Chapter

Advances in Emotion Recognition: Link to Depressive Disorder

Xiaotong Cheng, Xiaoxia Wang, Tante Ouyang and Zhengzhi Feng

Abstract

Emotion recognition enables real-time analysis, tagging, and inference of cognitive affective states from human facial expression, speech and tone, body posture and physiological signal, as well as social text on social network platform. Recognition of emotion pattern based on explicit and implicit features extracted through wearable and other devices could be decoded through computational modeling. Meanwhile, emotion recognition and computation are critical to detection and diagnosis of potential patients of mood disorder. The chapter aims to summarize the main findings in the area of affective recognition and its applications in major depressive disorder (MDD), which have made rapid progress in the last decade.

Keywords: emotion recognition, computational modeling, machine learning, depressive disorder

1. Introduction

Making computers capable of emotional computing was first proposed by Minsky (one of the founders of artificial intelligence) of the MIT. In his book *The Society of Mind* he proposed that "The question is not whether intelligent machines can have any emotions, but whether machines can be intelligent without emotions" [1]. Picard [2] proposed the concept of affective computing (AC) in 1995. Her monograph "Affective Computing" published in 1997 defined affective computing as "calculation related to, derived from or capable of emotions." She divided the research content of affective computing into nine aspects: mechanism of emotion, acquisition of emotion information, recognition of emotion pattern, modeling and understanding of emotion, synthesis and expression of emotion, application of emotion computing, interface of emotion computer, transmission and communication of emotion, and wearable computer. Among these aspects, the practical research of emotion recognition is largely based on theories of mechanism of emotion and acquisition of emotion information.

The mechanism of emotion is based on phenomenal and mechanistic views of emotion. The phenomenal views typical involved two approaches: discrete and dimensional views of emotion. The former proposed that emotion can be labeled as a limited set of basic emotions which could be combined into complex emotions. This method is problematic because the labels for emotions may be too restrictive to reflect complex emotions. Additionally, these labels may be culture dependent which could not reflect common substrates of different affective labels.

The latter proposed that emotions can be distributed in a multidimensional space which continuously evolves. Two common dimensions are valence (pleasantness) and arousal (activation level). The emotion recognition algorithms using emotion representation based on emotional labels are intuitive which are ambiguous for computer processing. Additionally, recognition of emotion pattern involves the classification of emotional data according to a large group of labels. For these reasons, researchers developed a number of dimensional model of emotions, such as Russell's circumplex model, Whissell's evaluation-activation space model, and Plutchik's wheel of emotions [3].

According to the mechanistic views of emotion, emotion pattern recognition not only relies on semantic labels but also physiological signals which originate in the peripheral nervous system (PNS) and central nervous system (CNS) dynamics [3]. (1) *The PNS emotion patterns*. The PNS included the autonomic and the somatic nervous systems (ANS and SNS). According to Schachter and Singer's peripheral theories of emotion (or cognition-arousal theory), people assess their emotional state by physiological arousal. Emotion states are inherent in these physiological dynamics and feasibly recognized by using PNS physiological data, according to the work from the lab led by Picard. Ekman and colleges provided the first evidence of PNS differences (including hand temperature, heart rate, skin conductance, and forearm tension) among four negative emotions [4]. However, their algorithms are based on intentionally expressed emotion and are user dependent, which may restrict generalization to other users [5]. (2) *The CNS emotion patterns*. The large majority of computational models of emotion stem from appraisal theory of emotion, which emphasized the CNS process of emotion. Frijda criticized the arousal theory of emotion and proposed that awareness of autonomic responding is not prerequisite for emotional experience or behavior. The differentiation of the emotions is explained as the result of the sequential appraisal for affective stimulus. Scherer suggests that there may be as many emotions as there are different appraisal outcomes. Thus there exists the minimal set of appraisal criteria necessary to the differentiation of primary emotional states. However, it should be noted that physiological changes is not only determined by appraisal meaning but also by factors outside of the appraisal or emotion realm. Therefore, there is not adequate evidence for consistent and specific PNS response during emotional episodes [6].

