

<https://doi.org/10.1038/s44184-025-00164-4>

Emoface: AI-assisted diagnostic model for differentiating major depressive disorder and bipolar disorder via facial biomarkers

Check for updates

Jiahui Yu^{1,2,11}, Jingkai Chen^{3,4,11}, Yutong Zhang², Hailong Lyu^{3,4}, Tianyu Ma², Huimin Huang⁵, Zhong Wang^{3,4}, Xin Xu^{6,7,8} ✉, Shaohua Hu^{3,4,7,8,9} ✉ & Yingke Xu^{1,2,10} ✉

Affective disorders, including Major Depressive Disorder (MDD) and Bipolar Disorder (BD), exhibit significant mood abnormalities, making rapid diagnosis essential for social stability and healthcare efficiency. Traditional diagnostic solutions, including medical history collection and psychological assessments, often struggle to differentiate their similar clinical presentations, leading to time-consuming, laborious, and a high rate of misdiagnosis. Here, we propose Emoface, an AI-assisted diagnostic model that reads the emotional activities of faces in affective disorders. By analyzing videos from 353 participants exposed to various emotional stimuli, Emoface identified unique facial digital biomarkers distinguishing BD from MDD. Based on this, Emoface contributed to develop the first digital facial mapping for clinical and teaching use. In clinical practice with 347 patients, Emoface achieved precise diagnosis based on various facial visual signals, with accuracy rates of 95.29% for BD and 87.05% for MDD, offering a reliable face-based AI solution in a new era of affective disorders.

Affective disorders, mainly including Major Depressive Disorder (MDD) and Bipolar Disorder (BD), are among the most significant public health challenges worldwide. Precision diagnosing BD and MDD in clinical practice requires senior-trained psychiatrists. The diagnostic process involves: (1) conducting a structured interview using the MINI International Neuropsychiatric Interview (MINI)¹ to evaluate alcohol dependency before testing, (2) diagnosing the disorder based on the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) or the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-V) criteria² while excluding other major psychiatric disorders, a history of severe head trauma, or other serious neurological diseases, and 3) integrating all clinical scale data and patient history to make a differential diagnosis and provide a final diagnosis.

Due to similar clinical presentations and a lack of rapid diagnostic biomarkers, BD patients who have experienced depressive episodes are

often misdiagnosed as having MDD (approximately 40% to 50% according to existing literature). This misdiagnosis often results in significant delays in receiving appropriate treatment, up to 8 to 10 years of procrastination, further exacerbating the progression of conditions. In the United States, only 47.3% of primary care physicians are able to accurately identify cases of depression, and among these, only 33.6% are correctly documented. Moreover, the number of misdiagnosed cases exceeds the number of missed cases^{3,4}. Misclassifying BD as MDD and applying inappropriate treatment protocols can worsen patient prognoses, increasing hospitalization rates and healthcare expenditures. Hence, a precise diagnostic system that can efficiently discriminate between BD and MDD is essential to improving clinical care and management of patients with psychiatric disorders.

Recently, AI-driven digital solutions have shown demonstrated potential in improving diagnostic efficiency, especially between mental disorder phenotypes with similar underlying etiologies. By analyzing

¹Department of Biomedical Engineering, MOE Key Laboratory of Biomedical Engineering, State Key Laboratory of Extreme Photonics and Instrumentation, Zhejiang Key Laboratory of Intelligent Sensing and Advanced Medical Instrument, Zhejiang University, Hangzhou, China. ²Innovation Center for Smart Medical Technologies and Devices, Binjiang Institute of Zhejiang University, Hangzhou, China. ³Department of Psychiatry, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China. ⁴The Zhejiang Key Laboratory of Precision Psychiatry, Brain Research Institute of Zhejiang University, Zhejiang Engineering Center for Mathematical Mental Health, Hangzhou, China. ⁵Department of Psychiatry, The Third Affiliated Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, China. ⁶School of Public Health, the Second Affiliated Hospital of School of Medicine, Zhejiang University, Hangzhou, Zhejiang, China. ⁷Nanhu Brain-computer Interface institute, Hangzhou, China. ⁸The State Key Lab of Brain-Machine Intelligence, MOE Frontier Science Center for Brain Science and Brain-Machine Integration, Zhejiang University School of Medicine, Hangzhou, China. ⁹Liangzhu Laboratory, Zhejiang University School of Medicine, Hangzhou, China. ¹⁰Department of Endocrinology, Children's Hospital of Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou, China. ¹¹These authors contributed equally: Jiahui Yu, Jingkai Chen. ✉e-mail: xuxinsummer@zju.edu.cn; dorhushaohua@zju.edu.cn; yingkexu@zju.edu.cn

medical imaging and historical data, AI-powered tools can assist in identifying subtle emotional, behavioral, and linguistic expressions, supporting early and accurate identification of mental health illness^{5,6}. Therefore, exploring effective digital diagnostic solutions and digital biomarkers is key for rapidly diagnosing affective disorders.

Recent works have reported significant advancements in AI-driven digital diagnostic solutions. Facial motion has been validated as a feasible approach for the analysis of depression^{7,8}. Previous studies have successfully extracted and integrated facial appearance and motion features to characterize depressive symptoms⁹. By incorporating a Global Average Pooling layer into the regression framework, they have also achieved accurate prediction of depression severity and pinpointed the most salient facial regions associated with symptomatology¹⁰. However, most of these focused on depression rating, with few using facial activity features to classify BD and MDD. Additionally, due to privacy concerns and time/human costs, face-based depression data are limited. This leads to low discrimination by models when handling facial expressions of samples with similar depression scores, especially for difficult cases. AI-assisted analysis of emotional tendencies and intensities has been achieved by extracting speech and text information from social media and everyday conversations¹¹. Multimodal AI frameworks can process diverse digital signals in a collective and complementary manner¹², including language, facial expressions, heart rate, skin conductance¹³, and electroencephalograms^{14,15} to enhance diagnostic accuracy¹². However, multimodal interactions must simultaneously consider various patient physiological characteristics¹⁶, which poses challenges in diagnostic precision and efficiency. Additionally, jointly collecting multimodal information on patients to build personalized diagnostic solutions for BD and MDD is not easily applicable in clinical practice¹⁷.

In face-to-face clinical evaluation of affective disorder, experienced psychiatrists are adept at capturing microsecond changes in patients' facial activities, particularly when exposed to multiple emotional stimuli¹⁸. Facial activity changes, including muscle group movements, organ movements, and temporal dynamic changes, can quickly and accurately reveal the patient's mental state¹⁹. This provides a more rapid and accessible means of data collection compared to other biomarkers. AI analysis of facial movements, incorporating emotional features, holds promise for developing rapid digital screening technologies for bipolar depression, enhancing the accuracy of clinicians' diagnoses within short outpatient hours. These digital technologies can be considered a useful adjunct to existing clinical practice, thus driving the digital development of the mental health field²⁰.

In this paper, we collected the largest single-center facial dataset for affective disorders to date, comprising facial movement videos of 353 individuals under emotional stimuli. Leveraging this dataset, we developed a deep learning model for discriminating between BD and MDD that excels in reading the emotions reflected in patients' faces, called Emoface. Emoface analyzes patients' facial movements, including 68 facial key points, 16 facial regions, and 9 facial organs, to achieve rapid and precise diagnoses of emotional disorders and identify facial digital targets. Moreover, we created a standardized digital facial mapping for emotional disorders, offering a new digital scheme for clinical practice, patient privacy protection, and education. In clinical evaluations of 347 patients, the results demonstrated the potential of digital facial features in diagnosing emotional disorders. Emoface represents the diagnostic model that relies exclusively on patients' facial movements, enabling quick and effective BD and MDD diagnosis, promising to enhance mental health services globally.

