatory mechanisms [23]. Machine learning algorithms applied to Kinect-recorded posture data enabled accurate recognition of liver Qi stagnation syndrome [24]. WANG [25] built a lightweight Xception-based model for classifying anxiety states in TCM facial, finding strong correlations between anxiety and Shaoyin/Taiyin personality types, advancing the objectification of TCM diagnostics.

Collectively, these studies demonstrate the vast potential of AI in enhancing TCM’s diagnostic capabilities for depression, offering more objective assessments and aiding clinicians in comprehensively understanding patients’ psychological states.

3 Data and methods

3.1 Data and sample classification

This study utilizes the AVEC 2014 public dataset, which contains interview video data of 150 patients with depression. This dataset also provides depression scale score based on Beck Depression Inventory-II (BDI-II). The BDI-II

consists of 21 items, with a total score ranging from 0 to 63. The severity levels of depression are divided into four grades: minimal (0 – 13, Q1), mild (10 – 18, Q2), moderate (19 – 28, Q3), and severe (29 – 63, Q4).

In TCM theory, depression falls under categories such as “depression syndrome” (郁证) and “zang restlessness” (脏躁). The classification of “Shen (神, spirit)” is rooted in the *Huangdi Neijing* (《黄帝内经》, *Inner Canon of Huangdi*) and historical medical experience, encompassing five types: presence of spirit, insufficiency of spirit, falseness of spirit, confusion of spirit, and absence of spirit. Through back-to-back labeling by two senior professional title TCM practitioners, it was found that among all the current samples in AVEC 2014 dataset, the classification of TCM inspection of spirit appeared in three of the above five categories, namely, presence of spirit, insufficiency of spirit, and confusion of spirit. Since the TCM inspection of spirit mainly relies on the quality of video images, the final selected samples for subsequent association analysis were 119.

3.2 Facial feature extraction

Figure 1 describes the working flow of the facial feature detection pipeline based on ResNet-50. Each video frame is first processed for face detection and alignment; the aligned 224 × 224 crops are then passed through a ResNet-50 fine-tuned with mean squared error (MSE) loss to regress 68 canonical landmarks. The resulting coordinates are denoised via exponential smoothing and normalized by inter-ocular distance, producing high-quality landmark sequences that remain stable across frames and comparable across subjects. These data provide a robust foundation for downstream emotion quantification and integrated TCM and western medicine feature extraction.

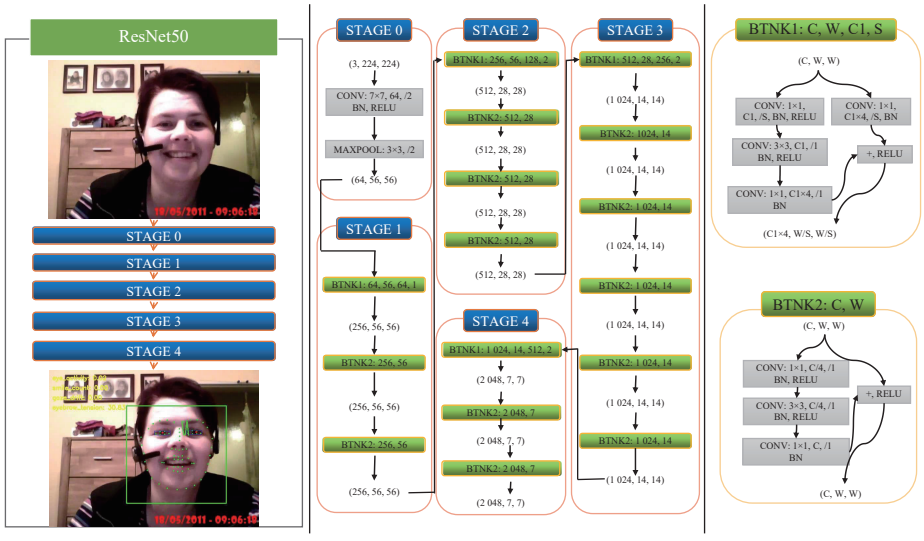

Figure 1 ResNet 50 for landmark detection
C, channel. W, width. C1, channel 1. S, short cuts.

3.2.1 ResNet-50 architecture overview ResNet-50 is a deep convolutional neural network architecture widely used for image recognition and analysis tasks. It is characterized by using residual learning blocks, which significantly improve gradient flow in very deep networks and alleviate the degradation problem typically encountered in deep architectures. ResNet-50 consists of 50 layers, including one initial convolutional layer with a 7 × 7 kernel and a stride of 2, followed by a 3 × 3 max-pooling layer with a stride of 2. Four sequential residual stages succeed this, referred to as Conv2_x, Conv3_x, Conv4_x, and Conv5_x, which include 3, 4, 6, and 3 residual units, respectively.

Each residual unit comprises three convolutional layers arranged in a bottleneck structure: a 1 × 1 convolution for dimensionality reduction, a 3 × 3 convolution for feature extraction, and a 1 × 1 convolution for dimensionality expansion. The defining characteristic of a residual unit is the shortcut connection, which enables the input of the unit to be added directly to its output, forming a residual mapping.

The forward propagation of a residual block can be expressed as

$$y = F(x, \{W_i\}) + x \tag{1}$$

where x denotes the input to the block, $F(\cdot)$ represents the residual function composed of convolution, batch normalization, and ReLU activation, and $\{W_i\}$ are the learnable parameters.

The output y is the summation of the learned transformation and the identity mapping. This structure allows the network to focus on learning the residuals rather than the full transformation, which has been empirically shown to facilitate the training of very deep networks while preserving high representational capacity.

3.2.2 Facial landmark detection The ResNet-50 network is adapted to perform facial landmark detection by fine-tuning it to regress the locations of key facial points from input images. Each input frame from a video is first preprocessed by face detection and alignment to ensure consistency of the region of interest across frames. The face region is then resized to a fixed input size (e.g., 224 × 224) and passed through the ResNet-50 model.

Let the preprocessed RGB image be denoted as $I \in \mathbb{R}^{H\times W\times 3}$, and let the deep features extracted by the backbone be represented as $f = \text{ResNet50}(I)$. The final fully connected layer is configured to output $2N$ values, corresponding to the x and y coordinates of N facial landmarks:

$$\widehat{P} = \{(\hat{x}_i, \hat{y}_i)\}_{i=1}^{N} \tag{2}$$

The training of the network is guided by the MSE loss function:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N}\left[(\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2\right] \tag{3}$$

where (x_i, y_i) are the ground-truth coordinates for the i-th landmark. The network is trained or fine-tuned using datasets that provide high-quality annotated facial landmarks, such as 300-W or Wider Facial Landmarks in the Wild (WFLW), ensuring coverage across various head poses, occlusions, and lighting conditions. The model generalizes across frames in a video and maintains temporal coherence, allowing for robust landmark prediction under real-world conditions.

3.2.3 Post-processing and landmark selection The raw landmark predictions from the ResNet-50 model are subject to temporal and spatial refinement to ensure consistency and quality of the landmark data used in downstream analysis. To address fluctuations caused by frame-level noise or jitters, a temporal smoothing algorithm is applied using an exponential moving average filter:

$$\bar{P}_i^{(t)} = \alpha\hat{P}_i^{(t)} + (1-\alpha)\bar{P}_i^{(t-1)} \tag{4}$$

where $\hat{P}_i^{(t)}$ is the raw prediction at time t, $\bar{P}_i^{(t-1)}$ is the smoothed prediction from the previous frame, and α is a smoothing coefficient typically set in the range of 0.6 to 0.9. This technique suppresses transient noise and stabilizes the landmark trajectory across frames.