Practically, the acquisition of emotional information is required for emotion recognition. Emotional information characteristics included a variety of physiological or behavioral reactions concurrent with emotional state changes, including internal and external emotional features. (1) *Internal emotional information*. It refers to physiological reactions that cannot be detected from the outside of human body, such as the electrical or mechanical/chemical output of human brain activity (EEG), heart muscle activity(ECG, heart rate, pulse), skeletal muscle activity (EMG), breathing activity (respiration), and blood vessel activity (blood pressure, hemangiectasis). (2) *External emotional information*. It refers to the reactions that can be directly observed from the appearance, such as facial expression, speech, and posture. The extraction of common features for highly individualized emotion information constitutes the fundamental basis of emotion recognition. A great amount of features could be extracted from internal and external emotional signals, by calculating their mean, standard deviation, transformation, wave band power and peak detection, and others.

2. Methods for emotion recognition

The main methods for emotion recognition involve the following emotion indexes: (1) emotional behavior, namely, facial expression recognition, speech

emotion recognition, and posture recognition (see Sections 2.1–2.3); (2) physiological pattern, which means objective emotional index after measuring PNS and CNS physiological signals (see Section 2.5); and (3) psychological measures and multimodal emotion signals, such as textual information and multimodal emotion information (see Sections 2.3 and 2.6).

2.1 Facial expression recognition

Faces may be one of the most important methods for visual communication of emotion. Though started from the 1970s, facial expression recognition is the most studied field in natural emotions machine recognition, especially in the USA and Japan, wherein studies on facial expression recognition have grown to be a hotspot of AI emotion recognition. In 1971, American psychologists Ekman and Friesen categorized facial expression into six types: anger, disgust, happiness, fear, surprise, and sadness. They also established the Facial Action Coding System (FACS) in 1978 [7], which is the earliest research of facial expression recognition. Facial expressions were deemed as observable indicators of internal emotional states, which make emotion differentiation possible.

Currently, the most-used facial expression databases included Ekman's FACS and its updated version, automated facial image analysis (AFIA) developed by Carnegie Mellon University, Japanese female expression database JAFFE and its expansion set in Japan ATR Media Information Science Laboratory, Cohn-Kanade expression database, CK+ expression database, and Rafd facial expression database established by CMU Robotics Research Institute, USA. Common facial expression picture libraries in China include the USTC-NVIE image library [8], the CFAPS facial emotion stimulating materials [9], and the Chinese facial expression intensity grading picture library [10].

The facial expression recognition included the following steps: (1) facial image acquisition, (2) image preprocessing, (3) feature extraction, and (4) emotion classification (**Table 1**).

Apart from Ekman's discrete emotion model, facial expression recognition was also conducted under the other emotional models (such as dimensional model). Ballano et al. proposed a method for continuous facial affect recognition from videos based on evaluation-activation 2D model proposed by Whissell [11].

Processes	Sub-processes and related work
Facial image acquisition	Facial images are obtained from images and videos, including static expressions and dynamic expressions
Image preprocessing	Face detection and positioning, face adjustment, editing, scale normalization, histogram equalization, dimming, light compensation, homomorphic filtering, graying, Gaussian smoothing
Feature extraction	(a) Static image. Gabor wavelet transformation, local binary patterns (LBP), scale-invariant feature transformation (SIFT), discrete cosine transformation (DCT), regional covariance matrix. (b) Dynamic image. Optical flow method, difference image method, feature point tracking, model-based method, elastic graph matching.
Emotion classification	Canonical correlation analysis, sparse representation classification, expert rule-based method[a]

[a] *An expert rule-based method classifies the emotion according to a set of "if-then" statements based on expert experience or sampling with rule acquisition algorithm.*

Table 1.
Facial expression recognition process.