Methods

Participants and clinical investigations

In this study, we established the largest facial movement dataset to date for patients with affective disorders. All experiments and protocols in this study were approved by the Institutional Review Board of the First Affiliated Hospital of Zhejiang University School of Medicine. All ethical regulations relevant to human research participants were followed in this study. We recorded the dynamic facial changes of 353 participants, including 158 with MDD, 128 with BD, and 67 healthy controls (HCs), before and after

watching emotionally stimulating videos. For clinical validation, 347 additional patients were recruited in China, comprising 130 BD patients, 132 MDD patients, and 85 healthy controls. The dataset includes diagnostic records by physician-provided, which were analyzed to examine the correlation between patients' responses during the diagnostic process and the medical procedures employed (Ethics Approval Number: #2019-1181 and #2021-382).

The diagnostic video data were analyzed using an emotion induction study method. Psychiatrists presented various emotion-eliciting videos to the subjects and recorded their facial expression changes throughout the process to facilitate a comprehensive analysis of facial features^{21–23}. In the field of emotion research, video stimuli are commonly used as effective tools for inducing emotions. In the pre-experiment, the self-assessment manikin was used to assess the subjects' emotional pleasure, arousal, and dominance following exposure to emotional video content, to ensure that these videos could play a role in emotional arousal. In the diagnosis of emotional disorders, emotion induction methods involve presenting five distinct video stimuli (happiness, anger, sadness, fear, and neutral) to elicit specific emotional responses and evaluate the participants' affective reactions. These videos are carefully selected by clinicians and initially tested on healthy controls to ensure their reliability in eliciting the target emotions. Subsequently, the facial expressions, physiological responses, and subjective emotional experiences of the participants are observed and quantified. By comparing the responses of patients with BD, MDD, and HCs to these stimuli, researchers can gain a deeper understanding of the nature of emotional regulation abnormalities²⁴.

All patients underwent structured interviews conducted by professionally trained mental health clinicians using MINI-International Neuropsychiatric Interview²⁵. Diagnoses were established in accordance with the criteria outlined in the DSM-IV. Exclusion criteria for the patient group included: (1) having severe mental disorders other than BD and MDD, such as schizophrenia or related spectrum disorders, intellectual disability, etc.; (2) having a history of severe head trauma (loss of consciousness for more than 5 minutes), current or past epilepsy, intracranial hypertension, or other serious neurological conditions; (3) having a history of alcohol or substance abuse/dependence within 6 months prior to testing; (4) determination by the researchers of unsuitability for participation or refusal to engage in the study. Healthy participants were required to confirm they had no history of mental illness and no family history of bipolar disorder or depression among first-degree relatives²⁶. They were also excluded if they had a history of severe head trauma (loss of consciousness for more than 5 minutes), current or past epilepsy, intracranial hypertension, or other serious neurological disorders²⁷. These videos not only provide scientific evidence for clinicians treating patients in diagnostic scenarios but also offer valuable biometrics for the development of our systems.

Creating visual-based face dataset for affective disorders

To accommodate real-time vision-assisted diagnostic scenarios, we utilized the transform module in the torchvision library for image normalization and enhancement. Video frames were upsampled from 720 × 576 pixels to 512 × 512 pixels, and the frame rate was set to 5 frames per second (FPS). We observed that the dynamic changes in video content directly influence the optimal frame rate selection. Specifically, in tests involving unipolar patients, where facial expression changes are relatively limited, the video scenes exhibit low dynamics. In such cases, a higher frame rate (e.g., 10 FPS) may lead to data redundancy without significantly improving model performance. Conversely, in scenes with rapid motion or pronounced emotional changes, particularly in bipolar patients under sad stimulation, facial expressions change more noticeably. In these instances, a higher frame rate can effectively capture key expression changes. To address these variations, we adjusted the frame rate settings of the video, testing rates of 5, 10, and 15 FPS. We found that at 5 FPS, the model's training and feature learning performance was superior, resulting in lower validation loss and faster processing times compared to using longer clips at higher frame rates. To ensure data quality, we designed a strict quality assessment module to