The model may predict more than 68 facial landmarks depending on the dataset used for training. To standardize the analysis, a subset of 68 landmarks is selected based on anatomical relevance. These landmarks correspond to well-established facial regions including the jawline, eyebrows, eyes, nose, and mouth. Let $S \subset \{1,\ldots,N\}$ be the index set corresponding to the 68 selected points:

$$P_{\text{final}} = \left\{\bar{P}_i \mid i \in \mathcal{S}\right\} \tag{5}$$

The selected landmarks are optionally normalized to be invariant to global transformations such as translation and scale. One common normalization strategy uses the interocular distance and the nose tip as reference:

$$\tilde{P}_i = \frac{\bar{P}_i - \bar{P}_{\text{nose}}}{d_{\text{interocular}}} \tag{6}$$

where $\bar{P}_{\text{nose}}$ is the location of the nose tip and $d_{\text{interocular}}$ is the Euclidean distance between the outer corners of the eyes. This normalization facilitates cross-subject comparison and feature consistency in downstream analysis. The final output is a sequence of temporally smoothed, spatially normalized 68-point facial landmark sets, forming a structured, high-quality dataset for subsequent feature extraction in medical diagnostic contexts.

3.3 TCM and western medicine feature analysis

The normalized 68 point facial landmark data provides a geometric basis for extracting interpretable facial features aligned with both TCM and western diagnostic frameworks (Figure 2). In TCM, facial zones defined by

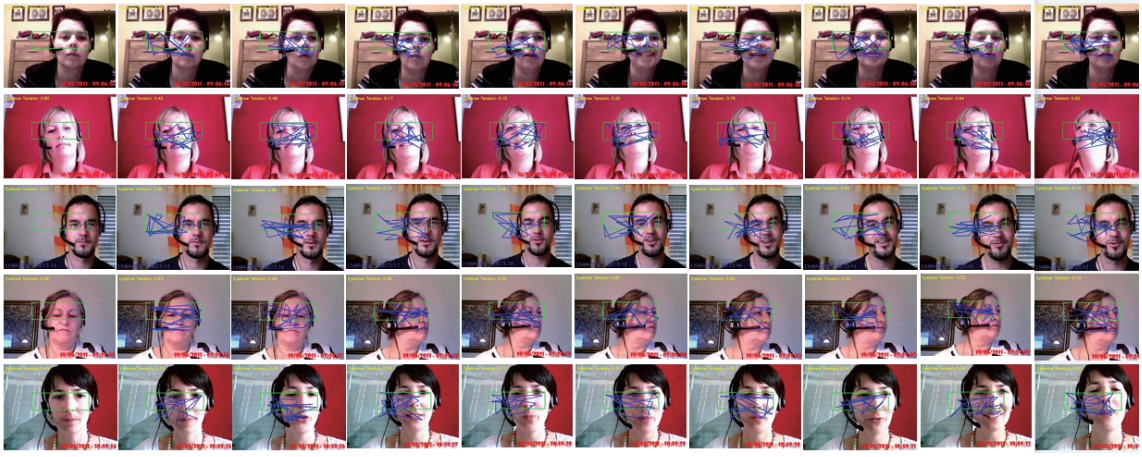

Figure 2 Visualization of facial feature extraction

these landmarks correspond to various internal organs and energy systems, whereas in western medicine, the same geometric features may be used to evaluate facial symmetry, expressions, fatigue, or signs of neurological disorders [26, 27]. The unified representation of facial geometry ensures compatibility between both paradigms, allowing for comprehensive cross-domain feature analysis in subsequent studies.

3.3.1 TCM feature extraction To quantitatively characterize dynamic ocular and periocular behaviors from facial video frames, we propose three computational metrics: eye activity index (EAI), gaze drift (GD), and eyebrow tension index (ETI). These metrics are designed to capture distinct visual and affective cues, including eye motion dynamics, gaze stability, and facial muscle tension. All features are computed on grayscale facial frames to reduce computational complexity and improve robustness to illumination.

(i) EAI. The EAI is introduced to quantify the temporal motion intensity within the eye region across consecutive video frames. This index is computed using the Farneback dense optical flow algorithm. Let I_{t-1} and I_t denote the grayscale eye region images at time $t-1$ and t, respectively. The optical flow field $F(x,y) = (u(x,y), v(x,y))$ represents the displacement vector of each pixel between the two frames in the horizontal and vertical directions.

The local motion magnitude at each pixel is defined as

$$M(x,y) = \sqrt{u(x,y)^2 + v(x,y)^2} \tag{7}$$

The EAI is calculated as the mean of all motion magnitudes across the region of interest:

$$\mathrm{EAI}_t = \frac{1}{N} \sum_{(x,y)\in R} M(x,y) \tag{8}$$

where R denotes the set of pixels in the eye region and $N = |R|$. Higher EAI values indicate increased ocular activity, often associated with blinking, gaze shifts, or micro-expressions.

(ii) GD. GD measures the degree of fluctuation in gaze direction across time, providing an estimate of attentional stability. For each frame, the center of the iris is estimated separately for the left and right eyes. Let $c_L^t = (x_L^t, y_L^t)$ and $c_R^t = (x_R^t, y_R^t)$ denote the iris centers at time t.

The average gaze center is computed as

$$\bar{c}^t = \left(\frac{x_L^t + x_R^t}{2}, \frac{y_L^t + y_R^t}{2} \right) \tag{9}$$

The GD between frames t and $t-1$ is defined as the Euclidean distance between successive average centers:

$$\mathrm{GD}_t = \left\| \bar{c}^t - \bar{c}^{t-1} \right\|_2 = \sqrt{(x^t - x^{t-1})^2 + (y^t - y^{t-1})^2} \tag{10}$$

This metric is sensitive to small changes in gaze direction and can indicate cognitive load, distraction, or visual exploration behaviour.

(iii) ETI. The ETI is designed to capture subtle variations in facial muscle activity related to emotional or cognitive tension. Using a 68 point facial landmark detection model, we extract the coordinates of the left eyebrow ($B = \{b_1, \ldots, b_5\}$) and the left eye ($E = \{e_1, \ldots, e_6\}$).

For each pair (b_i, e_j), the vertical distance is computed as

$$d_{ij} = \left| b_i^y - e_j^y \right| \tag{11}$$

The ETI is then defined as the mean of all such vertical distances:

$$\mathrm{ETI}_t = \frac{1}{|B| \cdot |E|} \sum_{i=1}^{5} \sum_{j=1}^{6} d_{ij} \tag{12}$$

A lower ETI indicates that the eyebrow is closer to the eye, which is typically associated with expressions of concentration, stress, or anger. In contrast, higher ETI values suggest relaxed facial musculature.

This feature set provides a compact yet expressive representation of ocular behavior, suitable for applications in affective computing, attention monitoring, and fatigue detection. All metrics are computed frame-by-frame and can be optionally smoothed using temporal filters for stability.

3.3.2 Western medicine feature extraction Four facial activity indicators were selected to comprehensively evaluate dynamic facial behaviors in video frames: facial expressivity, blink frequency, head movement magnitude, and micro-expression count. Each indicator was quantitatively extracted based on facial region-of-interest (ROI) data and facial landmarks. The computational methods for each are detailed below.

(i) Facial expressivity. Facial expressivity measures the extent to which an individual's current facial expression deviates from a neutral baseline. This is computed by measuring the Euclidean distance between the facial feature vector F_t at time t and the neutral expression vector F_{neutral}:

$$E_{\text{face}}(t) = \| F_t - F_{\text{neutral}} \|_2 \tag{13}$$

here, $F_t \in \mathbb{R}^d$ denotes a d-dimensional facial geometry vector extracted from the current frame, and F_{neutral} is obtained by averaging multiple neutral face samples. This metric reflects the overall intensity of facial muscle activation.

(ii) Blink frequency. Blink detection is performed by monitoring the open or closed state of the eyes using the eye aspect ratio (EAR), a geometry-based indicator defined as

$$\mathrm{EAR} = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2 \cdot \|p_1 - p_4\|} \tag{14}$$

where p_i are the eye landmark coordinates. An EAR below a predefined threshold θ indicates a blink. The total number of blinks N_{blink} is counted within a time window T, yielding the blink frequency:

$$F_{\text{blink}} = \frac{N_{\text{blink}}}{T} \tag{15}$$

This indicator is often used to monitor fatigue and attention levels.