The evaluation dimension defines the valence of emotion, while the activation dimension defines the action tendencies (e.g., active versus passive) under the emotional state. Their model extended the emotion information to continuous emotional trajectory.

Micro-expressions are quick, unconscious, and spontaneous facial movements that occur when people experience strong emotions. The duration of micro-expression is about 1/25 to 1/2 s. The fleeting micro-expression has small movement and does not appear in the upper face and the lower face at the same time, so it is quite difficult to observe and recognize correctly. Therefore, the collection and selection of micro-expression data sets are very important. Micro-expression recognition requires (1) the image acquisition and preprocessing of face image, (2) the detection of micro-expressions from the face and the extraction of its features, and (3) classifying and recognizing the categories of the micro-emoticon. Different research teams have developed different automatic micro-expression recognition systems and established databases.

Polikovsky et al. [12] explored the 3D gradient histogram method for feature extraction of facial micro-expressions in video sequences based on the Polikovsky expression library. They proposed a new approach to capture micro-expression using 200fps high-speed camera. Shreve et al. [13] established the USF-HD database and applied the optical flow method for automatic micro-expression recognition research. They developed a method of automatically spotting continuously changing facial expressions in long videos. The University of Oulu in Finland developed the spontaneous micro-expression corpus (SMIC) and SMIC2. Yan et al. [14] improved the micro-expression elicitation paradigm and developed the Chinese micro-expression database CASME. Later on they further expanded the sample number, improved the frame rate and image quality of CASME, and created CASMEII. They differentiated full suppression of facial movements from self-perceived suppression of facial movements. The micro-expressions were elicited in a well-controlled laboratory context and had high temporal resolution (200 frames/s). The best performance is 63.41% for 5-class classification.

2.2 Speech emotion recognition

As the easiest, most basic, and direct way of information communication, speech contains rich emotional information. Speech could not only convey semantic information but also reveal speaker's emotional state, for instance, a person may have a voice with high volume, heavy tones, and accelerated speed when getting angry, but sullen intonation and slow speed when feeling sad. Therefore, in order to make the computer understand people's emotions better and interact more naturally with people, it is necessary to study speech emotions. Speech emotion recognition is widely applied in man-machine interaction, such as automatic customer service system, which can transfer emotional users to manual service [15]; it monitors the driver's emotional fluctuations based upon his speech speed and volume to remind him of staying calm, thus preventing him from a car accident [16]; it helps the disabled to speak [17]; and it detects emotional state of patients with mental disorders based upon context analysis [18].

Most of the studies use prosodic features as characteristic parameters of speech emotion recognition. For example, Gharavian et al. [19] extracted parameters such as fundamental frequency, resonance peak, and Mel coefficient and then analyzed the correlation among them. The obtained 25-dimensional vectors were classified by FAMNN classification algorithm to gain a more credible emotion recognition result. Devi et al. [20] summarized speech signal preprocessing techniques, common short-term energy, MFCC features, and their applications in speech emotion

recognition. Zhang et al. [21] use a multilayer deep belief network (DBN) to automatically extract the emotional features in speech signals, piece together consecutive multi-frame speeches to form an abstract high-dimensional feature, use features trained by the deep belief network as the input end of the extreme learning machine (ELM) classifier, and ultimately establish a speech emotion recognition system. Zhu et al. [22] propose a track-based space-time spectral signature speech emotion recognition method and obtain relatively accurate results. Liu and Qin [23] study the application of speech emotion recognition in manned space flight, establish a stress emotion corpus, and build speech emotion recognition model and software through feature extraction and Gaussian mixture model (GMM) to verify the accuracy of speech emotion recognition.

The emotional speech features extracted in the abovementioned study are mostly targeted at personalized speech emotion recognition, while the feature extraction for non-personalized speech emotion recognition is still a challenge. Recent efforts have been made toward development of large corpus [24]. The current speech emotion recognition study is limited by lack of unified, public and standard mandarin emotion corpus, as well as an authoritative and unified standard for building emotion corpus. Many researches are conducted based on self-recorded databases which vary in terms of age, gender, number of participants, text information, and the scale of the final corpus, making it difficult to compare between different research results. Furthermore, most of the studies are conducted based on discrete emotions, taking into consideration limited emotional dimension corpus.

