Contents lists available at ScienceDirect

Digital Chinese Medicine

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

A machine learning-based depression recognition model integrating spirit-expression features from traditional Chinese medicine

Minghui Yao[a†], Rongrong Zhu[a†], Peng Qian[a], Huilin Liu[a], Xirong Sun[b], Limin Gao[b*], Fufeng Li[a, b*]

a. School of Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China

b. Department of Traditional Chinese Medicine, Shanghai Pudong New Area Mental Health Center, Shanghai 200124, China

A R T I C L E   I N F O

Article history
Received 11 August 2025
Accepted 05 January 2026
Available online 25 March 2026

Keywords
Traditional Chinese medicine
Spirit
Expression
Feature fusion
Depression
Recognition model

A B S T R A C T

Objective To develop a depression recognition model by integrating the spirit-expression diagnostic framework of traditional Chinese medicine (TCM) with machine learning algorithms. The proposed model seeks to establish a TCM-informed tool for early depression screening, thereby bridging traditional diagnostic principles with modern computational approaches.

Methods The study included patients with depression who visited the Shanghai Pudong New Area Mental Health Center from October 1, 2022 to October 1, 2023, as well as students and teachers from Shanghai University of Traditional Chinese Medicine during the same period as the healthy control group. Videos of 3 – 10 s were captured using a Xiaomi Pad 5, and the TCM spirit and expressions were determined by TCM experts (at least 3 out of 5 experts agreed to determine the category of TCM spirit and expressions). Basic information, facial images, and interview information were collected through a portable TCM intelligent analysis and diagnosis device, and facial diagnosis features were extracted using the Open CV computer vision library technology. Statistical analysis methods such as parametric and non-parametric tests were used to analyze the baseline data, TCM spirit and expression features, and facial diagnosis feature parameters of the two groups, to compare the differences in TCM spirit and expression and facial features. Five machine learning algorithms, including extreme gradient boosting (XGBoost), decision tree (DT), Bernoulli naive Bayes (BernoulliNB), support vector machine (SVM), and k-nearest neighbor (KNN) classification, were used to construct a depression recognition model based on the fusion of TCM spirit and expression features. The performance of the model was evaluated using metrics such as accuracy, precision, and the area under the receiver operating characteristic (ROC) curve (AUC). The model results were explained using the Shapley Additive exPlanations (SHAP).

Results A total of 93 depression patients and 87 healthy individuals were ultimately included in this study. There was no statistically significant difference in the baseline characteristics between the two groups ($P > 0.05$). The differences in the characteristics of the spirit and expressions in TCM and facial features between the two groups were shown as follows. (i) Quantispirit facial analysis revealed that depression patients exhibited significantly reduced facial spirit and luminance compared with healthy controls ($P < 0.05$), with characteristic features

†The authors contributed equally.

*Corresponding author: Fufeng Li, E-mail: li_fufeng@aliyun.com. Limin Gao, E-mail: healthybank@126.com.

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

DOI: 10.1016/j.dcmed.2026.02.006

Citation: YAO MH, ZHU RR, QIAN P, et al. A machine learning-based depression recognition model integrating spirit-expression features from traditional Chinese medicine. Digital Chinese Medicine, 2026, 9(1): 68-79.

such as sad expressions, facial erythema, and changes in the lip color ranging from erythematous to cyanotic. (ii) Depressed patients exhibited significantly lower values in facial complexion L, lip L, and a values, and gloss index, but higher values in facial complexion a and b, lip b, low gloss index, and matte index (all $P < 0.05$). (iii) The results of multiple models show that the XGBoost-based depression recognition model, integrating the TCM “spirit-expression” diagnostic framework, achieved an accuracy of 98.61% and significantly outperformed four benchmark algorithms—DT, BernoulliNB, SVM, and KNN ($P < 0.01$). (iv) The SHAP visualization results show that in the recognition model constructed by the XGBoost algorithm, the complexion b value, categories of facial spirit, high gloss index, low gloss index, categories of facial expression and texture features have significant contribution to the model.

Conclusion This study demonstrates that integrating TCM spirit-expression diagnostic features with machine learning enables the construction of a high-precision depression detection model, offering a novel paradigm for objective depression diagnosis.

1 Introduction

Modern psychosocial stressors, such as increasingly accelerated lifestyles and intense social competition, are contributing to a growing mental health crisis. Among these conditions, depression has become particularly prevalent, characterized by high incidence, disability rates, as well as substantial gaps in treatment coverage [1]. According to the estimates by the World Health Organization (WHO), approximately 3.8% of the global population is affected by depression [2]. Current diagnostic methods, which rely predominantly on clinician interviews and subjective rating scales, remain susceptible to recall bias and measurement variability [3], highlighting the urgent need for more objective and reliable assessment tools.

Computer-aided diagnosis has demonstrated significant potential in mental health applications, particularly for early prediction and screening of psychiatric disorders [4]. At present, a number of machine learning-based early-warning models for depression have been developed, most of which rely on questionnaire-based survey data [5], electroencephalogram (EEG) data [6], and other sources, to predict and diagnose depression, thereby providing early warning and timely intervention for at-risk individuals.

Even widely used self-report tools such as the Patient Health Questionnaire-9 (PHQ-9) possess inherent limitations, including reliance on subjective reporting, coverage of only a partial range of symptoms, and the inability to independently assess symptom duration or rule out other conditions [7]. Moreover, reliance on a single type of scale to assess depression severity may limit the comparability of different questionnaires and affect the reproducibility and generalizability of depression research [8], thus restricting its application in depression diagnosis. EEG relies on specialized equipment, strict experimental conditions, and trained technicians for operation, involves substantial cost, and may cause discomfort to patients, thereby reducing compliance and limiting its suitability for continuous daily monitoring. In contrast, facial feature-based assessments demonstrate distinct advantages. First, digital and objective assessment represents a key direction in the current development of mental health tools [9]. Second, face-to-face consultations can enable non-contact, low-cost data acquisition through conventional cameras, offering high clinical convenience and accessibility, features that support large-scale screening and remote health care applications.

In depression early-warning research, facial micro-expressions or changes in facial muscle movement are often analyzed through video-based facial action unit features to predict depression [10]. Study indicates that reduced activity of the cheek muscles is the most prominent facial manifestation of depression [11]. Other work has identified the corners of the mouth, cheeks, and eyes as key regions distinguishing depressed patients from healthy individuals, with differences also observed between males and females [12]. Additional study reports that depression-related facial characteristics primarily involve the mouth and eye regions [13]. Recently, the diagnostic and therapeutic advantages of traditional Chinese medicine (TCM) in depression have gained increasing recognition. Research has revealed distinct facial diagnostic features in depressed patients; for example, diminished pupillary oscillation (manifested as a vacant gaze) has been noted as a primary characteristic of facial expression and ocular spirit in liver-Qi stagnation syndrome [14], providing a reference for assessing depression from the TCM perspective of spirit-expression observation.