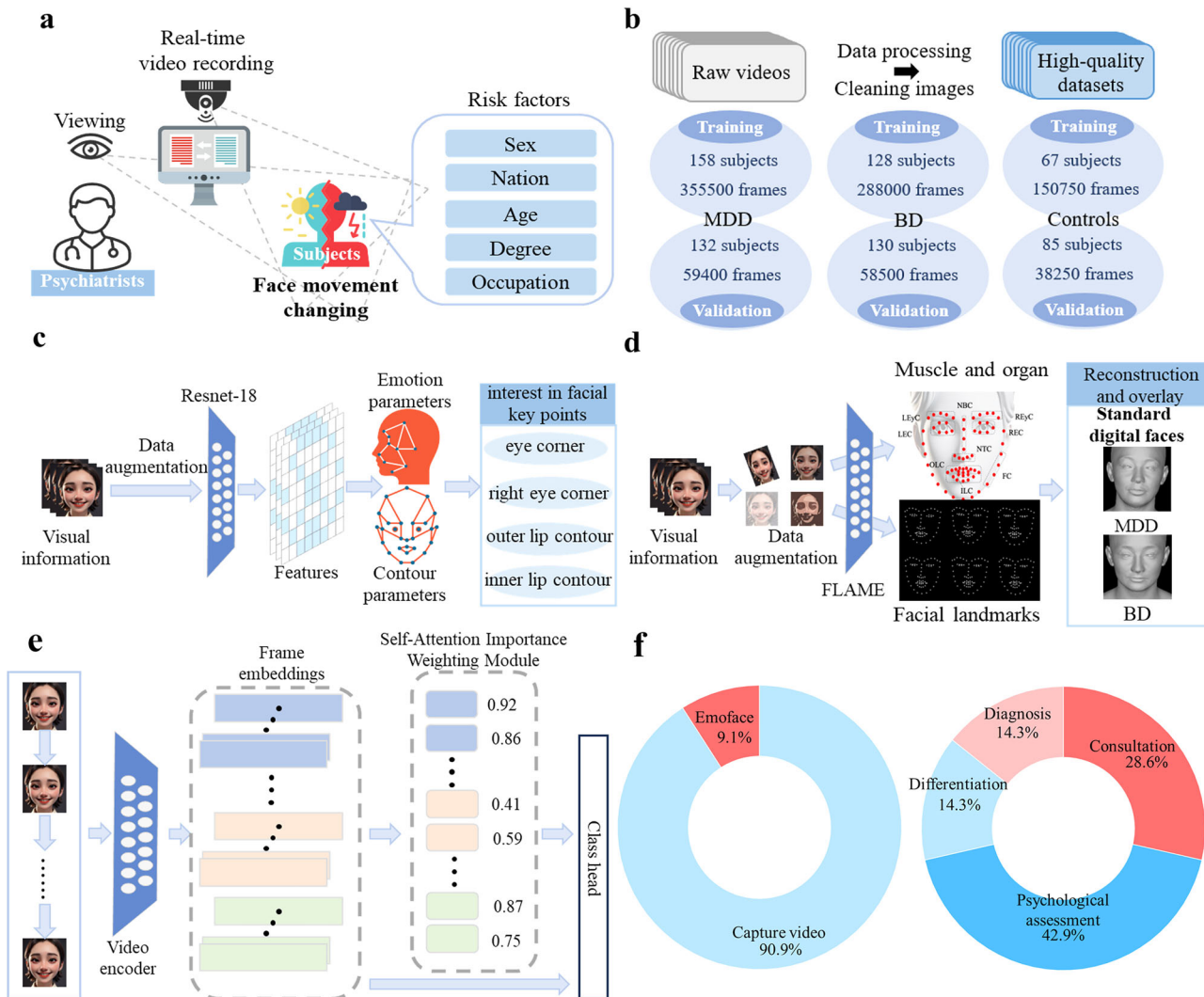


Fig. 1 | Overview of the work. a Graphical illustration of the face-to-face diagnosis, visual-based deep learning analysis, and data collection in clinical practice. **b** Overview of the collected facial datasets of MDD, BD, and HCs. **c**, Overview of diagnosing affective disorders and finding digital biomarkers with Emoface. **d** The

process of generating standard digital facial mapping with facial digital biomarkers in MDD and BD. **e** Proposed Emoface framework for automatic depression assessment with the attention calculation mechanism. **f** Comparison of Emoface and doctors in terms of diagnostic process and time.

analyze each frame and exclude images with severe facial occlusions. The final dataset comprises 355,500 frames of MDD patients, 288,000 frames of BD patients, and 150,750 frames of healthy controls. In clinical evaluation, Emoface tested 347 individuals using videos, including 59,400 frames of MDD (132 individuals), 58,500 frames of BD (130 individuals), and 38,250 frames of HCs (85 individuals) (Fig. 1a, b).

Emoface concept and implementation

The Emoface framework operates in a full-cycle loop, starting from raw image acquisition, progressing to digital biomarker identification, and culminating in disease classification. The implementation requires: 1) enhancing the quality and consistency of raw data through image processing modules; 2) differentiating various types of subjects and computing facial regions of interest; and 3) integrating multi-dimensional digital biomarkers to improve diagnostic comprehensiveness and reliability (Fig. 1c). In our implementation, facial movement information of affective disorders is processed through neural networks, generating feature maps. These feature maps are subsequently processed through activation functions to compute the output probability distribution. For regions of interest, scores from the probability distribution are used as decision parameters for the final diagnostic decision (Fig. 1d, e). Finally,

the subsequent framework response and optimization are triggered based on the final diagnostic decision.

Emoface excels at identifying features of interest in each frame and using them for diagnostic assistance. It effectively identifies multi-dimensional regions of interest in diseases and performs dimension integration, which enables it to have a unique diagnostic process and time (Fig. 1f). In this study, we first propose a concept: digital facial mapping, i.e., reconstructing 3D digital faces of MDD and BD patients based on key facial features. The digital face depicts the standard facial morphology of such patients in a generalized form (Fig. 1d).

Algorithm development

The diagnostic model is designed to classify emotional videos among BD, MDD, and HCs. The model is built on a deep residual network (ResNet-18), featuring a pre-trained feature extractor and a custom fully connected classifier. Data augmentation techniques, such as random sampling and random horizontal flipping, were applied to the input facial images to convert them into a format suitable for the model. During the training, we employed rank regularization loss to enhance the generalization ability²⁸. This involves sorting the samples in descending order by weight, classifying those with high weights into a high-confidence group, and those with lower

weights into a low-confidence group. The mean difference between these two groups is then calculated to focus on the high-confidence samples. Additionally, we implemented a dynamic relabeling strategy. Samples that may be misclassified are relabeled based on the difference between the prediction and label probabilities after a certain number of epochs. During the testing phase, we used digital biomarker features to assess the model's diagnostic performance. These digital biomarker features involved the extracted key points and feature matching methods.

Regarding the training strategy for the classification model, we first pre-trained the ResNet-18 model on the RAF-DB dataset. We then removed its last fully connected layer and replaced it with a new layer with 3D output to adapt to the classification diagnosis task. The 353 initial samples were pre-divided into training and validation sets at a 7:2 ratio. The model extracted features from video frames sampled at 5 FPS and classified them using this newly added layer. Additionally, we introduced an attention calculation mechanism, which computes attention weights using a linear layer and a Sigmoid function. During forward propagation, we applied a dropout method to prevent overfitting, ultimately outputting the attention weights and classification results. To accelerate model convergence, we initialized the parameters²⁹ of the convolutional layer, Batch Normalization layer, and fully connected layer using pre-trained weights. The model employed a cross-entropy loss function to measure the difference between the predicted results and the true labels. We used an exponential decay learning rate scheduler to dynamically adjust the learning rate throughout the training process^{30,31}. The model utilized a cross-entropy loss function to measure the discrepancy between predicted and true labels, combined with an exponential decay learning rate scheduler for dynamic learning rate adjustment. The initial learning rate was set to 0.01, with a momentum of 0.9 and weight decay of 1×10^{-4} . The model was trained for a total of 70 epochs. Within each training epoch, the training set was first used to compute loss and back-propagate to update model parameters. Subsequently, model performance was evaluated on the validation set by calculating accuracy and loss, with training strategies adjusted based on validation results³².

Generating facial landmarks

The backbone network is responsible for directly extracting features from input facial images and predicting key point coordinates. The auxiliary network uses these features to predict head pose (Euler angles), thereby improving key point accuracy under various pose conditions. During training, we employed the Adam optimizer to minimize the prediction error between the predicted and actual key point positions. The loss function combines Euclidean loss for key point coordinates with pose prediction loss, enhancing the adaptability to complex pose variations. To account for different lighting and environmental conditions, image preprocessing and augmentation techniques were applied throughout the training process. The implementation of this study is based on the PyTorch framework, with model parameters and training strategies configurable via command-line parameters. Detailed records of the training and evaluation processes are stored in log files, and the training process is visualized using TensorBoard.

We built on the original videos by utilizing the DLIB library for facial recognition and key point detection. A pre-trained model was employed to detect 68 key points on the face, marking critical facial regions such as the eyes, eyebrows, nose, mouth, and jawline. These key points are essential for conducting more advanced facial analyses, including expression recognition, age and gender estimation, and other biometric authentication tasks. By leveraging these 68 key points, researchers and developers can accurately capture the structural information of the face, facilitating detailed analysis and application of facial features.

Visualization

We employed Gradient-weighted Class Activation Mapping (GradCAM) technology to analyze the gradients with respect to the input image, identifying the areas of focus to generate facial key point positions and their corresponding gradient values. Specifically, we selected a target layer³³ within the pre-trained ResNet-18 model for gradient calculation. The facial

key point detector was used to identify the face position and key point coordinates within the image, and their effectiveness for the diagnostic model was evaluated through uniform manifold approximation and projection. Next, using GradCAM, the input image was preprocessed, and gradients were computed at the target layer. These gradients reflect the sensitivity to specific key point locations. For each detected key point, the corresponding gradient value was extracted from the generated gradient map and saved to a text file. Finally, facial key points were overlaid on the original images, and the annotated results were saved for visualization.