2.3 Posture emotion recognition

Posture refers to the expressional actions of other parts of human body than face. It can coordinate or supplement speech content and effectively convey emotional information. Postures can be divided into body expression and gestures. Body expression is one of the ways to express emotions. People would have different postures under different emotional state, such as belly laugh when happy, arched shoulders when scared, and being fidgeted when nervous. Postures such as raising hands and akimbo can express individual emotions. People may have different postures at different emotional state and level, hence it is possible to analyze and predict emotional state by observing different expressions and intensity of the expression. Researchers have pointed out at early times that posture and movements can only reflect intensity of emotion, but not the essence and type of the emotions. Later, some put forward that posture is conducive to the expression of emotional intensity although it cannot reflect accurately emotional state. Some scholars have studied the ability of subjects to understand six basic postures, the subjects were expressionless throughout the test, and the result showed that posture can be used to identify certain emotional state, such as sorrow and fear.

Generally, there are two posture recognition methods: (1) recognizing affective content of daily behavior through analysis and (2) using the temporal and spatial characteristics of gestures (such as the rhythm, amplitude, and strength of the motion) to analyze the affective content. For example, Castellano et al. [25] proposed a method for recognizing emotions based on human motion indicators (such as amplitude, velocity, and mobility) and establishing emotional models with image sequences and motion test indicators; Bernhardt and Robinson [26] used segmentation techniques to quantify high-dimensional motion into a set of simple motion data, extract motion features, and pair them with corresponding emotions; Liu et al. [27] classified the body movement and combined motion and velocity parameters for weighting function calculation to identify the emotion expressed by certain movement. Shao and Wang [28] extracted two 3D texture features by processing

the image sequences of body movement and used this as a basis for emotion classification. The recognition rate can reach 77.0% in experiment which tests seven common natural emotions in the FABO database.

The posture recognition process mainly includes four steps: motion data acquisition, preprocessing, motion feature extraction, and emotion classification. *Firstly*, motion data collection. Generally, there are two types of motion data collection methods: (1) contact type which is a wearable device embedded with various sensors, such as electronic gloves and data shoe covers, and (2) noncontact type, which generally obtains image information through the camera. The contact recognition technology has high equipment cost, uncomfortable user experience, and goes against the objective of natural man–machine interaction. *Secondly*, data preprocessing. This generally includes human body detection, image denoising, image segmentation, image binarization processing, time window, filtering processing, and others. Among them, human body detection mainly includes basic image segmentation, background difference method, interframe difference method, optical flow method, and energy minimization method. *Thirdly*, motion feature extraction. Generally speaking, motion features can be divided into four categories: (1) static features which include size, color, outline, shape, and depth; (2) dynamic features which include speed, optical flow, direction, and trajectory; (3) spatiotemporal features which include spatiotemporal context, spatiotemporal shape, and spatiotemporal interest points; and (4) descriptive features which include scenes, attributes, objects, and poses. There are three types of most-used methods for motion feature extraction, namely, time domain analysis, frequency domain analysis, and time-frequency domain analysis. *Fourthly*, emotional classification. Other classifiers than the commonly used ones are dynamic time warping, dynamic programming, potential Dirichlet distribution, probabilistic latent semantic analysis, context-free grammar, finite state machines, conditional random fields, and others.