Gesture, expression, and movement characteristics of individuals with depression can reflect the emotional state associated with depressed liver-Qi stagnation syndrome [15]. These findings confirm that the application of spirit-expression analysis in disease diagnosis has gained increasing attention, and alterations in spiritual state and expressive features are closely linked to the onset of depression, suggesting that spirit-expression analysis may offer a novel basis for predicting, identifying, and treating depression at an early stage. However, research specifically focused on depression recognition by integrating TCM spirit-expression features remains relatively limited in the current literature.

Based on the TCM spirit-expression features, this study explores the construction of a depression recognition model using machine learning algorithms, aiming to provide an efficient and low-cost approach for depression screening in contemporary populations—an effort with important implications for depression prevention and treatment. Following interdisciplinary paradigms, we apply computer-based information processing techniques to extract and analyze TCM spirit-expression features from healthy controls and depressed individuals, integrate the derived information, and develop a depression recognition model to offer auxiliary reference data for the identification of depression cases.

2 Materials and methods

2.1 Data source

Patients with depression who sought medical treatment at Shanghai Pudong New Area Mental Health Center were enrolled, from October 1, 2022 to October 1, 2023, along with students and faculty from Shanghai University of Traditional Chinese Medicine as the healthy control group. The study was conducted by the Institutional Review Board of Shanghai University of Traditional Chinese Medicine (Approval No. 2024-2-14-08), and the study was approved according to the Declaration of Helsinki principles. All written informed consents were obtained from participants.

2.2 Diagnostic criteria for depression

2.2.1 Western diagnostic criteria for depression The diagnostic criteria for depression are based on the Diagnostic and Statistical Manual of Mental Disorders (15th edition) [16] (DSM-V). The main symptom is a depressed mood, accompanied by at least four of the following symptoms for more than two weeks: (i) loss of interest or pleasure; (ii) decreased energy or fatigue; (iii) psychomotor retardation or agitation; (iv) low self-esteem, self-blame, or guilt; (v) concentration defects; (vi) thoughts or behaviours of death, or suicidal or self-harming; (vii) sleep disturbance, such as insomnia; (viii) decreased appetite or significant weight loss; (ix) decreased libido.

2.2.2 TCM diagnostic criteria for depression According to the Diagnostic Criteria and Treatment Plans of TCM Syndromes for Depression [17], the diagnosis requires persistent mental depression (⩾ 2 weeks) as the core symptom, accompanied by at least four of seven secondary symptoms: anhedonia, irritability, psychomotor retardation, fatigue/weakness, insomnia, memory impairment, and decreased appetite/libido.

2.3 Inclusion and exclusion criteria

2.3.1 Inclusion criteria The inclusion criteria for the depression are as follows: (i) fulfill all aforementioned diagnostic criteria; (ii) aged 15 – 60 years with symptom duration ⩾ 2 weeks; (iii) consensus diagnosis by at least 3 of 5 independent TCM practitioners through standardized video assessments; (iv) written informed consent was obtained independently from both participants (all ⩾ 18 years old) and legal guardians. Accompanying relatives were informed to assist with communication, with the final decision by participants.

The inclusion criteria for the healthy controls are as follows: (i) no psychiatric history; (ii) normal anxiety/depression scores [Self-Rating Anxiety Scale (SAS)/Self-Rating Depression Scale (SDS) < 50]; (iii) unremarkable physical examinations. Groups were age- and gender-matched to minimize confounding variables.

2.3.2 Exclusion criteria for depression The exclusion criteria for the depression are as follows: (i) failure to meet the aforementioned diagnostic and inclusion criteria; (ii) secondary depression attributable to organic pathologies or severe cardiopulmonary, hepatic, and renal dysfunction; (iii) active suicidal ideation with high-risk propensity; (iv) incomplete or non-standardized data collection; (v) unanalyzable facial images due to technical or quality limitations.

2.4 Sample collection and interpretation methods

Upon entering the assessment environment, participants assumed either a seated or standing position and remained at rest for precisely 60 s before evaluation commenced. Subsequently, a Xiaomi Pad 5 was positioned at a fixed distance of 50 cm from the participant's face, and the participant was instructed to maintain gaze at the camera lens while a 3 – 10 s video was recorded using the device's built-in camera under controlled conditions (lighting: color temperature 5 500 ± 300 K, illumination: 400 – 500 lx; ambient noise: 40 – 50 decibels). Portable intelligent analysis and TCM diagnosis equipment were then employed to collect participants' basic information, facial images, and TCM consultation data (Supplementary Figure S1). Participants were instructed to sit upright with the chin on a fixed frame, keep the face centered without deviation, and maintain a natural, relaxed expression for 3 – 5 s to capture facial images, with color judgment based on a TCM-specific color card developed from a 24-color standard color card (Patent No. CN110495888B) [18].

2.5 Data preprocessing and feature engineering

2.5.1 Image preprocessing flow Facial images and frames extracted from videos have undergone unified preprocessing. The standardized facial regions are detected and cropped from the original images. The images are converted from red, green, blue (RGB) color space to the Commission Internationale de I'Eclairage Lab (CIELAB) color space, and the values of the L, a, and b channels are linearly normalized to the range of 0 – 255 to eliminate illumination differences. Gaussian filters (kernel size = 5 × 5 pixels) are applied to suppress image noise. In the Lab color space, L represents brightness, ranging from 0 to 100, indicating from black to white, with a higher value indicating a brighter detection area. The a values range from – 128 to 127, represented from green to red, indicating the redness of the area. The b values represented from blue to yellow. This study is based on the Open CV computer vision library, and the Lab components have been normalized, with the value ranging from 0 to 255.

2.5.2 Feature extraction and quantification Facial features were selected based on the TCM theory of facial diagnosis [19]. According to TCM theory, regarding the 12 meridians and 365 collaterals, their blood and Qi all ascend to the face and flow to the orifices. The expressions, shapes, and colors of the face are important external manifestations of the internal organs and meridians. Therefore, we aimed to digitize and quantify the three core elements of spirit, color, and form. Based on this theory, we referred to our previous research methods [20, 21] and extracted the following three types of features.