Generating standard digital faces

We utilized a deep learning-based 3D facial reconstruction model to record and analyze the facial contour parameters and emotional information. The 3D facial model is based on a geometric structure, which excludes external interference such as accessories and lighting, thereby enabling precise quantification of facial muscle movements and capturing subtle asymmetric expressions with high accuracy. Emoface is designed not only to achieve basic 3D facial reconstruction using these parameters but also to enhance the reconstruction by adding fine details such as facial wrinkles. The initial phase involves capturing the facial contour and emotional information parameters. Emoface processes these inputs to generate a basic 3D reconstruction, including its overall shape and structure, providing a foundational 3D facial model. Building upon this, Emoface employs advanced algorithms to identify and reconstruct finer details such as wrinkles and other subtle features. We fine-tuned the 3D facial reconstruction model to focus on the specific facial features of MDD and BD patients. For MDD patients, the model emphasizes movements around the inner corners of the eyes and the mouth. Conversely, for BD patients, the focus is on the movements of the outer corners of the eyes and the mouth. Finally, we created digital faces representing the typical facial features of MDD and BD patients by summarizing the facial contour parameters. During the validation phase, we synthesized facial videos of unknown patients and extracted embeddings using a facial encoder. We then calculated their cosine similarity to the embeddings of a standardized digital face, thereby demonstrating the generalizability of the digital reference face. Specifically, we weighted and averaged their facial contour parameters and superimposed a series of related emotional feature parameters. Through this methodology, Emoface achieves high-fidelity 3D facial reconstructions and provides valuable insights into the facial characteristics and emotional expressions unique to MDD and BD patients.

Model assessment

In this study, we adopted multiple evaluation indicators to comprehensively measure the performance of the Emoface model in diagnosing affective disorders. These indicators included accuracy (ACC), recall, precision, F1-score, and the area under the receiver operating characteristic curve (AUC). These metrics are calculated using a binary confusion matrix. True Positives (TP) and True Negatives (TN) represent the number of correctly predicted samples for affective disorder patients and healthy controls, respectively. False Negatives (FN) and False Positives (FP) denote the number of misclassified samples for affective disorder patients and healthy individuals, respectively. The formulas are as shown in (1).

$$\begin{aligned} \text{ACC} &= \frac{TP+TN}{TP+FP+TN+FN} \\ \text{Recall} &= \frac{TP}{TP+FN} \\ \text{Precision} &= \frac{TP}{TP+FP} \\ F1 &= 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \end{aligned} \quad (1)$$

Results

Emoface helps find new visual-based digital biomarkers

Here, we answer the question: how does Emoface distinguish between MDD and BD? We visualized the facial activities and regions, that Emoface prioritizes on collected 353 participants, aiming to find effective digital

biomarkers (Fig. 2a–c). We divided the face into 16 regions (Fig. 2a). The main findings reveal that emotional states fluctuate with the alternating movements of the inner and outer corners of the eyes. For BD diagnosis, regions 5 and 8, i.e., the outer corners of both eyes, exhibited higher activation values, showing that Emoface places significant emphasis on these movements for identifying BD patients. This prominent activation represents a distinguishing feature of the complex dynamic expressions typical of BD patients (Fig. 2d). In contrast, for MDD patients, regions 6 and 7, i.e., the inner corners of both eyes, displayed higher activation values, with regions 5 and 8 following closely. The results show that Emoface demonstrates heightened sensitivity to subtle inner eye corner movements in MDD patients, whereas outer eye corners exhibit reduced expressivity (Fig. 2d). For the healthy controls, Emoface recorded more evenly distributed activation values across regions 5, 6, 7, 8, 14, and 15, showing that the facial expressions of healthy individuals are more neutral and broadly distributed (Fig. 2d).

As an alternative representation to recording facial movements, we divided the face into nine organ groups, each morphed by activating its associated muscle group (Fig. 2b). We analyzed the relationships and activation intensities among different facial groups. To refine facial organ movements, we further examined the interest in specific facial key points. For BD, Emoface identified the left eyebrow contour (LEyC, 16.92%), right eyebrow contour (REyC, 15.09%), and outer lip contour (OLC, 12.84%) as regions showing significant emotional changes. More intuitively, by combining facial gradient maps, we observed that areas with high gradient values in BD patients are concentrated at the junction of the eyebrows and eyes, and around the mouth (Fig. 2c). The results show that Emoface focuses on the joint movements of these regions to capture the complex emotional dynamics characteristic of BD. For MDD, Emoface primarily focused on LEyC (16.56%), REyC (15.28%), and Inner Lip Contour (ILC, 12.09%). High gradient values at the junction of the eyebrows and inner corners of the eyes, as well as the outer lip contour, also show the emotional states of MDD patients (Fig. 2c). ILC (14.16%) and OLC (14.25%) play crucial roles in identifying MDD. We also found that high gradient values, indicating emotional fluctuations in MDD, are predominantly distributed around the eyes, mouth, and forehead. The attention to facial features LEyC, REyC, and ILC was 13.56%, 13.28%, and 12.09%, respectively (Fig. 2e). In healthy controls, the eyes, mouth, and eyebrows serve as key emotional cues, with gradient values distributed more evenly, reflecting natural and diverse facial expressions (Fig. 2e).

Based on the facial movement data of these 353 participants, Emoface has identified digital biomarkers from specific facial regions and organ groups, which are used to differentiate individuals with MDD, BD, and healthy individuals. We also present all patients' digital biomarkers collected in real-world clinical practice. (Supplementary Figs. 1–3).

Emoface helps generate standard feature-wide facial mapping for distinctive affective disorders

AI-driven digital human technology has been widely applied in the medical field, offering the ability to simulate human appearance, language, and behavior to facilitate friendly human-computer interactions. In this study, we employ Emoface to generate 3D digital faces for MDD and BD. The digital faces emphasize key digital biomarkers and prominent facial features, creating new face-based disease maps for affective disorders. Through scanning facial movements and transforming the feature parameters into digital models, we reconstruct their 3D faces. Initially, we utilize emotional data from 286 patients, detecting digital biomarkers in each frame through Emoface and overlaying the regions. Following this, we document standardized facial contour parameters utilizing weighted average computations. The computations work on the emotional parameters and facial contour parameters that have been extracted from Emoface. Finally, by integrating digital biomarkers with facial contour and emotional parameters, we produce standardized face-based disease maps for MDD and BD (Fig. 3). We also recorded 347 patients' digital faces in clinical practice. (Supplementary Figs. 1–3). These digital facial profiles aim to capture

individual patient features while preserving clinical distinctiveness. We anticipate that the digital faces will offer innovative solutions for clinical practice, enhance patient privacy protection, and serve as valuable resources for medical education in the future.