2.4 Textual emotion recognition

Emotions are not exactly linguistic constructs. However the most convenient to emotion is through language. With the advent of social media, social media platforms are becoming a rich source of multimodal affective information, including text, videos, images, and audios. One of them is textual analysis. Affect recognition from text analysis is often used for a public opinion mining. The process of text recognition contains four steps: material collection, text preprocessing, feature extraction, and emotion classification. (1) The first step is material collection. Web crawlers are commonly used to collect materials from blogs, e-commerce sites, and news sites. (2) The second step is text preprocessing, which includes word segmentation, part-of-speech tagging, tag filtering, affix trimming, simplification and replacement, and so on. (3) The third step is feature extraction. Main text features involve words, phrases, n-gram, concepts, and others. Words containing general features can be automatically extracted, while others need to be identified by human efforts before creating emotional glossary. Other methods used are frequent pattern mining techniques and associated rule mining techniques. (4) The fourth step is emotion classification. In addition to some commonly used classifiers, it also includes central vector classification, maximum entropy, emotion-based words labeling, and word frequency-weighted statistics.

Domestic researches on text recognition mainly center around emotion recognition of social platforms such as microblog. For example, Hao et al. [29] proposed a microblog emotion recognition method based on wording features of microblogs and verified its validity. Hao et al. [30] proposed a classification method based on

supervised learning for the classification and prediction of emotional polarity in microblogs, and the accuracy of the experimental analysis reached 79.9%. Huang et al. [31] proposed a multifeature fusion-based microblog theme and emotion mining model TMMMF and verified its validity; Zhang et al. [32] proposed a joint model of microblog emotion recognition and emotion incentive extraction based on neural network. The experiment shows that the F value of the model in the emotion incentive extraction task is 82.70% and the F value in the emotion recognition task is 74.74%.

2.5 Physiological model recognition

William James [33] proposed that emotions derive from peripheral physiological responses. Kreibig [34] examined the patterns of autonomic nervous system activity under different emotions, showing the specificities in different physiological responses. For example, fear would cause accelerated heart rate and respiratory rhythm and strengthened galvanic skin response. The theory confirms the role of autonomous physiological activities in emotional expression but ignores the role of the brain center in emotions. In 1929, Cannon questioned James's theory and came up with the Cannon-Bard theory (also known as the thalamus theory) with Bard. According to this theory, emotions and their corresponding physiological changes occur simultaneously, both of which are controlled by the thalamus, and the central brain determines the nature of emotions, which affirms the central nervous system's role in regulating and controlling emotions. In conclusion, the occurrence of emotions is accompanied by certain degree of physiological activation of the central and peripheral nervous system. This provides a theoretical basis for studying emotion recognition in different physiological patterns.

Early studies mainly focused on the PNS physiological signals such as skin temperature, blood pressure, electrocardiogram, electromyography, respiratory action, galvanic skin response, and blood volume fluctuation for emotion recognition. Picard et al. [35] collected four physiological signals of galvanic skin response, blood volume fluctuation, electromyographic signal, and respiratory action under different emotional states and reached 81% in terms of recognition accuracy for eight emotions. Kim and Andre [36] developed a short-term monitoring emotion recognition system based on physiological signals of multiple users. They used support vector machine (SVM) to classify and identify four emotions including sadness, depression, surprise, and anger, achieving a classification rate at 95%. Yan et al. [37] collected a variety of physiological signals through multipurpose polygraph MP150: used Fisher, k-NN, and other intelligent algorithms for feature extraction and analysis; and identified six basic emotional states with recognition rate being at 60–90%. Li et al. [38] proposed emotion recognition based on recursive quantitative analysis of physiological signals. They extracted 10 sets of nonlinear features from the recursive graphs of skin conductance signals, myoelectric signals, and respiratory signals and achieved higher emotion recognition rate. Jin et al. [39] used the updated LSTSVM for emotion recognition based on the physiological signals of electroencephalography, skin conductance, myoelectricity, and respiration and obtained higher recognition accuracy.