(i) Color, luster, and texture features. Color, luster, and texture features were systematically qualified following standardized facial image acquisition, providing an objective analysis of two key dimensions of TCM facial inspection—overall complexion and specific lip color. Color and luster were measured separately for the entire facial region and for an automatically segmented lip region within the perceptually uniform CIELAB (L, a, b) color space using the OpenCV library. Mean L, a, and b values were calculated for each region, with channel values normalized to a 0 – 255 range. Within the TCM diagnostic framework, the L value objectively represents skin/lip luster, a key indicator of spirit, while the a value provides a quantitative measure of redness. This process yielded six continuous features, corresponding to the three color channels for both the facial and lip regions.

Texture features were extracted from model-attentive facial regions to quantitatively assess textural patterns associated with TCM spirit and expression. Specifically, features were extracted from facial regions automatically identified as significant through model interpretability analysis. Gradient-weighted class activation mapping (Grad-CAM) was applied to our integrated spirit-expression classification model, generating heatmaps that objectively highlight the facial areas most influential for model decisions, which primarily included the eye, eyebrow, and perioral regions. From grayscale patches of these high-activation areas, textural features were extracted using the gray-level co-occurrence matrix (GLCM) [22], yielding contrast and energy statistics to quantify local intensity variation and textural uniformity within these diagnostically salient regions. This approach generated a set of continuous texture features for each subject, directly derived from the model-identified key facial areas.

(ii) Spirit classification label. Spirit labels were determined by a panel of five certified TCM practitioners based on the classification norms presented in the first chapter of the nationally authorized planning textbook, *Diagnosis of Traditional Chinese Medicine* [23], which outlines the qualitative criteria for assessing spirit and facial expressions. A consensus from the majority of practitioners (⩾ 3/5 agreement) served as the gold-standard categorical variable for model input.

(iii) Expression classification label. Quantification method: similar to the assessment of spirit, five certified TCM practitioners independently evaluated the expression observed in the key video segments, guided by the standards outlined in the Induction and Assessment of Expression [24]. The majority consensus (defined as agreement by at least three of the five practitioners marked the same) was adopted as the gold-standard label. Feature form: this expressive label was input as another categorical variable into the model.

2.5.3 Feature selection and selection basis Given that the feature set extracted in this study, based on prior TCM theoretical knowledge, is of relatively low dimension and all the features have clear clinical significance and theoretical support. We did not adopt algorithm-driven feature selection. Instead, all theoretically selected features were included in the model to avoid discarding any potential markers, thereby ensuring a comprehensive exploration of their predictive value. The final feature set used is shown in Table 1.

2.6 Machine learning approach for recognition model construction

Given the limited sample size in this study, we adopted the following research design strategies to ensure model robustness and control overfitting. First, the core strategy

Table 1 Definitions and label values of facial information features

Feature label	Label value
Categories of facial spirit	Reduced spirit (0), loss of spirit (1), presence of spirit (2)
Categories of facial expressions	Joy (0), sadness (1), neutral (2)
Texture features	GLCM features
Categories of overall facial complexion	White (0), black (1), red (2), yellow (3), bluish/cyan (4), normal (5)
CIELAB color space values of overall facial complexion	L, a, b
Facial gloss categories	Glossy (0), reduced gloss (1), non-glossy (2)
Levels of facial skin gloss index	high gloss index (0), moderate gloss index (1), low gloss index (2)
Lip color categories	Pale white (0), light red (1), red (2), dark red (3), purple (4)
CIELAB color space values of lip color	L, a, b

involved employing a low-dimensional feature set predefined based on TCM theory, which inherently reduces model complexity. Second, we prioritized classical machine learning methods, such as decision trees (DT) [25], extreme gradient boosting (XGBoost) [26], k-nearest neighbor (KNN) [27], support vector machine (SVM) [28], and Bernoulli naive Bayes (BernoulliNB) [29], to construct interpretable, relatively simple models rather than complex deep-learning architectures. All implementations used Python V3.8.5. Finally, all models were evaluated using stratified k-fold cross-validation, a well-established method for obtaining reliable performance estimates in small-sample learning scenarios.

The performance of the classification models was evaluated using the following metrics: accuracy, precision, recall, and F1-score. The area under the receiver operating characteristic (ROC) curve (AUC) was also employed as indicators of model performance. The corresponding calculation formulas have been listed in Supplementary Table S1. Images and their corresponding labels from the entire dataset were used as input for the classification methods described above. A five-fold cross-validation scheme was implemented, in which the dataset was randomly partitioned into five equal subsets. In each fold, four subsets (80% of the data) were used for training, and the remaining subset (20%) was reserved for testing. This cross-validation approach provides a nearly unbiased estimate of generalization error, thereby helping to obtain a reliable and stable model. It also enhances the model's learning capacity and mitigates overfitting, which is particularly important for small-sample datasets. Test results were recorded, and evaluation metrics such as accuracy were examined accordingly.

In this initial exploratory study, we utilized the default hyperparameters for all models as provided by the respective machine learning libraries (e.g., scikit-learn and XGBoost). This approach was adopted to establish baseline performance and to mitigate the risk of overfitting the limited training data through an extensive hyperparameter search. The default parameters settings for key models are detailed in Supplementary Table S2. To interpret the model's predictions and assess the contribution of individual features, we utilized Shapley Additive exPlanations (SHAP). This method quantifies the influence of each feature on the model's output, offering interpretable insights into the decision-making process.

2.7 Statistical analysis

All statistical analyses were performed using SPSS 26.0. Normality of continuous variables was evaluated using the Shapiro-Wilk test. Data following a normal distribution are presented as mean ± standard deviation (SD), while non-normally distributed data are expressed as median with interquartile range (IQR). Between-group comparisons were performed using independent-samples t tests for parametric data (with Levene's test for equality of variances) or the Mann-Whitney U test for non-parametric distributions. $P < 0.05$ was considered statistically significant.

3 Results

3.1 Baseline data

Baseline demographic and anthropometric data, including gender, age, height, weight, and body mass index (BMI), were collected from all participants upon enrollment. A total of 93 depression patients and 87 healthy controls were enrolled, and no statistically significant differences were observed in general data between the two groups ($P > 0.05$) (Table 2).