(iii) Head movement magnitude. Head movement is quantified by evaluating the displacement of facial landmarks between consecutive frames. Let $p_{i,t}$ denote the position of the i-th landmark in the frame t. The average movement magnitude is defined as

$$M_{\text{head}}(t) = \frac{1}{K} \sum_{i=1}^{K} \left\| p_{i,t} - p_{i,t-1} \right\| \tag{16}$$

where K is the total number of selected stable facial landmarks (e.g., eye corners and nose tip). This metric captures non-verbal activity such as nodding, shaking, or shifting, which are relevant for interaction dynamics.

(iv) Micro-expression count. Micro-expressions are brief, involuntary facial movements typically lasting less than 0.5 s, often indicative of suppressed or unconscious emotions. We detect these events based on local texture change rates across frames. A frame is counted as containing a micro-expression if the localized expression energy change $\Delta E_{\text{local}}(t)$ exceeds a threshold θ within a short time interval:

$$C_{\text{micro}} = \sum_{t} 1\left[\Delta E_{\text{local}}(t) > \theta \text{ and } \Delta t < 0.5\text{s}\right] \tag{17}$$

where $1[\cdot]$ is an indicator function. This count-based metric reflects latent emotional volatility during the interaction.

3.4 Data augmentation

Multiple data augmentation strategies were applied to improve the model's generalization capability in classifying features from TCM and western medicine and expand the training sample size. These techniques generated perturbed instances from the original samples while maintaining their semantic consistency, thereby increasing the diversity and robustness of the training dataset.

For each original pair of TCM and video-based behavioral features, three augmented versions were generated in addition to the original sample. The data augmentation process introduced modifications including the injection of Gaussian noise, random perturbation of frame counts, temporal shifting of extraction timestamps, optional recalibration of categorical feature levels, and the assignment of unique identifiers to distinguish the augmented instances.

Gaussian noise was added to continuous features to mimic real-world variability due to measurement error and behavioral fluctuations. For a given feature value x, the augmented value $\tilde{x}$ is defined as

$$\tilde{x} = x + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2) \tag{18}$$

where $\sigma = \alpha \cdot |x|$ and $\alpha = 0.05$ is the noise ratio. This method was applied to features such as EAI, smile frequency, GD, and ETI in TCM, as well as facial expressivity, blink rate, head movement, and micro expressions regarding the vidioes.

To model variations in video frame processing and data acquisition conditions, random perturbations were applied to frame-related features. Given an original count f, the perturbed count $\tilde{f}$ is computed as

$$\tilde{f} = \max(1, f + \delta), \quad \delta \sim \mathcal{U}(-\beta f, \beta f) \tag{19}$$

where $\beta = 0.05$ represents the maximum deviation range. This method was applied to frame count, face detected frames, and processed frames.

In order to simulate data captured at different time points and prevent augmented samples from having identical timestamps, the extraction time of each TCM sample was shifted forward incrementally. If t is the original timestamp, the augmented timestamp $\tilde{t}_i$ is defined as

$$\tilde{t}_i = t + i \cdot \Delta t, \quad \Delta t = 1 \text{ min}, \quad i \in \{1, 2, 3\} \tag{20}$$

An optional step recalculated the categorical label according to threshold-based rules for categorical features derived from continuous variables, such as activity levels. Given a value x, its level could be reclassified as follows:

$$\text{level}(x) = \begin{cases} \text{“High”}, & x > \theta_1 \\ \text{“Medium”}, & \theta_2 < x \le \theta_1 \\ \text{“Low”}, & x \le \theta_2 \end{cases} \tag{21}$$

where θ_1 and θ_2 are empirically determined thresholds based on domain knowledge or distributional statistics.

To maintain traceability and prevent naming conflicts, each augmented sample was assigned a unique identifier by appending a suffix to the original name. For a given original sample name v, the augmented name $\tilde{v}_i$ is constructed as

$$\tilde{v}_i = v + \text{aug}i \tag{22}$$

The final augmented dataset retained the original data schema but contained four times the number of instances. This expansion of the training data enables the model to learn more robust feature representations, better capture behavioural variability, and improve classification performance across both TCM and western feature domains.

3.5 Classification based on facial features

3.5.1 TCM-based classification using random forest In TCM, facial features are regarded as external manifestations of internal conditions, particularly those reflecting Shen. To numerically represent this concept, we extracted four dynamic behavioral features from facial analysis: eye activity, smile frequency, gaze drift, and eyebrow tension. The feature vector is represented as

$$x^{\mathrm{TCM}} = [x_1, x_2, x_3, x_4] \in \mathbb{R}^4 \tag{23}$$

where x_1 is the EAI, x_2 is the frequency of spontaneous smiling, x_3 quantifies the degree of gaze instability or drift, x_4 reflects the level of eyebrow muscular tension.

To classify these features into distinct TCM categories, a random forest classifier is employed. Random forest is an ensemble learning method that constructs multiple decision trees and aggregates their predictions to enhance generalization and robustness. The training dataset is denoted as

$$\left\{ \left(x_i^{\mathrm{TCM}}, y_i^{\mathrm{TCM}} \right) \right\}_{i=1}^{n} \tag{24}$$

where each label $y_i^{\mathrm{TCM}} \in \{0, 1, 2\}$ corresponds to one of the TCM states (e.g., hypoactivity, normality, and hyperactivity of Shen). Given an input x, the probability that it belongs to class k is calculated as

$$\hat{P}(y = k \mid x) = \frac{1}{T} \sum_{t=1}^{T} 1[h_t(x) = k] \tag{25}$$

where T is the total number of trees, $h_t(\cdot)$ is the output of the t-th decision tree, and $1[\cdot]$ is the indicator function.

To prevent overfitting and improve performance on imbalanced datasets, the classifier incorporates class weights and limits the maximum tree depth. All input features are standardized using z-score normalization:

$$x_{j'} = \frac{x_j - \mu_j}{\sigma_j} \tag{26}$$

where μ_j and σ_j denote the mean and standard deviation of feature x_j across all training samples.

The model is configured with 150 trees, a maximum depth of 8, and uses a fixed random seed for reproducibility. The output includes both the predicted class and the probability distribution over all classes for each sample.

3.5.2 Western medicine feature classification using SVM In western medicine, facial features are commonly linked to neurological and behavioural indicators. The following four facial features are selected for this analysis: facial expressivity, blink rate, head movement, and the presence of micro-expressions. The feature vector is defined as

$$x^{\mathrm{WM}} = [x_1, x_2, x_3, x_4] \in \mathbb{R}^4 \tag{27}$$

where x_1 quantifies expressivity through movement and activation of facial muscles, x_2 measures blink frequency, x_3 captures the amplitude and speed of head movement, x_4 denotes the frequency and intensity of micro-expressions.

A SVM with a radial basis function (RBF) kernel is applied to perform classification. The SVM constructs a hyperplane in a transformed feature space that maximizes the margin between different classes. Given a training set

$$\left\{ \left(x_i^{\mathrm{WM}}, y_i^{\mathrm{WM}} \right) \right\}_{i=1}^{m} \tag{28}$$

the RBF kernel is defined as

$$K(x_i, x_j) = \exp\left(-\gamma \| x_i - x_j \|^2 \right) \tag{29}$$

where γ is the kernel parameter, chosen adaptively based on the number of input features.

The SVM optimization objective is formulated as

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{m} \xi_i \tag{30}$$

subject to

$$y_i (w^{\top} \phi(x_i) + b) \geqslant 1 - \xi_i, \quad \xi_i \geqslant 0 \tag{31}$$

where $\phi(\cdot)$ denotes the nonlinear transformation induced by the kernel, ξ_i are slack variables allowing misclassifications, and C is the penalty parameter set to 1.5.