In recent years, with the development of neurophysiology and the rise of brain imaging technology, CNS brain signals have attracted the attention of researchers and been used in emotion recognition because of their high temporal resolution and strong functional specificity. In the early stage of the study, the most common measurement index was electroencephalogram (EEG). Some scholars pointed out that the frontal brain asymmetry is closely related to emotional valence. Studies have shown that high-frequency parts of EEG can reflect people's emotional and

cognitive states and the γ and β bands can better tell the change of emotional state than the low-frequency band [40]. Jie et al. [41] realized the recognition of high and low arousal and high and low pleasure through nonlinear feature sample entropy. Duan et al. [42] used differential entropy in machine classification learning for emotion recognition, and the classification accuracy rate was up to 84.22%. It is shown that as a nonlinear EEG feature, differential entropy shows higher classification efficiency. Later, some scholars combined spontaneous physiological signals with EEG and used comprehensive information to improve the recognition rate [43, 44].

However, the EEG acquisition process is relatively complicated and often has the interference with external noise and electromyography. The cerebral blood oxygen parameter measurement method based on functional near-infrared spectroscopy (NIRS) is gaining greater popularity in emotion recognition because of its high portability, insensitivity to noise and motion, and high possibility for long-term continuous measurement. Tai and Chau [45] extracted the time domain features of prefrontal signals during affective states to identify positive and negative emotions elicited by emotional pictures. The recognition rate of 13 subjects was within the range of 75.0–96.67%.

The most critical steps in emotion recognition based on physiological signals are signal preprocessing, feature extraction and optimization, and classification identification.

1. **Emotion signal preprocessing**. This step mainly retains valid data segments during emotion induction process at its highest level and then removes noise and artifacts from the signal. The artifact removal methods mainly include filtering, normalization, independent component analysis, and so on. (a) Filters with different frequency band parameters, such as adaptive filters and Butterworth filters, are commonly used for denoising physiological signals, such as smoothing filtering of the galvanic skin to remove high-frequency glitch. (b) Normalization could reduce the adverse effects of baseline individual differences on emotion recognition [46]. (c) Independent component analysis or principal component analysis may remove electro-oculogram and artifacts [47].

2. **Feature extraction**. There are four main types of features: time domain, frequency domain, time-frequency, and nonlinear features.

    a. **Time domain**. Time domain feature extraction is found first and relatively simple. It obtains information in amplitude, mean value, standard deviation, partiality, and kurtosis by analyzing the time domain waveform of signal. In this processing, less information is lost. Common time domain analysis methods include zero-crossing analysis, histogram analysis, analysis of variance, correlation analysis, peak detection, waveform parameter analysis, and waveform recognition. Emotion recognition studies using cerebral blood oxygen parameters more often involve time domain feature analysis and extraction.

    b. **Frequency domain**. Frequency domain feature extraction is based on power spectrum analysis and widely used in analysis of ECG, respiration, EEG, and other signals, such as power spectrum ratio, power spectrum energy, and sub-band power spectral density in different frequency bands.

c. **Time-frequency feature**. The time-frequency feature extraction considers joint distribution information in terms of time domain and frequency domain. This method describes the changing relationship between signal frequencies and time and contains more comprehensive contents. Commonly used analysis methods are wavelet transform, short-time Fourier transform, Hilbert-Huang transform, and others. Wavelet transform has multiresolution, adjustable sliding time window, has good resolution in both time domain and frequency domain, and has become an effective tool for analyzing nonstationary signals, such as EEG, ECG, EMG, and other signals underlying emotion processes.

d. **Nonlinear feature**. EEG signals are created in complex limbic system with noticeable nonlinearity and chaos characteristic, so the extraction of EEG features is more complex and diverse than other physiological signals. In recent years, the analysis of nonlinear features such as entropy, correlation dimension, and fractal dimension has gradually increased in the study of emotional EEG recognition. Konstantinidis et al. [48] calculated the correlation dimension of emotional EEG for online recognition research; Liu et al. [49] extracted the nonlinear features such as the fractal dimension of EEG to obtain the ideal recognition effect and built an online application.