3.2 Comparison of the facial diagnosis features between the depression patients and the healthy controls

Comparison of facial diagnostic features between depression patients and healthy controls revealed that the complexion L value, luster index, lip color L value, and lip color a value were significantly lower in the depression

Table 2 Comparison of general data between the depression patients and the healthy controls

Factor	Healthy control (n = 87)	Depression patient (n = 93)	Statistical value	P value
Gender (male/female)	23 (26.44%)/64 (73.56%)	36 (38.71%)/57 (61.29%)	χ^2 = 3.073	0.080
Age (year)	29.75 ± 10.72	38.16 ± 18.89	Z = – 1.920	0.055
Height (cm)	165.98 ± 8.39	165.41 ± 9.11	Z = – 0.304	0.761
Weight (kg)	58.75 ± 10.04	60.89 ± 12.93	Z = – 1.121	0.262
BMI	21.24 ± 2.65	22.15 ± 3.65	Z = – 1.929	0.054

group (P < 0.05). Conversely, the values of complexion a and b, low gloss index, glossless index, and lip color b value were all significantly higher in depression patients compared with healthy controls (P < 0.05). Details are provided in Table 3.

Table 3 Comparison of facial diagnostic features between the depression patients and the healthy controls (mean ± SD)

Factor	Healthy control (n = 87)	Depression patient (n = 93)	P value
Complexion L value	107.87 ± 12.16	94.79 ± 13.12	< 0.001
Complexion a value	135.59 ± 2.83	136.52 ± 2.75	0.029
Complexion b value	128.68 ± 2.86	132.85 ± 2.73	< 0.001
Gloss index	83.51 ± 7.11	73.88 ± 9.24	< 0.001
Low gloss index	6.94 ± 2.04	8.03 ± 2.18	< 0.001
Matte index	3.45 ± 1.30	5.21 ± 1.79	< 0.001
Labial L value	75.66 ± 13.87	59.89 ± 12.09	< 0.001
Labial a value	141.83 ± 3.90	137.46 ± 3.02	< 0.001
Labial b value	128.10 ± 2.46	130.02 ± 2.19	< 0.001

3.3 Construction and validation of the TCM spirit-expression integrated model for depression recognition

3.3.1 Accuracy comparison of depression recognition algorithms in depression patients and healthy controls

Figure 1 compares the performance of the five machine learning algorithms for depression classification. The XGBoost-based model demonstrated the best performance, achieving perfect classification (100% accuracy) for depression patients and 96.83% accuracy for healthy controls. Other algorithms showed varied accuracy: DT (92.59% for depression patients, 88.89% for healthy controls), BernoulliNB (88.89% for depression patients, 82.54% for healthy controls), and KNN (38.27% for depression patients, 71.43% for healthy controls). Notably, the SVM failed to discriminate between the two groups, classifying all samples as depressed (100% sensitivity, 0% specificity). Among the five evaluated TCM spirit-expression depression recognition models evaluated, the XGBoost algorithm demonstrated optimal diagnostic performance, significantly outperforming all comparative models (DT, BernoulliNB, SVM, and KNN).

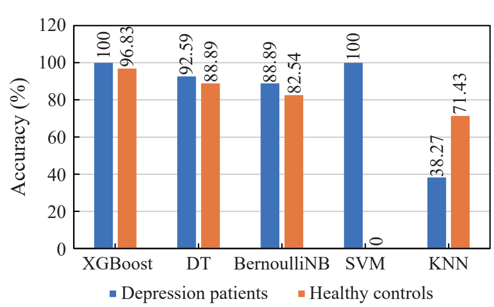

Figure 1 Recognition accuracy of machine learning algorithms in depression patients and healthy controls

3.3.2 Algorithmic performance comparison for depression classification

Comprehensive algorithm evaluation (Figure 2) revealed XGBoost's superior performance across all diagnostic metrics (accuracy: 98.61%, precision: 98.61%, recall: 98.61%, F1-score: 98.61%, AUC: 0.984), significantly exceeding DT, BernoulliNB, SVM, and KNN in depression recognition capability (all P < 0.01). The specific values for depression patients and healthy controls are provided in Figure 3.

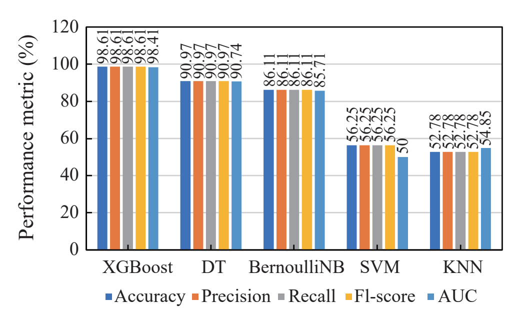

Figure 2 Algorithm comparison for TCM spirit-expression-based depression recognition models

To evaluate the generalization ability of the model on unseen data, this study employed the five-fold cross-validation. The dataset was randomly divided into five mutually exclusive subsets. Each time, four subsets were used for training and the remaining subset for testing, and this process was repeated five times. The confusion matrix presented in the figure represents the aggregated results across all test folds. Due to the exclusion of some original samples owing to unclear labels or missing features, only a portion of valid samples were retained. Among the confusion matrix corresponding to XGBoost, the true

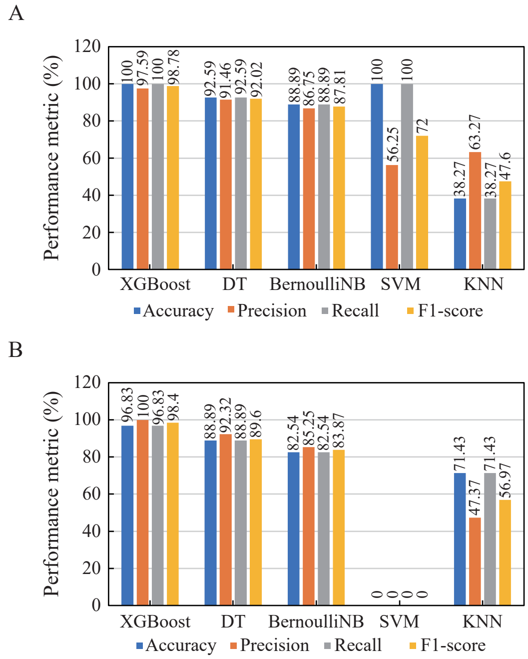

Figure 3 Model performance across depression patients and healthy controls for TCM spirit-expression-based depression recognition models

A, depression patients. B, healthy controls.

positives (TP) were 61, the false positives (FP) were 2, the false negatives (FN) were 0, and the true negatives (TN) were 81. This combined result provides a stable and comprehensive reflection of the model's classification performance on the test data (Figure 4).