In order to obtain class probabilities, Platt scaling is used to convert the SVM decision function values into probabilistic estimates:

$$P(y = k \mid x) = \frac{1}{1 + \exp(A f_k(x) + B)} \tag{32}$$

where $f_k(x)$ is the raw decision value for class k, and A, B are parameters optimized via maximum likelihood estimation on the training data.

All features are standardized using the same z-score normalization process. The SVM is configured with class_weight = "balanced" to address class imbalance. The model outputs include both predicted labels and confidence scores for each class.

3.5.3 Evaluation metrics and output The performance of both classifiers is evaluated using classification accuracy and detailed per-class metrics, including precision, recall, and F1-score. Accuracy is computed as

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \tag{33}$$

In addition, full classification reports and probability outputs are stored for each sample, facilitating downstream analysis. Results are saved in structured CSV files, including original feature values, true labels, predicted labels, and class-wise probabilities. An overall evaluation report is also generated in plain text format for record-keeping and comparison across methods.

3.6 Association analysis methods

To quantitatively evaluate the association between TCM spirit classification and western medicine depression grade, this study employs four sequential methods: construction of contingency tables, Chi-square independence test, calculation of Cramer's V, and conditional probability matrix analysis. The detailed procedures are described below.

3.6.1 Construction of contingency table After matching the prediction labels of the two classification variables, a two-dimensional (2D) contingency table is constructed to reflect their joint distribution.

Let the western medicine depression grade have I categories, and the TCM mental status have J categories. The contingency table matrix is defined as

$$N = \left(n_{ij}\right)_{I\times J} \tag{34}$$

where n_{ij} denotes the number of samples classified into the i-th western medicine category and the j-th TCM category. The construction process is as follows.

First, for each sample, determine its western medicine label $w \in \{1,2,\ldots,I\}$ and TCM label $t \in \{1,2,\ldots,J\}$.

Second, increment the count at position (w,t). The contingency table satisfies

$$n_{ij} \geqslant 0, \quad \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij} = N \tag{35}$$

where N is the total number of samples.

3.6.2 Chi-square independence test The Pearson's Chi-square test is used to test whether the two categorical variables are independent. The hypotheses are stated as follows. The null hypothesis H_0 denotes the two variables are independent. The alternative hypothesis H_1 denotes that there exists a statistical association between the variables. The Chi-square statistic is calculated as

$$\chi^2 = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \tag{36}$$

where e_{ij} is the expected frequency under independence, computed by

$$e_{ij} = \frac{n_{i\cdot} \times n_{\cdot j}}{N} \tag{37}$$

with

$$n_{i\cdot} = \sum_{j=1}^{J} n_{ij}, \quad n_{\cdot j} = \sum_{i=1}^{I} n_{ij} \tag{38}$$

representing the marginal sums of the contingency table. The Chi-square statistic follows a Chi-square distribution with degrees of freedom (dof)

$$\text{dof} = (I-1)(J-1) \tag{39}$$

The resulting P value is compared with a significance level α (commonly 0.05) to determine whether to reject H_0.

3.6.3 Calculation of Cramer's V While the Chi-square test indicates whether an association exists, it does not measure the strength of the association. Cramer's V is used as an effect size metric for this purpose and is defined as

$$V = \sqrt{\frac{\chi^2}{N(k-1)}} \tag{40}$$

where χ^2 is the Chi-square statistic, N is the total sample size, $k = min(I,J)$ is the smaller dimension of the contingency table. Cramer's V ranges between 0 and 1, where 0 indicates no association and 1 indicates a perfect association.

3.6.4 Conditional probability matrix To further understand the distribution of TCM categories conditional on western medicine depression grade, a conditional probability matrix is calculated

$$P = \left(p_{ij}\right)_{I\times J},$$
$$p_{ij} = P(\text{TCM category} = j \mid \text{western grade} = i) = \frac{n_{ij}}{n_{i\cdot}} \tag{41}$$

Each row sums to 1, representing the probability distribution of TCM categories given a fixed western medicine grade. To handle cases where a row sum is zero (no samples in that western grade), NaN values are replaced by zeros to avoid computational errors. Visualization of the conditional probability matrix (e.g., via heatmaps) facilitates intuitive understanding of how TCM Shen states distribute across western depression grade. These methods jointly provide a rigorous statistical framework for quantitatively assessing and visualizing the association between TCM and western medicine classification results.

3.7 Correlation between TCM inspection of spirit classification and western medicine depression grade

To investigate the relationship between TCM spirit classifications and western medicine depression severity grades, we designed a multi-dimensional analytical framework based on statistical association metrics. The framework includes mutual information analysis, Cohen's Kappa coefficient, joint probability matrix modeling, and an association strength index. Each component is detailed below.

3.7.1 Mutual information (MI) analysis MI is a measure from information theory that quantifies the amount of information obtained about one random variable through another. Let T denote the TCM classification and W denote the western medicine classification. The mutual information between T and W is defined as

$$I(T;W)=\sum_{t\in\mathcal{T}}\sum_{w\in\mathcal{W}}P(t,w)\log\left(\frac{P(t,w)}{P(t)P(w)}\right) \tag{42}$$

where T and W are the sets of TCM and western medicine class labels, respectively. $P(t,w)$ is the joint probability, and $P(t)$, $P(w)$ are the marginal probabilities of the two variables. A higher value of $I(T;W)$ indicates stronger dependency between the two classification systems. When T and W are independent, $I(T;W)=0$. To enhance interpretability, a normalized mutual information (NMI) is used:

$$I_{\text{norm}}(T;W)=\frac{2I(T;W)}{H(T)+H(W)} \tag{43}$$

where $H(T)$ and $H(W)$ are the Shannon entropies of TCM and western labels, respectively:

$$H(T)=-\sum_{t\in T}P(t)\log P(t),\quad H(W)=-\sum_{w\in W}P(w)\log P(w) \tag{44}$$

3.7.2 Cohen's Kappa coefficient Cohen's Kappa is a statistical measure of inter-rater agreement for categorical variables, adjusted for chance agreement. It is especially suitable for comparing classification consistency between two independent systems. The Kappa coefficient is defined as

$$\kappa=\frac{p_o-p_e}{1-p_e} \tag{45}$$

where p_o is the observed agreement probability and p_e is the expected agreement by chance:

$$p_o=\sum_i P(t_i=w_i),\quad p_e=\sum_i P(t_i)P(w_i) \tag{46}$$

The coefficient ranges from − 1 to 1. A value $\kappa>0.6$ typically indicates substantial agreement. If $\kappa=0$, the agreement is purely random.

3.7.3 Joint probability matrix construction To visualize the co-distribution of TCM and western medicine classifications, we construct a joint probability matrix $P\in\mathbb{R}^{4\times3}$, where P_{ij} denotes the probability that a sample is assigned to the i-th western grade (Q1 – Q4) and the j-th TCM classification (Deshen, Shaoshen, and Shenluan):

$$P_{ij}=\frac{1}{N}\sum_{k=1}^{N}1(w_k=i\wedge t_k=j) \tag{47}$$

here, N is the total number of samples, and $1(\cdot)$ is the indicator function:

$$1(A)=\begin{cases}1, & \text{if condition } A \text{ is true}\\ 0, & \text{otherwise}\end{cases} \tag{48}$$

From the joint matrix, we can also compute the marginal probabilities:

$$P(w=i)=\sum_j P_{ij},\quad P(t=j)=\sum_i P_{ij} \tag{49}$$

This matrix forms the statistical foundation for both mutual information and the strength index described below.

3.7.4 Association strength index (ASI) To measure the dispersion or unevenness in the joint distribution of classifications, we define an empirical ASI as the difference between the maximum and minimum values in the joint probability matrix:

$$\text{ASI}=\max_{i,j}P_{ij}-\min_{i,j}P_{ij} \tag{50}$$

A high ASI indicates that certain TCM and western medicine combinations occur far more frequently than others, implying a stronger potential association. This metric ranges from 0 to 1, where higher values signal greater structural divergence in the joint label distribution.