2.6 Multimodal emotion recognition

Most recent researches have focused on multimodal emotion recognition using visual and aural information. Human expression of emotion is mostly multimodal, including visual, audio, and textual modalities for effective communication [3]. Furthermore, physiological signals can reveal emotional state objectively, even if the subject conceals his/her expression of emotion due to complex reasons. Hence emotion recognition integrating multiple modalities has gained increasing attention, and research hotspot has shifted from single modality to multimodal emotion recognition in practical applications. D'Mello and Kory [50] used statistical methods to compare the accuracy of single modality and multimodal on different databases. Multimodal expression recognition was superior to single modality performance in the experiments. The McGurk [51] phenomenon reveals that in the process of brain perception, different senses are automatically combined unconsciously to process the information, and any lack or inaccuracy of sensory information will lead to deviations in the brain's understanding of external information. Therefore, multimodal feature fusion recognition technology has become a research hotspot in the past few years.

The widely used multimodal emotion databases are HUMAINE database [52], the Belfast database [53], the large-scale audiovisual database SEMAINE [54], the IEMOCAP emotional database [55], the audiovisual database eNTERFACE [56], the Acted Facial Expression in the Wild database (AFEW) [57] composed of audio and video clips from English movies and TV programs, and the Chinese multimodal emotional data set CHEAVD [58].

Multichannel information fusion levels can be divided into three categories: data layer, feature layer, and decision layer: (1) Data layer fusion refers to the fusion of collected raw data and then extracting feature vector from the fused data, finally classifying the emotion; (2) feature layer fusion refers to conducting preprocessing and feature extraction of the collected data of each channel first, then obtaining the

feature vector by fusing extracted emotion features, and then finally classifying the emotion; and (3) decision layer fusion refers to making separate emotion classification decision for collected data of each channel and then fusing the single modality recognition result to obtain the final classification result. The commonly used information fusion methods are D-S evidence theory, artificial neural network, fuzzy set theory, Bayesian inference, cluster analysis, expert system method, and others.

Current studies on postures mainly concentrate on bimodal emotion recognition of facial expressions and postures. Castellano et al. [25] conducted a comparative study of the processing of body language and facial expression and found that body language and facial expression have similar visual processing mechanisms. The two are highly similar in terms of event-related potential (ERP) components, psychological functions, and influencing factors and are partially overlapping or adjacent to each other in potential neural bases. Gunes and Piccardi [59, 60] conducted long-term research on bimodal emotion recognition of facial expressions and postures and established the Bi-modal Face and Body Gesture Database for Automatic Analysis of Human Nonverbal Affective Behavior (FABO). Yan et al. [61] studied video-based bimodal emotion recognition of facial expression and postures and proposed an emotion recognition method based on bilateral sparse partial least squares which has low computational complexity but low recognition rate. In order to tell human emotions through video data, Wang and Shao [62] extracted emotional features of facial expression and body movements from the FABO database, used a fusion algorithm based on canonical correlation analysis (CCA) to fuse two features, and then used nearest neighbor classifier and support vector machine for emotion recognition. After using updated sparsity preserving CCA (SPCCA), they combined emotion features of facial expressions and body movements, achieving an emotion recognition rate at 90.48%. Wang et al. [63] focused on the problem of high computational complexity in video emotion recognition and proposed a bimodal emotion recognition method based on temporal-spatial local binary pattern moment (TSLBPM) which has been proven effective. Jiang et al. [64] proposed a spatiotemporal local ternary orientational pattern (SLTOP) feature description method and cloud-weighted decision fusion classification method for bimodal emotion recognition of facial expressions and postures in video sequences, achieving better recognition result than other classification recognition methods in the comparative experiments.