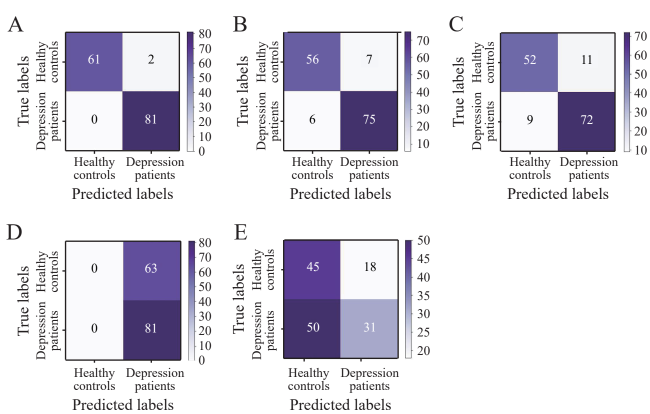

Figure 4 Confusion matrices of TCM spirit-expression-based depression recognition models across patients and healthy controls

A, XGBoost. B, DT. C, BernoulliNB. D, SVM. E, KNN.

3.3.3 Feature contribution analysis in the depression recognition model To visually illustrate the contribution of each selected feature to depression recognition, we applied SHAP to quantify and display their impacts. As shown in Figure 5, the top 15 risk factors are ranked by their mean absolute SHAP values, with the SHAP value on the *x*-axis reflecting the importance of each feature in the recognition model. The complexion b value demonstrated the highest predictive contribution (0.2641),

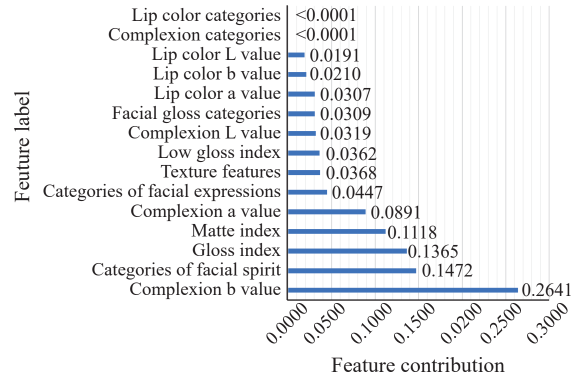

Figure 5 Feature contribution of the XGBoost model in the depression recognition model integrating the TCM spirit-expression features

followed by categories of facial spirit (0.1472), gloss index (0.1365), matte index (0.1118), and complexion a value (0.0891). Lower but significant contributions were observed for categories of facial expressions (0.0447) and texture features (0.0368). Notably, the combined categories of facial spirit and facial expressions, together with their associated texture features, accounted for 0.2287 of the total feature importance, while lip color categories showed minimal contribution.

4 Discussion

4.1 Main findings and integration of TCM theory with modern science

This study systematically analyzed facial images from 93 depression patients and 87 healthy controls, revealing distinct facial characteristic patterns in the depression group. Quantitative analysis demonstrated significant alterations across multiple color spaces and feature domains.

In the Lab color space, depression patients exhibited notably elevated complexion b values (indicating yellowness) across facial regions, along with decreased L values (lightness) and increased a values (redness). These findings bear similarities to a computational psychiatry study by TANISHIMA et al. [30], which quantitatively demonstrated that changes in facial chromaticity serve as reliable indicators for assessing psychological distress in clinical populations, including depression and catastrophic thinking.

Parallel changes were observed in lip color characteristics: depression patients displayed significantly decreased L and the values alongside increased b values, presenting as dark, greenish-yellow lips. These alterations may be linked to depression-associated inhibition of vasodilation [31], which can lead to peripheral microcirculation dysfunction, insufficient tissue perfusion, or reduced blood oxygen saturation [32]. Such physiological changes may ultimately contribute directly to darkening and cyanosis (bluish-purple discoloration) of the skin and mucous membranes—manifested in color space as reduced brightness and an increased yellow component.

A reduction in the facial gloss index was another prominent feature, quantitatively capturing the diminished skin luminosity characteristic of depressive states. This aligns with observations from animal studies in which fur quality in depression-model rats deteriorated significantly, appearing dull, coarse, and disordered [33]. Psychological conditions such as depression can affect skin health through neuroendocrine pathways—for example, via activation of the hypothalamic-pituitary-adrenal (HPA) axis—leading to decreased sebum secretion, impaired skin barrier function, and an overall decline in skin quality, thereby providing a pathophysiological basis for reduced skin luster [34]. Qualitative assessment of spirit further confirmed substantial intergroup differences, with patients in the depression group showing marked deficiency in spiritual manifestation, characterized by diminished vitality and reduced expressive dynamism.

The coordinated alterations in facial features provide quantitative validation for classical TCM observations. The elevated b value offers objective support for the "spleen deficiency leading to yellowish complexion" theory, while the reduced luster and altered spirit-related texture substantiate the concept of "impairment of spirit" [35]. Critically, the co-occurrence of these features objectively reflects the core TCM pathomechanism of "depression originating from heart and spleen disorders" [36] where spleen deficiency leads to malnourishment (quantified as yellowness and texture change), and heart disturbance manifests as diminished vitality (quantified as reduced luster) [37]. Thus, our multidimensional feature analysis bridges abstract TCM theory with measurable facial phenotypes.

Notably, while our color-based findings strongly converge with existing literature, our approach uniquely integrates these objective measurements with TCM diagnostic principles. Whereas most previous studies have focused primarily on Western medical interpretations of facial features in depression [38], our work establishes a novel bridge between quantitative facial analysis and TCM theory.

In contrast to these consistent color and vitality markers, expression features demonstrated limited discriminatory power in our analysis. This finding contrasts with certain earlier studies emphasizing dynamic expression analysis [39], but aligns with methodological reports highlighting the challenges of static image-based expression recognition [40]. This discrepancy likely reflects methodological constraints imposed by our standardized neutral-expression protocol rather than indicating the irrelevance of expressive features in depression. The requirement for a "relaxed expression" during image acquisition necessarily limited the capture of spontaneous expressive dynamics, thereby restricting the model's access to discriminative affective cues.

The comprehensive facial measurements obtained in this study provide an objective foundation for TCM diagnostic criteria and offer valuable markers for depression assessment. These quantifiable features not only bridge classical TCM observations with modern measurement techniques but also open promising avenues for developing objective diagnostic aids in psychiatric practice.

4.2 Model interpretability and feature contribution analysis based on TCM theory

In this study, multiple machine learning classifiers—XGBoost, SVM, KNN, BernoulliNB, and DT—were developed and compared for the diagnosis of depression based on facial expression and spirit fusion features. Under a stratified nested cross-validation framework and feature selection guided by TCM theory, XGBoost achieved the highest performance across multiple evaluation metrics.