Emoface helps diagnose affective disorders in clinical practice

Here, we deployed Emoface in real-world clinical settings to diagnose affective disorders across 120 cases. We assessed the diagnostic performance using facial videos, digital markers, and 3D standard faces (Fig. 4a–d). The results demonstrate that Emoface achieved robust diagnosis by analyzing visual signals from faces, particularly excelling in identifying BD, which presents diagnostic challenges in clinical practice. For BD identification, Emoface delivered optimal performance when analyzing facial videos, achieving 95.38% ACC and 95.41% F1-score for BD cases, significantly superior to its MDD detection rates of 85.61% ACC and 85.62% F1-score (Fig. 4b). This performance gap reflects BD patients' more pronounced facial activation patterns, particularly around the outer eye corners which serve as reliable biomarkers. For MDD detection, explicit digital marker extraction yielded the best performance. Emoface achieved 87.12% ACC and 87.12% F1-score for MDD classification (Fig. 4c), demonstrating that explicit biomarker extraction effectively captures the subtler facial expressions characteristic of depression while reducing diagnostic ambiguity. To evaluate the clinical utility of the 3D standard facial model, we extracted embeddings from aggregated videos and measured their similarity using the cosine similarity metric. Emoface maintained stable performance, achieving average ACC and F1 scores of 88.93% and 88.96% for patients with BD and MDD, respectively (Fig. 4d). Although slightly inferior to video analysis, this approach provides a scalable solution for scenarios where real-time video capture is impractical, without over-relying on any single feature type. Notably, the 3D standard facial model is optimized as a target model for unique facial biomarkers specific to BD and MDD, enabling more precise characterization of disorder-specific facial features and thereby enhancing differential diagnosis of BD and MDD in clinical practice. Besides, in the collected dataset, the AUROC values for all three classes exceeded 0.95. This indicates that the model has a relatively strong ability to recognize the facial movement features of different categories. The AUROC values for MDD and BD were 0.97, showcasing the robust capability to distinguish between these similar conditions (Fig. 4a).

Emoface improves face representations and visual interpretability

Here, we explore the interpretability of facial encoding and representation modeling using 347 clinical cases. For face videos, the facial encoder first computes high-dimensional embeddings (Fig. 5a). These embeddings are then reduced in dimensionality and visualized using uniform manifold approximation and projection (UMAP) (Fig. 5b–d). Notably, the facial encoder effectively differentiates facial regions between patients with BD and MDD. It demonstrates clear region-specific separation within key anatomical areas, including the inner and outer eye corners, nose, left and right cheeks, and mouth (Fig. 5b). Specifically, the diagnostic approach using the inner and outer eye corners as primary biomarkers shows distinct separation in the test set (Fig. 5c). The test set forms three well-defined clusters, each representing significant differences in facial feature space among BD, MDD, and healthy controls (HCs). For the standard 3D facial models, diagnosis remains effective through feature matching between aggregated patient facial images and the standard 3D face, due to the clear separation of BD and MDD in the feature space (Fig. 5d). These results indicate that Emoface enhances the representational capability of facial features in emotional disorder patients.

Discussion

We analyzed a longitudinal scientific dataset of affective disorders from 353 individuals and proposed a deep learning-assisted diagnostic solution called Emoface. The dataset was collected over 19 months, capturing the facial

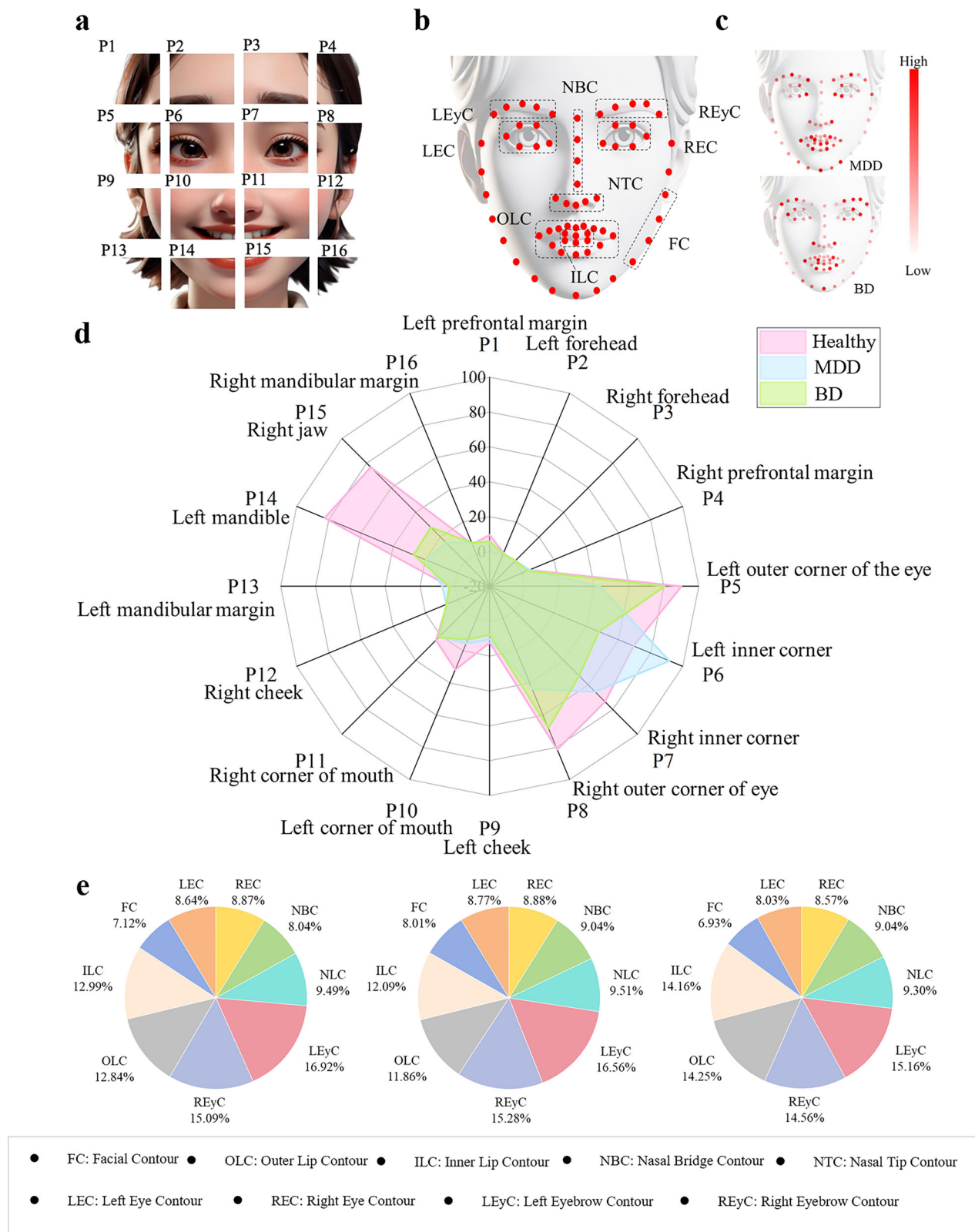


Fig. 2 | Digital biomarkers of affective disorders. **a** Graphical illustration of 16 regions of interest in faces. **b** Graphical illustration of facial landmarks and organs associated with muscle group movements. **c** Emoface's attention to facial landmarks

in MDD and BD. **d** Results of Emoface's attention to face regions in BD, MDD, and HCs, respectively. **e** Results of Emoface's attention to face organs in BD, MDD, and HCs, respectively.

movements (muscle groups and organs) before and after the participants watched emotion-eliciting videos. In clinical tests conducted on 347 patients, Emoface achieved diagnostic accuracy rates of 95.38% for BD and 85.61% for MDD, finding unique digital facial biomarkers for respective

conditions. Furthermore, we generated 3D standard digital faces of affective disorders based on facial features. The digital analysis of facial movements in patients presented in this study can also serve as a novel supplement to other existing traditional diagnostic solutions.