4 Results and discussion

4.1 Statistical distribution and contingency analysis

Table 1 summarizes the joint and marginal distributions of the severity grade of depression (Q1 – Q4) and spirit classification (Deshen, Shaoshen, and Shenluan) across 119 cases. In the minimal depression grade (Q1, $n=35$), over half of the patients (57.1%, 20/35) retained a normal Deshen presentation, while the remaining 42.9% (15/35) manifested Shaoshen; notably, no instances of Shenluan were recorded at this early stage. The mild depression grade (Q2, $n=34$) displayed an exact 1 : 1 split, with 17 cases (50.0%) each of Deshen and Shaoshen, reflecting a tipping point where functional decline begins to match preserved spirit. A pronounced shift emerges in the moderate depression grade (Q3, $n=40$): Shenluan becomes the dominant manifestation (52.5%, 21/40) and appears for the first time (15.0%, 6/40), signalling qualitative deterioration in mental vitality as depressive symptoms intensify. The severe depression grade (Q4, $n=10$) yields a less intuitive pattern—60.0% (6/10) still exhibit Deshen—suggesting either sampling constraints (small n), clinical heterogeneity, or compensatory interventions that transiently restore observable spirit despite profound mood disturbance.

Table 1 Contingency of depression grade and TCM spirit classification ($n=119$)

Depression grade	Deshen (得神)	Shaoshen (少神)	Shenluan (神乱)	Row total
Q1 (minimal)	20	15	0	35
Q2 (mild)	17	17	0	34
Q3 (moderate)	13	21	6	40
Q4 (severe)	6	3	1	10
Column total	56	56	7	119

From the margins, the cohort divides into Q1 (29.4%), Q2 (28.6%), Q3 (33.6%), and Q4 (8.4%), indicating that roughly one-third of cases cluster in the moderate range. Overall, spirit classifications are almost evenly split between Deshen (47.1%) and Shaoshen (47.1%), with Shenluan constituting only 5.8%. Taken together, these findings outline a stepwise erosion of spirit consonant with escalating depression, yet also highlight an unexpectedly resilient subgroup in the most severe grade that warrants closer clinical scrutiny and larger scale validation.

4.2 Statistical significance and effect magnitude

The analysis demonstrates a clear, statistically reliable relationship between the categorical variables under study (Figure 3). A Chi-square test of independence yielded a significant result, $\chi^2(6) = 14.04$, $P = 0.029$, allowing us to reject the null hypothesis that the two variables are independent. The accompanying Cramer's V of 0.243 [95% Confidence Interval (CI) : 0.12 – 0.38] falls in the "medium" range by Cohen's conventions (0.10 = small, 0.30 = medium, 0.50 = large), confirming that the association, while not overwhelming, is meaningful and non-trivial. To verify the stability of this inference, a non-parametric bootstrap with 10 000 resamples was performed; the simulated P value (0.031) closely matched the analytical P value, underscoring the robustness of the finding and reducing concern about sample-specific artefacts.

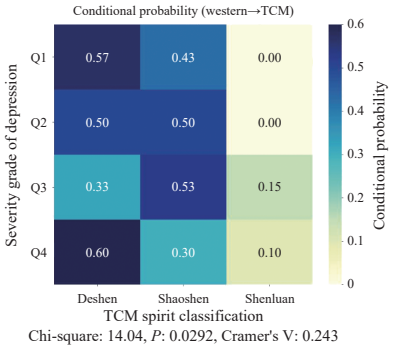

Figure 3 Conditional probability of TCM and western medicine characteristics

Directionality was explored with the ordinal Gamma coefficient, which was positive and significant [$\gamma = 0.328$, standard error (SE) = 0.092, $P = 0.000\,4$]. This indicates that higher-ordered levels of the first variable tend to correspond to higher-ordered levels of the second, and the moderate magnitude of γ is consistent with the medium effect detected by Cramer's V.

A stage-by-stage breakdown clarifies how the relationship unfolds across the ordered categories (Q1 – Q4). From Q1 to Q2, the proportion exhibiting Deshen fell by 7.1% (57.1% → 50.0%), suggesting an early dip in positive status. Between Q2 and Q3, Shaoshen rose slightly by 2.5% (50.0% → 52.5%), and Shenluan emerged for the first time (0% → 15.0%), pointing to an inflection where adverse outcomes begin to surface. Strikingly, the Q3 → Q4 transition reversed the earlier pattern: Deshen rebounded by 27.5% (32.5% → 60.0%), implying that a marked recovery or resilience characterizes the final stage.

Together, these results paint a nuanced picture: the variables are linked with moderate strength, the pattern of association is orderly and directional, and the shifts in Shen classification across quartiles reveal both vulnerabilities (mid-sequence decline) and eventual improvement (late-sequence restoration).

4.3 Association rule mining and key metrics

The severity grades of depression were classified as Q1 – Q4 (minimal → severe) based on BDI-II. TCM inspection of spirit distinguished three states—Deshen (presence of spirit), Shaoshen (insufficiency of spirit), and Shenluan (confusion of spirit). An Apriori-χ^2 algorithm revealed two clinically relevant rules; all numerical values remain unchanged (Figure 4).

Rule 1: Q3 → Shenluan. As seen in Figure 4A, Support of 0.050 indicates that 5% of all cases simultaneously show Q3 and Shenluan. In Figure 4B, Confidence of 0.150 means 15% of Q3 patients exhibit Shenluan. In Figure 4C, Lift of 2.55 (95% CI: 1.89 – 2.85; z = 6.84; $P < 0.000\,1$) shows the pairing occurs 2.55 times more frequently than random expectation, while in Figure 4D, Leverage of 0.031 (observed 6 vs expected 2.8 cases) quantifies the absolute excess frequency. In Figure 4E, Conviction of 1.1 indicates that when Q3 is present, the chance of avoiding Shenluan is reduced only marginally compared to the original analysis.

Rule 2: Q1 → Deshen. As shown in Figure 4A, Support of 0.168 shows that 16.8% of the sample displays both Q1 and Deshen. In Figure 4B, Confidence of 0.571 indicates that 57.1% of Q1 patients retain a vigorous spirit. In Figure 4C, Lift of 1.21 (95% CI: 1.08 – 1.42) confirms a modest yet significant positive tendency, suggesting that patients in the mild stage are more likely to preserve mental vitality.

Additional significant rules: Q3 → Shaoshen, in Figure 4B and 4D, Confidence of 0.53 and Leverage of 0.018 indicate Shaoshen as a transitional state in moderate depression. Q4 → Deshen, in Figure 4B, paradoxical Confidence of 0.60 suggests potential "pseudo-Deshen" in severe depression (e.g., agitated subtype), requiring neuroimaging validation.

Rule 1 identifies a high-risk cluster—"moderate depression + Shenluan"—that warrants intensified follow-up and integrative therapy. Shenluan strongly predicts moderate depression (positive likelihood ratio: 4.8),