Although XGBoost is a relatively complex ensemble-learning model, its application to structured biomedical datasets has become increasingly well-established. For instance, DUTTA et al. [41] compared XGBoost, SVM, and KNN on a small-sample breast cancer dataset and reported that XGBoost attained the highest accuracy. Similarly, YAQOOB et al. [42] applied XGBoost to high-dimensional, small-sample cancer gene-expression datasets and observed that it outperformed SVM and KNN. In another study, MIAO et al. [43] employed machine learning algorithms to extract high-dimensional features from MR images for predicting pathological staging in 139 breast cancer cases. Their results demonstrated that XGBoost yielded the best performance in a data structure analogous to that used in the present work.

The structure of our dataset-characterized by interactions among TCM features, potential nonlinearity, and a moderate sample size—aligns with that of the cited studies. In this context, simpler distance-based or linear-margin models such as KNN or SVM, may have limited ability to capture complex feature relationships. In contrast, appropriately regularized ensemble methods, including XGBoost, demonstrate a stronger capacity to model nonlinear interactions, which likely explains their superior performance in our analysis.

Feature selection is a critical step for optimizing predictive models, as it identifies a subset of the most informative features that best represent the essential characteristics of the dataset. In this study, we utilized the built-in feature importance analysis of XGBoost to evaluate the contribution of different facial diagnostic features in depression recognition [44]. The analysis revealed a clear hierarchy of feature importance that provides valuable insights into both the model's decision-making process and the underlying TCM pathophysiology.

Among all features evaluated, the complexion b value demonstrated the highest predictive importance, strongly supporting the classical TCM observation that "spleen deficiency leads to a yellowish complexion" theory [45]. This finding gains further relevance when considered alongside the concurrent increase in the a value and reduction in the L value, which together characterize the clinically recognized "dusky" and "lusterless" appearance often associated with chronic depression. These colorimetric alterations may reflect physiological disturbances including inflammation-mediated alterations in bilirubin metabolism [46], impaired cutaneous microcirculation, and the oxidative stress accumulation [47]—each representing a potential modern correlate of the TCM spleen deficiency pattern.

Features related to spirit and gloss emerged as the second most influential predictors, with our analysis revealing substantial intergroup differences in spiritual manifestation. In TCM theory, spirit-rooted in congenital essence and governing all vital activities—serves as a comprehensive indicator of health status, traditionally assessed through spiritual consciousness, facial luster, and ocular expression [48]. Depression fundamentally involves heart-spirit disturbance, in which chronic Qi stagnation and expressive overexertion progressively deplete heart-spleen function, ultimately leading to Qi-blood deficiency and impaired spiritual nourishment.

These pathophysiological changes manifest clinically as the characteristic facial dullness, diminished ocular brightness, and loss of vitality observed in depressive states. Quantitatively, our model captures this spiritual impairment through reduced gloss indices, which reflect diminished skin luminosity associated with Qi-blood deficiency, while spirit classification features effectively encode the characteristic loss of expressive dynamism. Together, these measures objectively substantiate the fundamental TCM diagnostic principle of "observing the spirit".

Regarding textural characteristics, periocular texture features demonstrated supplementary diagnostic value, ranking seventh in overall importance. The increased texture heterogeneity observed around the eyes physically manifests the chronic stress and sleep disturbances commonly associated with depression [49], while simultaneously aligning with the TCM concept that the spleen governs the fleshy orbiculus. The manifestation of texture changes in this region thus provides objective evidence for the TCM theory that spleen dysfunction leads to inadequate tissue nourishment.

Notably, expressive features showed relatively lower contribution weights, ranking 6th among the 15 diagnostic indicators. This finding should be interpreted primarily as a methodological consideration rather than a refutation of the importance of expression in depression. Our standardized data collection protocol, which required participants to maintain a "relaxed expression", necessarily limited the capture of spontaneous expressive dynamics, thereby restricting the model's access to discriminative affective cues. This insight clearly indicates that future research incorporating dynamic video analysis will be essential to fully leverage the diagnostic potential of expressive features.

The feature importance hierarchy established in this analysis provides robust, data-driven validation for core TCM theories regarding depression. The prominence of complexion, spirit, and periocular features objectively supports the pathological mechanism of "depression is rooted in the heart and spleen", while also demonstrating how contemporary computational methods can quantify traditional diagnostic observations.

4.3 Limitations and future perspectives

This study has several limitations. First, the model's specificity remains unverified due to the absence of disease control groups, such as patients with anxiety disorders or other chronic conditions. Second, all participants were recruited from a single center without stratification by depression severity, which may limit the model's generalizability and clinical precision. Third, although patients with confirmed comorbidities were excluded to maintain sample purity, the actual clinical populations often present with comorbid conditions. This gap between our study cohort and the general clinical population may restrict the immediate generalizability of the current model. Fourth, standard hyperparameters were employed, prioritizing the validation of the proposed analytical framework's feasibility over optimizing for peak predictive performance, a step that represents an important direction for model refinement following future sample size expansion. Future studies should validate the model in multi-center cohorts with severity stratification and incorporate disease control groups to verify depression-specificity and conduct external validation. Integrating dynamic video analysis and multimodal data including biochemical and inflammatory biomarkers will further enhance the model's robustness and clinical applicability. Additionally, applying these features to differentiate depression from other primary mental disorders, such as anxiety disorders, will help to expand their diagnostic utility in broader clinical scenarios.

5 Conclusion

This study established a depression recognition model integrating TCM "spirit-expression" facial features via machine learning. Among five algorithms, the XGBoost-based model performed best, identifying distinct facial feature differences between depression patients and healthy controls. These findings underscore unique advantages of TCM in depression diagnosis, as TCM facial features (spirit, expression, and texture) serve as effective markers and provide quantitative support for TCM diagnostic concepts. Integrating TCM facial features with machine learning offers a reliable depression screening approach, laying a methodological foundation for objective TCM evaluation and promising prospects in modern auxiliary diagnosis.

Fundings

General Program of National Natural Science Foundation of China (82474390), Construction Project of Pudong New Area Famous TCM Studios (National Pilot Zone for TCM Development, Shanghai) (PDZY-2025-0716), and Shanghai Municipal Science and Technology Program Project Shanghai Key Laboratory of Health Identification and Assessment (21DZ2271000).

Ethical approval

This study was approved by the Institutional Review Board of Shanghai University of Traditional Chinese Medicine (Approval No. 2024-2-14-08). All participants were aged ⩾ 18 years and provided written informed consent independently prior to study enrollment.