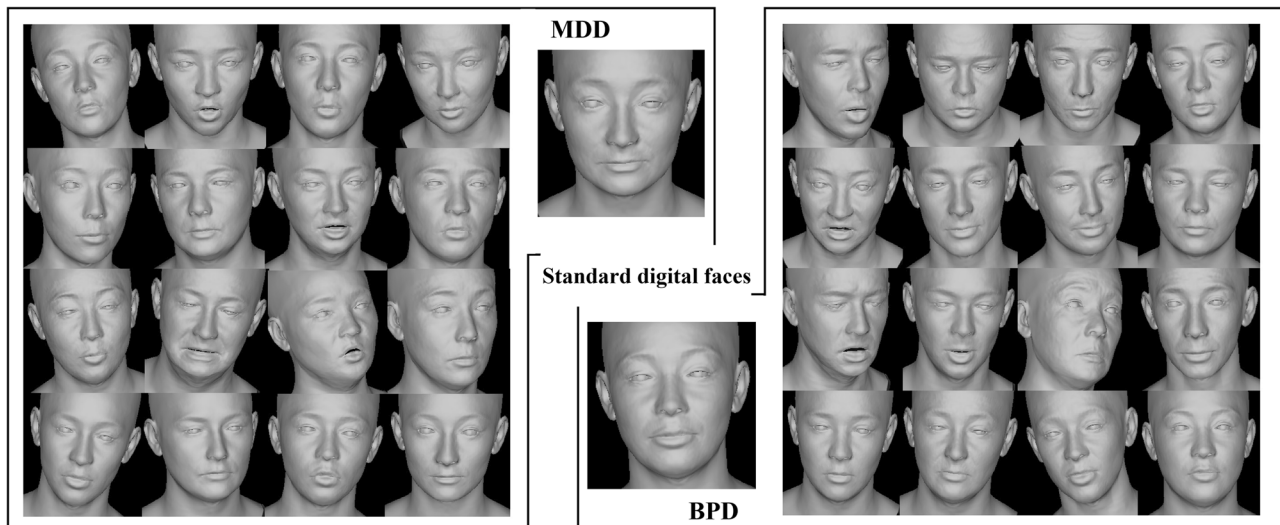


Fig. 3 | Standard digital facial mapping of MDD and BD. 32 patients with MDD and BD were randomly selected for 3D face reconstruction. 3D standard digital faces of affective disorder were generated by superimposing digital biomarkers.

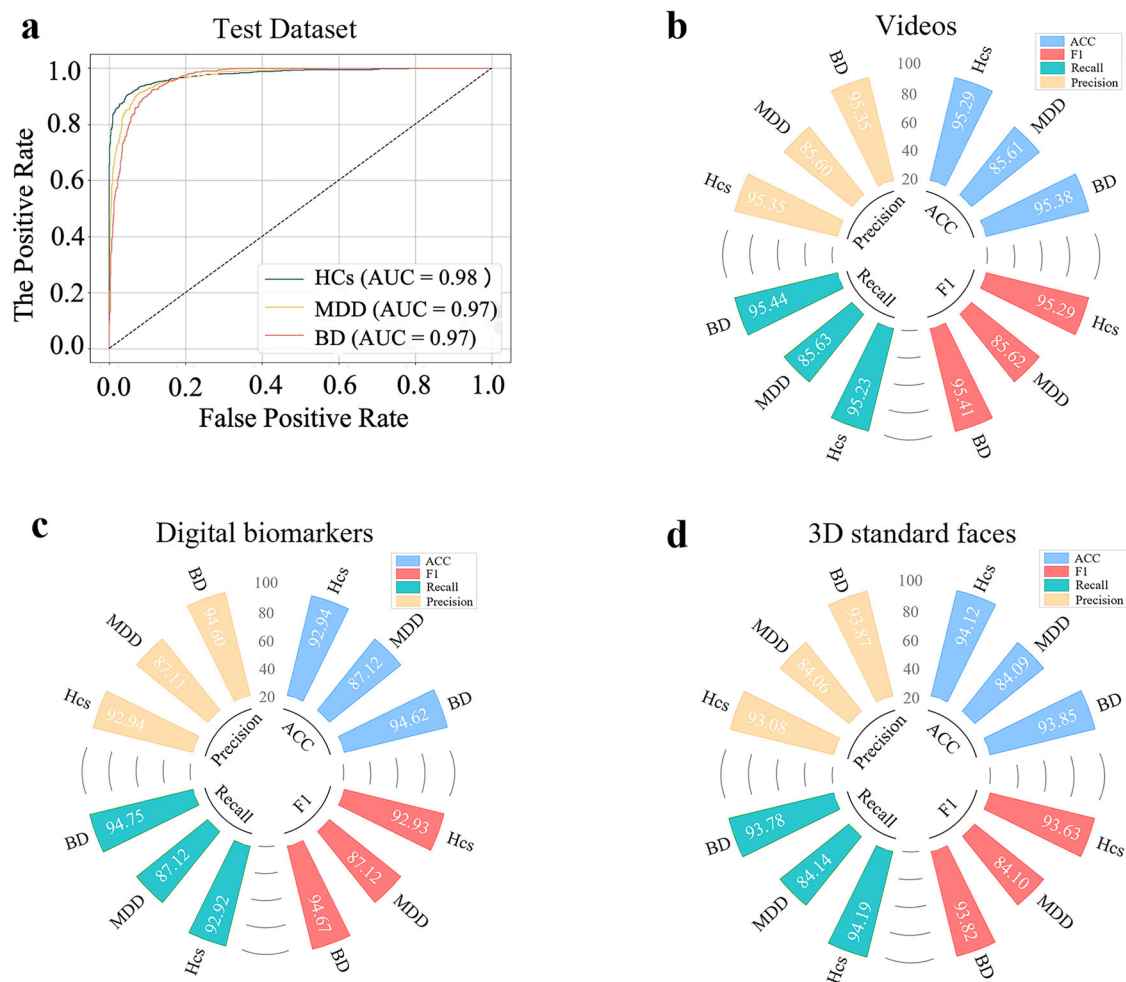


Fig. 4 | Emoface diagnoses affective disorder in real-world clinical settings. **a** ROC curves for testing the collected dataset. **b** Performance of Emoface using face videos. **c** Performance of Emoface using digital biomarkers. **d** Performance of Emoface using 3D standard faces.

In clinical practice, most psychiatrists diagnose affective disorders using structured interviews, medical history, and biochemical markers. Interestingly, experienced psychiatrists also observe changes in facial movements in response to various emotional stimuli. While many AI

models can analyze facial movements, most overlook facial details and the reconstruction of subtle expressions³⁴. Besides, there remains room for improvement in data requirements, detail capture, deeper facial information collection, generalizability, and computational resources³⁵. In the study, we