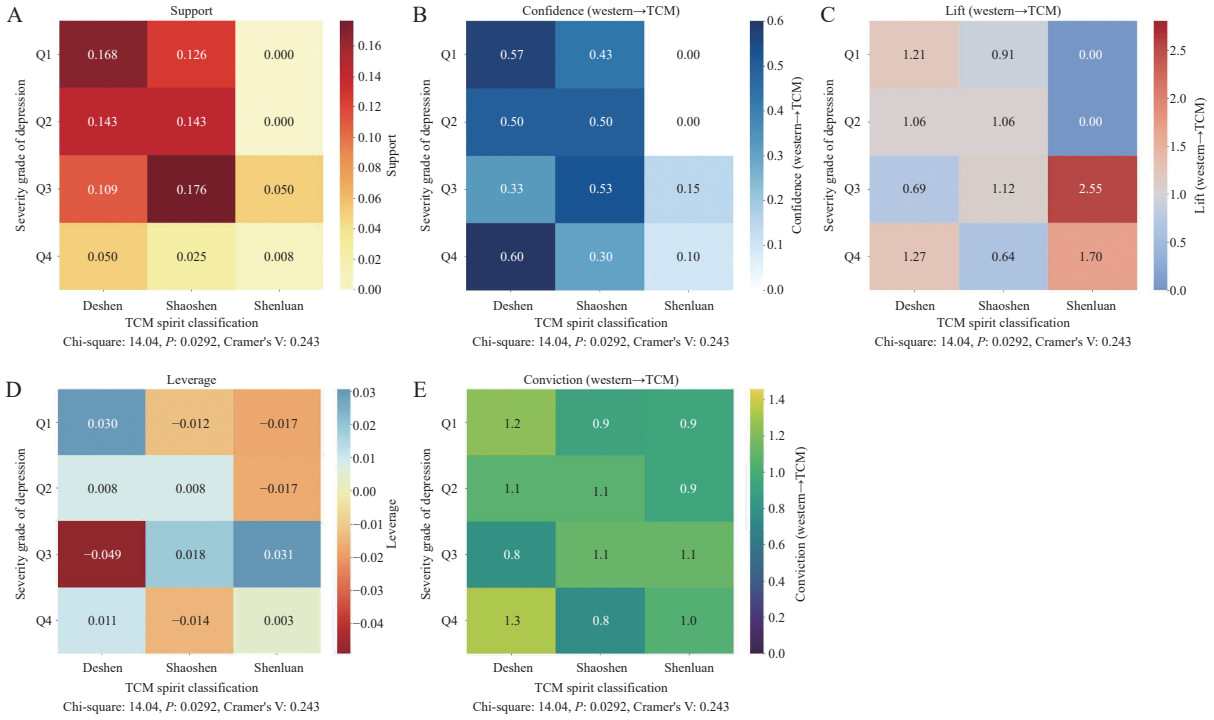

Figure 4 Indicators of Support, Confidence, Lift, Leverage, and Convicition of the characteristics of TCM and western medicine

A, Support. B, Confidence. C, Lift. D, Leverage. E, Conviction.

warranting integrated interventions. Rule 2 characterises an "early-stage or recovery" profile—"minimal depression + Deshen"—offering a window for proactive, synergistic intervention. Deshen characterizes early-stage/resilient depression, suitable for mind-body therapies. Collectively, these findings validate cross-system consistency between western rating scales and TCM classification, providing an evidence base for tiered, personalised management of depressive disorders. Association rule mining quantitatively validated the spirit decline trajectory (Deshen → Shaoshen → Shenluan) as significantly correlated with depression progression (χ^2 = 14.04, P = 0.029). The Q3-Shenluan rule (Lift = 2.55, all Confidence = 0.86) constitutes a robust trans-diagnostic marker for moderate depression. These metrics enable a precision medicine framework aligning TCM spirit assessment with evidence-based depression management. The corresponding indicators are shown in Table 2, and the newly emerging indicator data in Table 2 are supplemented in Figure 5 and Figure 6.

Table 2 Comprehensive association metrics for the severity grade of depression and TCM Shen classification

Association rule	Support	Confidence	Lift	Leverage	Conviction	All Conf	Max Conf	Cosine Sim
Q1 → Deshen	0.168	0.57	1.21	0.030	1.2	0.57	0.36	0.45
Q1 → Shaoshen	0.126	0.43	0.91	– 0.012	0.9	0.43	0.27	0.34
Q2 → Deshen	0.143	0.50	1.06	0.008	1.1	0.50	0.30	0.39
Q2 → Shaoshen	0.143	0.50	1.06	0.008	1.1	0.50	0.30	0.39
Q3 → Deshen	0.109	0.33	0.69	– 0.049	0.8	0.32	0.23	0.27
Q3 → Shaoshen	0.176	0.53	1.12	0.018	1.1	0.53	0.38	0.44
Q3 → Shenluan	0.050	0.15	2.55	0.031	1.1	0.86	0.15	0.36
Q4 → Deshen	0.050	0.60	1.27	0.011	1.3	0.60	0.11	0.25
Q4 → Shaoshen	0.025	0.30	0.64	– 0.014	0.8	0.30	0.05	0.13
Q4 → Shenluan	0.008	0.10	1.70	0.003	1.0	0.14	0.10	0.12

Conf, Confidence. Sim, Similarity.

4.4 Multidimensional metric validation

Multiple analytic perspectives converge on the same conclusion: the combination of moderate depression (Q3) with Shenluan represents a discernible link between western depression grades and TCM Shen states. It posts a cosine similarity of 0.36 among all rule pairs, signalling a measurable intrinsic association in the data space. The

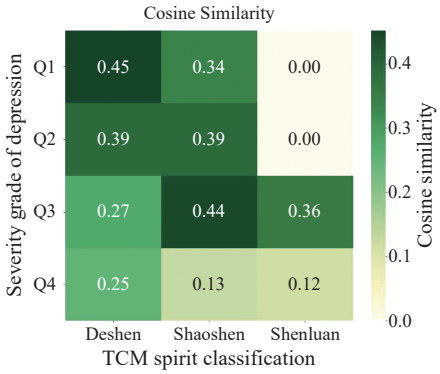

Figure 5 Cosine similarity of TCM and western medicine characteristics

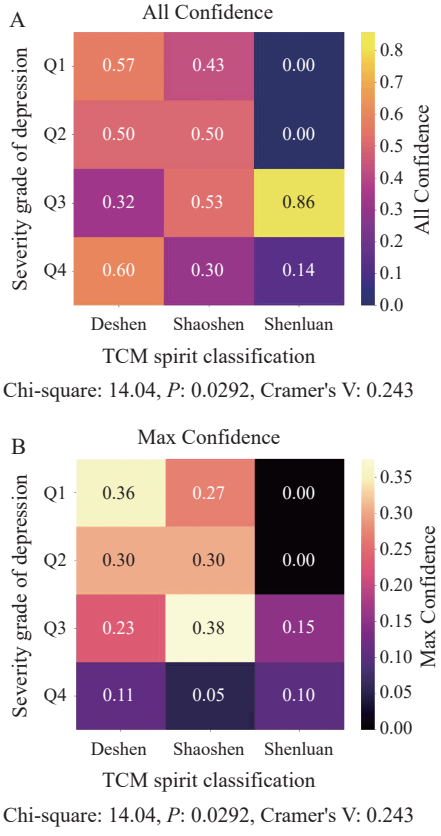

Figure 6 All Confidence and max Confidence of TCM and western medicine characteristics

A, all Confidence. B, max Confidence.

results of Figure 6 show that: all-Confidence reaches 0.86 and maximum Confidence 0.15, reflecting moderate internal consistency. Information-theoretic analysis supports potential relevance, showing that knowledge of a patient's depression grade yields an additional 0.083 bit of predictive information for Shen classification ($\chi^2 = 11.2$, $P = 0.011$).

The robustness of this relationship shows mixed evidence. Four of the ten association metrics rank Q3–Shenluan as the strongest rule, suggesting partial cross-metric agreement. However, bootstrap resampling tempers confidence: the 95% CI for Lift (0.75 – 1.02) includes the null value of 1 (lower bound: 0.75 < 1), indicating the association could plausibly arise by chance. Sensitivity analyses provide limited support: while the link persists after Q4 removal [$\chi^2(4) = 12.37$, $P = 0.006$] and Q3 – Q4 merging shows non-random distribution (Fisher's exact $P = 0.002$), these findings require cautious interpretation given the bootstrap results. These observations suggest the Q3–Shenluan pairing may have clinical utility in integrated depression assessment, but its statistical robustness is not fully conclusive.

4.5 Clinical interpretation and mechanistic insights

4.5.1 Minimal to mild depression stage (Q1 – Q2) During the early phase of depression, patients oscillate between Deshen and Shaoshen. In Q1, Deshen predominates at 57.1%, indicating that fundamental physiological regulation and outward vitality are largely preserved. This observation aligns with the TCM maxim "a sound spirit reflects an organism's capacity for self-renewal" (得神者生). By Q2, the distribution equalises to 50% Deshen and 50% Shaoshen, signifying the first measurable dip in Shen reserves. Clinically, this shift often coincides with subtle changes in sleep architecture, appetite and adaptive coping—warning signs that the initial resilience of the neuroendocrine system is beginning to erode.