Author contributions

Minghui Yao and Rongrong Zhu: data curation, writing – original draft, and writing – review & editing. Qian Peng: formal analysis. Huilin Liu: data curation, methodology, and software. Xirong Sun: resources. Limin Gao: funding acquisition, project administration, and supervision. Fufeng Li: conceptualization, funding acquisition, project administration, and supervision. All authors approved the submission and take responsibility for this manuscript.

Competing interests

The authors declare no conflict of interest.

References

[1] JI J, WANG YW, DU SH, et al. A brief discussion on the application of "Shen" in acupuncture and moxibustion for the diagnosis and treatment of depression in "Huangdi Neijing". Journal of Traditional Chinese Medicine, 2023, 38(9): 1840–1844.

[2] World Health Organization. Depressive disorder (depression). Geneva: World Health Organization, 2025-08-29. Available from: https://www.who.int/news-room/fact-sheets/detail/depression.

[3] YANG YT, FU CW, WANG JY. Research progress on depression early warning methods primary health care in China. Primary Health Care in China, 2023, 37 (4) : 76–80.

[4] BRADLEY MM, LANG PJ. The International Affective Picture System (IAPS) in the Study of Emotion and Attention. Handbook of Emotion Elicitation and Assessment. Oxford: Oxford University Press, 2007: 29–46.

[5] ZHANG DN. Theoretical and experimental research on the correlation between emotional "anger" and pulse change characteristics. Beijing: Beijing University of Chinese Medicine, 2017.

[6] MULC D, VUKOJEVIC J, KALAFATIC E, et al. Opportunities and challenges for clinical practice in detecting depression using EEG and machine learning. Sensors, 2025, 25(2): 409.

[7] LEVIS B, BENEDETTI A, THOMBS BD. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. BMJ, 2019, 365: l1476.

[8] FRIED EI. The 52 symptoms of major depression: lack of content overlap among seven common depression scales. Journal of Affective Disorders, 2017, 208: 191–197.

[9] WRIGHT JH, OWEN JJ, RICHARDS D, et al. Computer-assisted cognitive-behavior therapy for depression: a systematic review and meta-analysis. The Journal of Clinical Psychiatry, 2019, 80(2): 18r12188.

[10] WILLIAMSON JR, YOUNG D, NIERENBERG AA, et al. Tracking depression severity from audio and video based on speech articulatory coordination. Computer Speech & Language, 2019, 55: 40–56.

[11] GIRARD JM, COHN JF, MAHOOR MH, et al. Social risk and depression: evidence from manual and automatic facial expression analysis. 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). IEEE, 2013. doi: 10.1109/FG.2013.6553748.

[12] YANG NT, ZHOU X, MA QY, et al. Objective parameter analysis of face-to-face diagnosis for patients with liver depression and spleen deficiency syndrome of depression. Chinese Journal of Traditional Chinese Medicine, 2023, 38(8): 3836–3839.

[13] WEI W. Research on depression recognition based on facial features. Lanzhou: Lanzhou University, 2020.

[14] LI Y, XU Y, XIA M, et al. Eye movement indices in the study of depressive disorder. Shanghai Arch Psychiatry, 2016, 28(6): 326–334.

[15] LIU XX. Exploration on posture and expression recognition in patients with liver Qi depression syndrome of depression. Jinan: Shandong University of Traditional Chinese Medicine, 2022.

[16] American Psychiatric Association. The Diagnostic and Statistical Manual of Mental Disorders. Fifth Edition. Arlington: American Psychiatric Publishing, 2013.

[17] TANG QS. Diagnostic criteria and treatment plans of TCM syndromes for depression. Journal of Beijing University of Chinese Medicine, 2011, 34(12): 810–811.

[18] Shanghai University of Traditional Chinese Medicine. A Standard Color Card Based on Tongue and Facial Images in Traditional Chinese Medicine and Its Application. Patent: 201910798364.1, 2022-03-29.

[19] DENG TT. Diagnostic Techniques in Traditional Chinese Medicine. Shanghai: Shanghai Science and Technology Press, 2006.

[20] ZHANG HK, SUN CY, QIAN P, et al. Research on the extraction and classification methods of eye features in traditional Chinese medicine. Chinese Journal of Traditional Chinese Medicine, 2018, 33(3): 886–889.