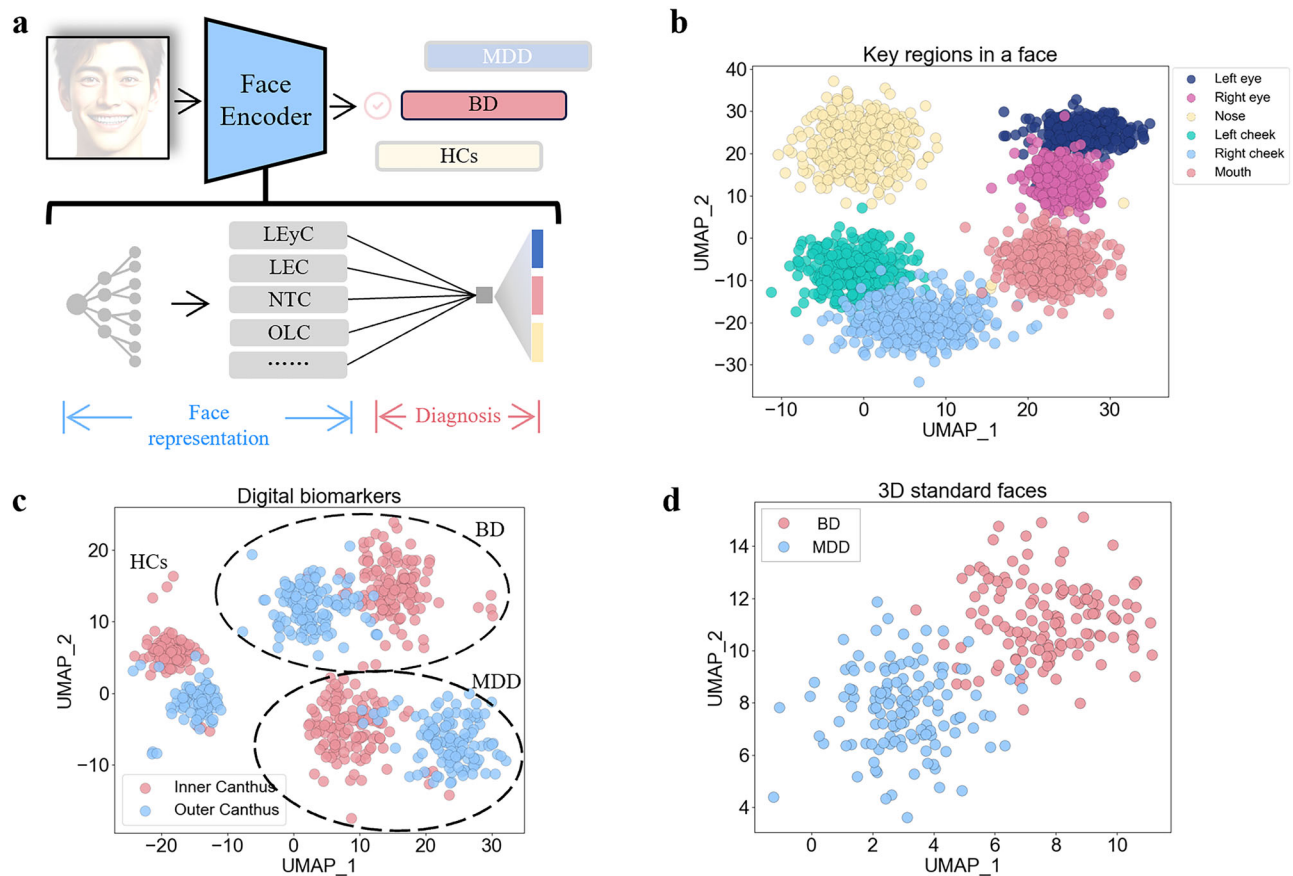


Fig. 5 | Face encoder and embedding analysis. **a** Graphical illustration of the face encoding process in the diagnostic model. **b** Image embeddings of key regions generated by the face encoder. **c** Image embeddings of the inner and outer eye corners generated by the face encoder. **d** standard face image embeddings generated by the face encoder.

focus on specific digital facial biomarkers³⁶, particularly organ groups and key points, to develop a novel deep-learning approach, which provides a more intuitive and dynamic diagnostic perspective. This approach helps identify characteristic patterns associated with subtle facial changes during emotional fluctuations, such as abnormalities. Emoface can also analyze facial movements, including 68 key points, 16 regions, and 9 organs, to achieve digital facial generation for individuals with affective disorders. We have integrated facial movement analysis with the generation of 3D standard digital faces, forming a comprehensive diagnostic system for emotional disorders: In the context of hospital outpatient services, facial expressions of patients can be captured via cameras. Compared with other biomarkers, this provides a faster and more accessible means of data collection. Subsequently, by combining traditional clinical consultation methods, artificial intelligence can be utilized to analyze this information in real-time. This approach enhances the accuracy of clinicians' diagnosis of emotional disorders within the limited time frame of outpatient services and reduces the incidence of missed diagnoses.

This study has several limitations. First, analyzing complex micro-expressions, such as the combination of happiness and surprise, poses a challenge for the Emoface. These expressions involve coordinated, rapid, and nuanced activations across multiple muscle groups, constituting complex and highly variable neural activation patterns. We anticipate that increasing the volume of patient facial data and implementing stringent data quality filtering processes will effectively mitigate this issue. Additionally, the current standard digital face, generated from existing feature parameters, has not yet achieved optimal realism. We anticipate that expanding the set of facial parameters, such as emotional parameters, will enhance the preservation of skin texture, the natural continuity of muscle dynamics, and the fidelity of expression rendering. Incorporating a broader range of clinical manifestations into the evaluation framework will further improve the

retention of finer details in facial reconstruction. In the future, digital faces can serve as a starting point for more advanced diagnostic tasks, such as real-time tracking of emotional changes in clinical diagnoses and early warnings of emotional risks. Moreover, since our study utilized only Han Chinese faces, the facial expression markers used may lack cultural diversity. Given that the expression and interpretation of facial cues can vary significantly across different ethnic and cultural groups, expanding the dataset to include diverse populations will be an important direction for future work. Finally, more powerful deep learning models, such as visual foundation models, will be needed in the future. We expect that designing generative AI based on the digital targets identified by Emoface will further improve performance.

Data availability

Pre-trained dataset RAF-DB public dataset is available at "<http://www.whdeng.cn/RAF/model1.html#dataset>". The use of private datasets strictly adheres to the privacy protection policies of the hospital and relevant ethical review guidelines to ensure patient privacy and security. The use of this privacy dataset was approved for use by the hospital (Ethics Approval Number: #2019-1181 and #2021-382) and complies with all relevant laws and regulations. The datasets used in the study are not all publicly available. We distribute small batches of emotionally disturbed face data (incl. 10 MDDs, 10 BDs, 5 HCs) at "https://drive.google.com/drive/folders/10jzFBapCTnDG_h2_nnS5QBU_oB0lbKYT?usp=sharing", which supports the plots and other finding of the study. Because of privacy protection, the corresponding feature information files of the dataset were provided. And the feature parameters in it are represented as NumPy arrays. This feature information includes vertices, shape parameters, expression parameters, and pose parameters. More samples for reasonable academic evaluation can be requested by contacting the corresponding author. All requests will be reviewed promptly by the author and processed according to

departmental guidelines. All codes used to implement the diagnostics of affective disorders and obtain biomarkers are available from "<https://github.com/hvp3100/Emoface>". Diagnostic model weights are available at "<https://drive.google.com/file/d/12qmNE0gCcpflSTSmP4iGRqkP-lv0ahfn/view?usp=sharing>".