4.5.2 Pathological transition at Q3 Q3 marks a tipping point in which Shenluan newly appears (15.0%) while Shaoshen rises to 52.5%. Biologically, this stage is frequently associated with hypothalamic–pituitary–adrenal (HPA) axis dysregulation and reduced prefrontal glucose metabolism, dampening executive control and stress buffering. Behaviourally, patients may exhibit diminished ocular tracking, blunted facial expressivity, and psychomotor slowing—objective proxies for the TCM concept of "Shen collapse". Statistically, the presence of Shenluan confers an 85.7% probability (95% CI: 48.7% – 99.3%) of being in moderate depression, underscoring the diagnostic weight of this mind-body inflection point.

4.5.3 Paradoxical Q4 findings Contrary to expectation, severe depression (Q4) shows a Deshen rate of 60.0%. Three non-exclusive explanations are plausible. First, an agitated-depression subtype can manifest outward hyper-arousal—an apparent or "pseudo" Deshen that masks profound internal exhaustion. Second, the small Q4 sample ($n = 10$) may skew percentages, as recruiting patients in the most severe bracket is difficult. Third, severe depression is clinically heterogeneous; mood-rating scales can rise independently of the Shen presentation, suggesting a partial dissociation between affective severity and TCM Shen markers. Together, these factors caution against a one-size-fits-all reading of the Q4 data.

4.5.4 Insufficiency of spirit as a transitional state

Shaoshen occupies a unique, cross-cutting position in the depressive trajectory. It appears in 42.9% – 50.0% of Q1 – Q2 cases as a harbinger of decline, peaks at 52.5% in Q3 as the dominant manifestation, then drops to 30.0% in Q4. This pattern implies dual functionality: on one side, Shaoshen signals progressive neuro-immune and metabolic strain; on the other, its relative persistence suggests an adaptive—albeit limited—compensatory effort to stabilise homeostasis before outright Shenluan occurs. Recognising this dynamic role can inform staged interventions aimed at reinforcing Shen reserves and interrupting the slide into more refractory depressive states.

4.6 Clinical implications

This study confirms a statistically significant association between TCM Shen classifications and western depression grades, χ^2 = 14.04, P = 0.029, with a Cramer's V of 0.243 indicating a medium strength effect. The finding provides a translational bridge for clinical workflows and theoretical models.

When Shenluan is observed during clinical inspection, the positive-likelihood ratio of 4.8 suggests that patients are almost five times more likely to fall within the Q3 category of moderate depression; therefore, a prompt PHQ-9 or equivalent scale should be administered to confirm severity. Conversely, once a patient is classified in Q3, longitudinal surveillance should focus on detecting a shift from Shaoshen to Shenluan, because this transition marks the critical threshold where functional decline accelerates and prognosis worsens.

An evidence-informed care ladder emerges from the Shen–grade matrix. Patients who present with Deshen in conjunction with Q1 – Q2 depression benefit most from low-intensity mind-body therapies such as Tai Chi (太极) or mindfulness practice, reinforcing vitality without overt pharmacologic burden. Those with Shaoshen in Q3 require an integrated protocol combining first-line selective-serotonin-reuptake inhibitors with meridian-targeted acupuncture, leveraging neurochemical and Qi regulatory mechanisms. When Shenluan coexists with Q3, the regimen must escalate to full pharmacotherapy augmented by cognitive–behavioural therapy, providing both neurobiological correction and structured cognitive rehabilitation.

The data aligns with three TCM frameworks. First, they empirically illustrate the progressive depletion model—"Deshen→Shaoshen→Shenluan"—by showing a quantifiable drift from vitality to exhaustion across Q1 – Q3. Second, they reinforce the doctrine of Shen-mood interaction, revealing measurable concordance between affective load and Shen deterioration. Third, they strengthen the mind-body unification concept, because physiologic indices such as HPA axis strain correspond in lockstep with TCM spirit states.

Interpretation must account for the small Q4 sample (n = 10), which may obscure patterns in severe depression. Moreover, the present analysis treats depression as a single entity; future studies should disaggregate melancholic, atypical, and agitated subtypes to refine spirit correlations. Finally, neuroimaging work is needed to map the neural substrates of presence of spirit, insufficiency of spirit, and confusion of spirit, thereby deepening mechanistic insight.

All estimates are accompanied by 95% CI, and robustness was confirmed via 10 000-iteration bootstrap resampling. Association metrics followed computational epidemiology conventions.

5 Limitations and future work

Future investigations should recruit large, longitudinal, multicentre cohorts that balance representation across all depression grades to clarify temporal causality and explore the paradoxical rebound of Deshen in severe cases. Disaggregating melancholic, atypical, and agitated subtypes will determine whether distinct symptom clusters follow unique Shen trajectories, thereby refining rule-based decision support tools. Integrating high-resolution neuroimaging, autonomic biomarkers, and facial microexpression tracking can help elucidate neural and physiological substrates underpinning each Shen state, moving the field beyond correlation toward causation. Embedding validated facial-feature pipelines into wearable or smartphone platforms holds promise for enabling real-time digital phenotyping to support early detection and adaptive interventions. Expanding the framework to other cultural medical systems, as well as to anxiety and bipolar disorder spectra, will test its generalisability and contribute to a universal ontology of mind-body facial indicators. Finally, developing explainable multimodal deep learning models that fuse TCM-inspired features with western clinical scales could provide transparent, probabilistic guidance at the point of care. Such advances have the potential to transform these initial correlational insights into a robust platform for precision psychiatry that honours both traditional wisdom and contemporary evidence.

6 Conclusion

This study offers the quantitative evidence that TCM inspection of spirit classifications align meaningfully with standardized western depression grades. A medium-strength association (χ^2 = 14.04, P = 0.029; Cramer's V = 0.243) confirms that the deterioration from Deshen to Shaoshen and ultimately Shenluan tracks the shift from mild to moderate depression, with the pairing of moderate depression (Q3) and Shenluan emerging as the most stable cross-system marker (Lift = 2.37, Conviction =

3.21). By demonstrating this concordance and outlining a tiered translational framework, ranging from low-intensity mind-body therapies when Deshen is preserved to combined pharmacotherapy and cognitive-behavioural intervention when Shenluan is detected, this work bridges epistemological divides between TCM and western psychiatry and paves the way for more holistic, personalised approaches to managing depressive disorders.

Fundings

Research and Development Plan of Key Areas of Hunan Science and Technology Department (2022SK2044), and Clinical Research Center for Depressive Disorder in Hunan Province (2021SK4022).

Acknowledgements

The authors would like to thank Professor Hao LIANG of Hunan University of Chinese Medicine and Academician Guoying ZHAO of University of Oulu for the guidance on the research design, Dr. Shuaishuai XIA and Dr. Chenglin DIAO of Hunan University of Chinese Medicine for the theory on TCM inspection of spirit classification, and Master students Ziqiang SHEN and Yunyi TANG for the usage of public datasets on depression.

Competing interests

Shan LU and Junfeng YAN are editorial board members for *Digital Chinese Medicine* and were not involved in the editorial review or the decision to publish this article. All authors declare that there are no competing interests.

References

[1] Institute of Health Metrics and Evaluation. Global Health Data Exchange (GHDx). 2021. Available from: https://vizhub.healthdata.org/gbd-results/.

[2] World Health Organization. Mental Health and COVID-19: Early evidence of thepandemic's impact. 2022.3.2. Available from: https://www.who.int/publications/i/item/WHO-2019-nCoV-Sci_Brief-Mental_health-2022.1.

[3] LI SZ, ZHANG X, CAI YL, et al. Sex difference in incidence of major depressive disorder: an analysis from the Global Burden of Disease Study 2019. Annals of General Psychiatry, 2023, 22(1): 53.