[21] ZHU RR. Research on classification and feature recognition of qualitative indicators of gaze and expression in traditional Chinese medicine. Shanghai: Shanghai University of Traditional Chinese Medicine, 2024.
[22] HARALICK RM, SHANMUGAM K, DINSTEIN I. Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics, 1973, SMC-3(6): 610–621.
[23] LI CD. Diagnostic techniques in traditional Chinese medicine. Beijing: China Publishing House of Traditional Chinese Medicine, 2016.
[24] LIU WZ, ZHANG F. The Induction and Evaluation of Emotions. Shanghai: Shanghai Science Popularization Press, 2022.
[25] ROSS QUINLAN J. C4.5: Programs for Machine Learning. California: Morgan Kaufmann Publisher, 1993.
[26] LI R, YANG JY, SONG SY, et al. Research on prediction of potential hyperlipidemia based on XGBoost algorithm. Electronic Technology and Software Engineering, 2021(2): 160–161.
[27] CUNNINGHAM P, DELANY SJ. K-nearest neighbour classifiers - a tutorial. ACM Computing Surveys, 2022, 54(6): 1–25.
[28] CRISTIANINI N, SHAWE-TAYLOR J. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge: Cambridge University Press, 2000.
[29] AMIRI S, AKBARABADI M, ABDOLALI F, et al. Radiomics analysis on CT images for prediction of radiation-induced kidney damage by machine learning models. Computers in Biology and Medicine, 2021, 133: 104409.
[30] TANISHIMA S, KOTANI Y, TAKEDA C, et al. Relationship between facial color changes and psychological problems associated with lower back pain. Medicina, 2022, 58(10): 1471.
[31] JOYNER MJ. Depression depresses vasodilation. Circulation Research, 2019, 124(4): 465–466.
[32] VAN AGTMAAL MJM, HOUBEN AJHM, POUWER F, et al. Association of microvascular dysfunction with late-life depression: a systematic review and meta-analysis. JAMA Psychiatry, 2017, 74(7): 729.
[33] ZHANG LW, WANG SY, GAO Y, et al. The neurotransmitter basis of negative attention bias: evidence from a rat model of depression. Journal of China Medical University, 2019, 48(9): 807–811.
[34] GUPTA MA, GUPTA AK. Psychiatric and psychological comorbidity in patients with dermatologic disorders. American Journal of Clinical Dermatology, 2003, 4(12): 833–842.
[35] MA YY, HAN J, LI FF, et al. Analysis of facial and tongue diagnostic image characteristics in patients with schizophrenia. New Chinese Medicine, 2025, 57(17): 73–77.
[36] ZHANG JP, FU X. Discussion on the TCM mechanism of acupuncture in treating depression combined with insomnia. Chinese Journal of Basic Medical Sciences in Traditional Chinese Medicine, 2025, 31(5): 889–891.
[37] DU LL, ZHANG DN, XUE M, et al. Research on the establishment and evaluation methods of animal models for traditional Chinese medicine syndromes. Journal of Nanjing University of Chinese Medicine, 2025, 41(1): 11–18.
[38] TADALAGI M, JOSHI AM. AutoDep: automatic depression detection using facial expressions based on linear binary pattern descriptor. Medical & Biological Engineering & Computing, 2021, 59(6): 1339–1354.
[39] HARATI S, CROWELL A, HUANG Y, et al. Classifying depression severity in recovery from major depressive disorder via dynamic facial features. IEEE Journal of Biomedical and Health Informatics, 2020, 24(3): 815–824.
[40] BARRETT LF, ADOLPHS R, MARSELLA S, et al. Emotional expressions reconsidered: challenges to inferring emotion from human facial movements. Psychological Science in the Public Interest, 2019, 20(1): 1–68.
[41] DUTTA M, MEHEDI HASAN KM, AKTER A, et al. An interpretable machine learning-based breast cancer classification using XGBoost, SHAP, and LIME. Bulletin of Electrical Engineering and Informatics, 2024, 13(6): 4306–4315.
[42] YAQOOB A, VERMA NK, AZIZ RM, et al. Optimizing cancer classification: a hybrid RDO-XGBoost approach for feature selection and predictive insights. Cancer Immunology, Immunotherapy, 2024, 73(12): 261.
[43] MIAO Y, TANG SY, ZHANG ZQ, et al. Application of deep learning and XGBoost in predicting pathological staging of breast cancer MR images. The Journal of Supercomputing, 2024, 80(7): 8933–8953.
[44] BENTÉJAC C, CSÖRGŐ A, MARTÍNEZ-MUÑOZ G. A comparative analysis of gradient boosting algorithms. Artificial Intelligence Review, 2021, 54(3): 1937–1967.
[45] LIU XJ, LI T. A brief discussion on the Zangfu organ differentiation of depression from the perspective of the five Zangfu organs governing the five vital organs. Global Chinese Medicine, 2021, 14(11): 1986–1989.
[46] JANGI S, OTTERBEIN L, ROBSON S. The molecular basis for the immunomodulatory activities of unconjugated bilirubin. The International Journal of Biochemistry & Cell Biology, 2013, 45(12): 2843–2851.
[47] BLACK CN, BOT M, SCHEFFER PG, et al. Is depression associated with increased oxidative stress? A systematic review and meta-analysis. Psychoneuroendocrinology, 2015, 51: 164–175.
[48] JIA HX, YIN DQ. The theory of “Five Divine Organs” in traditional Chinese medicine and the TCM locations of common mental symptoms. Modern Clinical Practice of Traditional Chinese Medicine, 2020, 27(5): 68–72.
[49] HUANG YJ, LI YN, SU T, et al. Effects of sleep quality, acute sleep deprivation, and napping on facial emotion recognition accuracy and speed. Nature and Science of Sleep, 2024, 16: 1637–1651.

(Editor-in-Charge Jie Deng)

融合中医神情特征的抑郁症机器学习识别模型

尧明慧[a†], 朱蓉蓉[a†], 钱鹏[a], 刘慧琳[a], 孙喜蓉[b], 高利民[b*], 李福凤[a, b*]

a. 上海中医药大学中医学院, 上海 201203, 中国
b. 上海市浦东新区精神卫生中心中医科, 上海 200124, 中国

【**摘要**】**目的** 本研究旨在融合中医"神与表情"诊断框架与机器学习算法，构建一个抑郁症识别模型。该模型致力于建立一种融合中医理论的早期抑郁筛查工具，从而将传统诊断原则与现代计算方法相连接。**方法** 研究纳入 2022 年 10 月 1 日至 2023 年 10 月 1 日在上海市浦东新区精神卫生中心就诊的抑郁症患者以及同期的上海中医药大学学生和老师作为健康对照组，采用 Xiaomi Pad 5 拍摄 3 – 10 s 视频，通过中医专家判读中医神与表情（5 名专家中至少 3 名专家判读一致以确定中医神与表情类别）。通过便携式中医智能分析与诊断设备采集基本信息、面部图像和问诊信息，运用 Open CV 计算机视觉库技术提取面诊特征。应用参数检验和非参数检验等统计分析方法对两组人群的基线资料和中医神与表情及面诊特征参数进行分析，比较中医神与表情及面部特征差异。运用极值梯度提升（XGBoost）、决策树（DT）、伯努利朴素贝叶斯（BernoulliNB）、支持向量机（SVM）和 k-近邻分类（KNN）五种机器学习算法，构建基于中医"望神与表情"特征融合的抑郁症识别模型，以准确率、精准率和受试者工作特征（ROC）曲线下面积（AUC）等指标评估模型性能。使用夏普利加性解释（SHAP）来解释模型结果。**结果** 本研究最终纳入 93 例抑郁症患者及 87 例健康人群，两组人群的基线资料无统计学差异（$P > 0.05$）。两组人群的中医神与表情特征及面部特征差异显示：（1）抑郁症组以面部少神、少光泽为主，相较于正常人群，抑郁症组面部画像向悲伤表情、面红、面黑及唇红偏移。（2）抑郁症患者面色 L 值、唇色 L 值、唇色 a 值、有光泽指数均小于正常人，面色 a 值、面色 b 值、唇色 b 值、少光泽指数、无光泽指数均大于正常人（$P < 0.05$）；（3）多模型结果显示，由 XGBoost 算法构建的联合中医"望神与表情"特征的抑郁症识别模型总体准确率达 98.61%，优于其他四种机器学习算法；（4）SHAP 可视化结果显示，在由 XGBoost 算法构建的识别模型中，面色 b 值、神的类别、有光泽指数、无光泽指数、表情与纹理特征对模型的贡献度较大。**结论** 本研究通过机器学习算法构建了基于中医"望神与表情"特征融合的抑郁症识别模型，该模型具有较高的识别准确率，为临床辅助诊断抑郁症提供了新的思路和方法。

【**关键词**】中医；神；表情；特征融合；抑郁症；识别模型