Received: 26 February 2025; Accepted: 1 October 2025;

Published online: 18 October 2025

References

- Rodrigues, G. S. et al. Diagnosis of bipolar and major depressive disorders: The appropriateness of MINI compared to the clinical interview in a sample of patients with mood disorders in tertiary mental health care. *J. Psychiatr. Res.* **169**, 341–346 (2024).
- Livne, O. et al. Agreement between DSM-5 and DSM-IV measures of substance use disorders in a sample of adult substance users. *Drug Alcohol. Depend.* **227**, 108958 (2021).
- Pascual, J. C., Arias, L. & Soler, J. Pharmacological management of borderline personality disorder and common comorbidities. *CNS Drugs* **37**, 489–497 (2023).
- Tomasetti, C. et al. Treating depression in patients with borderline personality disorder: clinical clues on the use of antidepressants. *Ann Gen Psychiatry* **23**, 21 (2024).
- O'Connor, R. C. et al. Gone Too Soon: priorities for action to prevent premature mortality associated with mental illness and mental distress. *Lancet Psychiatry* **10**, 452–464 (2023).
- Olawade, D. B. et al. Enhancing mental health with Artificial Intelligence: Current trends and future prospects. *J. Med. Surg. Public Health* **3**, 100099 (2024).
- De Sario, G. D. et al. Using AI to detect pain through facial expressions: a review. *Bioengineering (Basel)* **10**, 548 (2023).
- Alvarez Casado, C., Lage Cañellas, M. & Bordallo López, M. Depression recognition using remote photoplethysmography from facial videos. *IEEE Trans. Affect. Comput.* **14**, 3305–3315 (2023).
- Zhou, X., Jin, K., Shang, Y. & Guo, G. Visually interpretable representation learning for depression recognition from facial images. *IEEE Trans. Affect. Comput.* **11**, 542–552 (2020).
- Zhu, Y., Shang, Z., Shao, Z. & Guo, G. Automated depression diagnosis based on deep networks to encode facial appearance and dynamics. *IEEE Trans. Affect. Comput.* **9**, 578–584 (2017).
- Saha, P. et al. On the rise of fear speech in online social media. *Proc. Natl. Acad. Sci. USA* **120**, e2212270120 (2023).
- Tutun, S. et al. An AI-based decision support system for predicting mental health disorders. *Inf. Syst. Front.* **25**, 1261–1276 (2023).
- Ghosh, R. & Sinha, D. Human emotion recognition by analyzing facial expressions, heart rate and blogs using deep learning method. *Innov. Syst. Softw. Eng.* **20**, 499–507 (2024).
- Jui, S. J. J. et al. Application of entropy for automated detection of neurological disorders with electroencephalogram signals: a review of the last decade (2012–2022). *IEEE Access.* (2023).
- Liu, Q. et al. The feature, performance, and prospect of advanced electrodes for electroencephalogram. *Biosensors (Basel)* **13**, 101 (2023).
- Wu, W., Li, L., Li, Z., Sun, J. & Wang, L. J. Extensible integrated system for real-time monitoring of cardiovascular physiological signals and limb health. *Adv. Mater.* **35**, 2304596 (2023).
- Leichsenring, F. et al. Borderline personality disorder: a comprehensive review of diagnosis and clinical presentation, etiology, treatment, and current controversies. *World Psychiatry* **23**, 4–25 (2024).
- Hu, B., Tao, Y. & Yang, M. Detecting depression based on facial cues elicited by emotional stimuli in video. *Comput. Biol. Med.* **165**, 107457 (2023).
- Malgaroli, M., Hull, T. D., Zech, J. M. & Althoff, T. J. T. P. Natural language processing for mental health interventions: a systematic review and research framework. *Transl. Psychiatry.* **13**, 309 (2023).
- Smith, K. A. et al. Digital mental health: challenges and next steps. *BMJ Ment. Health* **26**, e300670 (2023).
- Cheong, J. H. et al. Py-Feat: Python Facial Expression Analysis Toolbox. *Affect. Sci.* **4**, 781–796 (2023).
- Gupta, T., Osborne, K. J., Nadig, A., Haase, C. M. & Mittal, V. A. J. P. Alterations in facial expressions in individuals at risk for psychosis: a facial electromyography approach using emotionally evocative film clips. *Psychol. Med.* **53**, 5829–5838 (2023).
- Karnati, M. et al. Understanding deep learning techniques for recognition of human emotions using facial expressions: A comprehensive survey. *IEEE Trans. Instrum. Meas.* **72**, 1–31 (2023).
- Pinto, A. M. et al. Emotion regulation and the salience network: a hypothetical integrative model of fibromyalgia. *Nat. Rev. Rheumatol.* **19**, 44–60 (2023).
- Lai, T. et al. Scaling up Global Mental Health Psychological Services with AI-based Large Language Models. *arXiv.* (2023).
- Zhang, H., Chen, J. & Fang, Y. Functional alterations in patients with bipolar disorder and their unaffected first-degree relatives: insight from genetic, epidemiological, and neuroimaging data. *Neuropsychiatr Dis. Treat* **19**, 2797–2806 (2023).
- Kluger, B. M. et al. Palliative care to support the needs of adults with neurological disease. *Lancet Neurol* **22**, 619–631 (2023).
- Yan, L., Shi, Y., Wei, M. & Wu, Y. Multi-feature fusing local directional ternary pattern for facial expressions signal recognition based on video communication system. *Alex. Eng. J.* **63**, 307–320 (2023).
- Reyad, M., Sarhan, A. M. & Arafa, M. A modified Adam algorithm for deep neural network optimization. *Neural. Comput. & Applic.* **35**, 17095–17112 (2023).
- Diawati, P., Gadzali, S. S., Abd Aziz, M. K. N., Ausat, A. M. A. & Suherlan, S. The role of information technology in improving the efficiency and productivity of human resources in the workplace. *J. Teknol. Dan sist. info. Bis.* **5**, 296–302 (2023).
- Wu, Y. & Liu, L. Selecting and composing learning rate policies for deep neural networks. *ACM Trans. Intell. Syst.* **14**, 1–25 (2023).
- Liu, X. et al. Emotion classification for short texts: an improved multi-label method. *Humanit Soc. Sci. Commun.* **10**, 1–9 (2023).
- Zmolikova, K. et al. Neural target speech extraction: An overview. *IEEE Signal Process. Mag.* **40**, 8–29 (2023).
- Adyapady, R. R. & Annappa, B. J. M. S. A comprehensive review of facial expression recognition techniques. *Multimed. Syst.* **29**, 73–103 (2023).
- Sajjad, M. et al. A comprehensive survey on deep facial expression recognition: challenges, applications, and future guidelines. *Alex. Eng. J.* **68**, 817–840 (2023).
- Fan, D.-P. et al. Facial-sketch synthesis: A new challenge. *Mach. Intell. Res.* **19**, 257–287 (2022).

Acknowledgements

The authors would like to thank the Key R&D Program of Zhejiang (2025C01104, 2025C02108, 2024C03098 and 2025C02109), the Zhejiang Provincial Natural Science Foundation (LZ23H180002, LQ23F030001 and LGC22H090001), the National Natural Science Foundation of China (62406280, 72274170 and 82571735), the Cao Guangbiao High-tech Development Fund (2022RC009), the Key Projects of Hangzhou Science and Technology Bureau (20231203A13), the National Key Research and Development Program of China (2023YFC2506200).

Author contributions

Conception and design: J.Y., S.H., Y.X.; Acquisition of data: J.C., H.L., Z.W., H.H., S.H.; Analysis and interpretation of data: J.Y., J.C., Y.Z., H.L., T.M., Z.W., Y.X.; First draft of manuscript: J.Y., J.C., Y.Z., S.H., Y.X.; Revision of manuscript: J.Y., J.C., Y.Z., T.M., S.H., Y.X. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at

<https://doi.org/10.1038/s44184-025-00164-4>.

Correspondence and requests for materials should be addressed to Xin Xu, Shaohua Hu or Yingke Xu.

Reprints and permissions information is available at

<http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025