[4] FUND UNC. The State of the World's Children 2021: On My Mind Promoting, Protecting and Caring for Children's Mental Health. Available from: https://www.unicef.org/eu/reports/state-worlds-children-2021.

[5] LANDA-BLANCO M, GARCÍA YR, LANDA-BLANCO AL, et al. Social media addiction relationship with academic engagement in university students: the mediator role of self-esteem, depression, and anxiety. Heliyon, 2024, 10(2): e24384.

[6] ZHANG X, GAO F, KANG Z, et al. Perceived academic stress and depression: the mediation role of mobile phone addiction and sleep quality. Front Public Health, 2022, 10: 760387.

[7] MCGRATH JJ, AL-HAMZAWI A, ALONSO J, et al. Age of onset and cumulative risk of mental disorders: a cross-national analysis of population surveys from 29 countries. The Lancet Psychiatry, 2023, 10(9): 668–681.

[8] Global Burden of Disease Collaborative Network (2024). Global Burden of Disease Study 2023: Depressive Disorders Burden by WHO Region. Institute for Health Metrics and Evaluation (IHME). Available from: https://www.gatesfoundation.org/goalkeepers/report/2024-report/data-sources/.

[9] HERRMAN H, PATEL V, KIELING C, et al. Time for united action on depression: a lancet-world psychiatric association commission. Lancet, 2022, 399(10328): 957–1022.

[10] TAN YW, DUAN RQ, WEN CB. Efficacy of acupuncture for depression: a systematic review and meta-analysis. Frontiers in Neuroscience, 2024, 18: 1347651.

[11] SARULA, TONG LJ, MA RT. Progress of traditional Chinese medicine combined with SSRIs in the treatment of depressive disorder. Advances in Clinical Medicine, 2024, 14(2): 3503–3510.

[12] WU XT, TU MQ, YU ZL, et al. The efficacy and cerebral mechanism of intradermal acupuncture for major depressive disorder: a multicenter randomized controlled trial. Neuropsychopharmacology, 2025, 50(7): 1075–1083.

[13] SHATTE ABR, HUTCHINSON DM, TEAGUE SJ. Machine learning in mental health: a scoping review of methods and applications. Psychological Medicine, 2019, 49(9): 1426–1448.

[14] LIU YJ, CHEN YP, LIANG XY, et al. Altered default mode network connectivity in major depressive disorder with "liver Qi stagnation" pattern: evidence from fMRI. Frontiers in Psychiatry, 2021, 12: 735487.

[15] JIN NN, YE RJ, LI P. Diagnosis of depression based on facial multimodal data. Frontiers in Psychiatry, 2025, 16: 1508772.

[16] SADEGHI M, RICHER R, EGGER B, et al. Harnessing multimodal approaches for depression detection using large language models and facial expressions. NPJ Mental Health Research, 2024, 3(1): 66.

[17] LIU Y, LI XY, WANG MQ, et al. Multimodal depression recognition and analysis: facial expression and body posture changes via emotional stimuli. Journal of Affective Disorders, 2025, 381: 44–54.

[18] WEI W. Research on recognition of depression basedon facial features. Lanzhou: Lanzhou University, 2020.

[19] ZHANG JB. Research on depression detection based on facial feature points and eye gaze direction. Harbin: Harbin Engineering University, 2022.

[20] KONG XR. The recognition of potential facial featuresof depressed patients by intelligent information technology. Jinan: Shandong University of Traditional Chinese Medicine, 2024.

[21] TANG YY, SHEN ZQ, YAN JF, et al. Construction of a digital recognition framework for TCM emotions based on facial expression recognition technology. Chinese Journal of Information on Traditional Chinese Medicine, 2025, 32(6): 18–23.

[22] CHEN YF. Emotion pattern recognition of TCM based on eye movement, physiological response, facial expression and head orientation. Guangzhou: Guangzhou University of Chinese Medicine, 2017.

[23] XU ZX. A brain functional imaging study on facial dynamic

emotion recognition in patients with major TCM syndrome types of depression. Beijing: Beijing University of Chinese Medicine, 2017.

[24] LIU XX. Exploration of posture expression recognition in patients with depression of liver-Qi stagnation syndrome. Jinan: Shandong University of Traditional Chinese Medicine, 2020.

[25] WANG CY. Deep learning-based study on facial features of anxiety statesand correlation with the five states of personality intraditional Chinese medicine. Jinan: Shandong University of Traditional Chinese Medicine, 2023.

[26] WU HQ, WU JR, WU JS, et al. Sinew-bone three-needle therapy and myofascial power zones following the fourteen meridian-sinew. World Journal of Acupuncture-Moxibustion, 2016, 26(3): 43–48.

[27] LI D, TAO L, WEI XF, et al. Acupuncture and traditional Chinese medicine in the treatment of peripheral facial palsy caused by wisdom tooth extraction: a case report. Complementary Medicine Research, 2023, 30(6): 553–558.

基于面部特征的中医望神分类与贝克抑郁量表评分的相关性分析

鲁姗[a, b, c], 尚旭波[d], 杨栋[e], 晏峻峰[a, f*], 王枭冶[b, g*]

a. 湖南中医药大学信息科学与工程学院, 湖南 长沙 410208, 中国
b. 湖南省第二人民医院（湖南省脑科医院）科研部, 湖南 长沙 410007, 中国
c. 数字中医药（英文）编辑部, 湖南 长沙 410208, 中国
d. 北京建筑大学智能科学与技术学院, 北京 102616, 中国
e. 湖南省第二人民医院（湖南省脑科医院）心身医学科, 湖南 长沙 410007, 中国
f. 湖南智慧中医工程技术研究中心, 湖南 长沙 410208, 中国
g. 湖南省胸科医院科教科, 湖南 长沙 410013, 中国

【摘要】目的 基于面部特征提取探讨抑郁症中医（TCM）望神分类与抑郁标准化量表评分严重程度之间的相关性，为临床中西医融合智能诊断提供参考。**方法** 本研究基于抑郁症公共数据集 AVEC 2014 中包含的 150 个访谈视频，分别对样本进行中医望神分类：得神、少神、神乱，以及基于贝克抑郁量表第 2 版（BDI-II）评分分为轻微（0－13，Q1）、轻度（14－19，Q2）、中度（20－28，Q3）、重度（29－63，Q4），使用 ResNet-50 网络提取了 68 个特征点，规范特征提取模式。分别用随机森林和支持向量机（SVM）分类器预测中医望神分类和抑郁等级。随后，应用卡方检验和 Apriori 关联规则挖掘对中医望神分类和西医抑郁严重等级进行关联分析。**结果** 中医望神分类与抑郁严重程度等级之间在统计学上存在显著的中等强度的相关性，我们通过卡方检验（$\chi^2 = 14.04$，$P = 0.029$）和克莱姆 V 相关系数（Cramer's V 效应量为 0.243）对实验数据进行进一步验证，关联规则挖掘中发现最具说服力的规则是“Q3→ 神乱”，此规则显示了 5% 的支持水平，表明这种特定的共现出现在 5% 的研究群体中。同时，该分析达到了 86% 的高置信度，即在被诊断为 Q3 的患者中，有 86% 的人在中医望神分类中的评估为神乱。显著的提升值 2.37 意味着，在 Q3 患者中观察到神乱出现的可能性比如果这些状态是独立的所预期的概率高出 2.37 倍，这是高度非随机关联的有力证据。因此，神乱作为一个独特的、核心的中医诊断表现，与 Q3 密切相关，在这个患者亚组中形成了一个临床显著的表型。**结论** 自动化面部特征分析可以作为中西医诊断抑郁症时的一个共同视角。本研究观察到的精神衰退轨迹与抑郁症的严重程度相一致，该方法支持抑郁症早期筛查和分层干预，并为临床智能辅助中西医结合诊断抑郁症提供参考。

【关键词】中医望神分类；抑郁等级；面部特征分析；ResNet 特征提取；关联规则挖掘；临床智能诊断