3. Application of emotion recognition in depressive disorder

Depressive disorder is characteristic of negative mood and anhedonia, which are two core symptoms for diagnosis of the disease. Traditionally, the clinical diagnosis for depression requires the clinicians to assess the severity of depressive symptoms according to verbal statements of patients as well as nonverbal indicators such as voices (pitch, speaking speed, and volumes) and facial expressions. Additionally, structured questionnaires (such as Beck Depression Inventory, Hamilton Depression Rating Scale) have been developed and validated in clinical populations to assess the severity of depressive symptoms. However, the physiological biomarkers of depression are still unclear. Since the 1950s the consensus has emerged that psychiatric diagnoses could be defined according to relevant biological characteristics. However, the empirical diagnostic categories such as depressive disorder failed to be reified and objectified by valid biological measures [65]. The Research Domain Criteria (RDoC) initiative attempted to link physiologic mechanisms (esp. circuit level) to dimensional constructs (e.g., positive/negative valence) rather than

diagnostic categories (e.g., MDD), with the potential for alternative diagnostic processes [66].

3.1 Physiological emotion recognition

Ample evidence showed that specific brain regions including the PFC, amygdala, anterior cingulate, and insula play a major role in the neuropathological basis of affective disorders. Recent meta-analyses found evidence which is against the locationist account of emotion and suggested that brain regions corresponding to basic psychological operations are involved in emotion processing across emotional categories and are not specifically localized to discrete brain networks [67]. With its advantages in superior soft tissue contrast, high spatial resolution, and noninvasive detection, magnetic resonance imaging (MRI) has become a promising tool for detection of neurological alterations in mental disorders such as depression.

Using an experimental therapeutics approach coupled with machine learning, Liu et al. investigated the effect of a pharmacological challenge aiming to enhance dopaminergic signaling on whole-brain's response to reward-related stimuli in MDD. Artificial intelligence technology combined with MRI technology was used to find the objective biological markers of depression. The brain regions with diagnostic value included anterior cuneate lobe, cingulate gyrus, inferior marginal angular gyrus, insular, thalamus, and hippocampus. The brain regions with preventive value included the precuneus, postcentral gyrus, dorsolateral prefrontal lobe, orbitofrontal lobe, and middle temporal gyrus. The brain regions with predictive therapeutic response included the precuneus, cingulate gyrus, inferior marginal angular gyrus, middle frontal gyrus, middle occipital gyrus, inferior occipital gyrus, and lingual gyrus [68].

Studies have shown that machine learning and deep learning techniques have been widely used in the diagnosis, prevention, and treatment of depression and other neuropsychiatric diseases in recent years. Abnormal brain regions may be used as predictors of diagnostic and therapeutic responses. Research hotspot mainly focused on cortical areas rather than the midbrain limbic system or dopamine system. Collectively, the literature review suggested that the cingulate gyrus and precuneus may be the most important candidate brain regions among the objective biological markers of depression. Due to complex pathophysiological changes and etiological heterogeneity of depression, combining imaging biomarkers with other indicators (e.g., biochemical, genetic) is necessary to achieve more objective assessment of course and prognosis of depression [69].

3.2 Textual emotion recognition

With the growing amount of emotional information from social media, including text, photos, and videos, emotion recognition through multimodal information using machine learning technique is becoming a trend. Absolutist thinking represents a form of cognitive distortion typical of anxiety and depression. Al-Mosaiwi and Johnstone conducted a text analysis of 63 Internet forums (over 6400 members) using the Linguistic Inquiry and Word Count software to examine absolutist thinking. The results suggested that absolutist words, rather than negative emotion words, tracked the severity of affective disorder forums. They found elevated levels of absolutist words in depression recovery forums. This suggests that absolutist thinking may be a vulnerability factor for relapse of affective disorder [70].

The project of Proactive Suicide Prevention Online (PSPO) identified suicide-prone individuals to provide further crisis management. A microblog group was

identified as a high-risk population, who commented around a Sina microblogger who committed suicide. They were assessed for suicidal thought and behavior. The frequency of death-oriented words significantly decreased after the intervention, while the frequency of future-oriented words significantly increased. This model may help people with suicidal thoughts and behaviors but with a low motivation to seek help [71].

The modeling of textual and visual features from Instagram photos successfully identified individuals diagnosed with depression. The results showed that depressed people are more likely to upload photos that are bluer, grayer, and darker. The human rating of photo attributes (happiness, sadness, interestingness, and likability) is a weak predictor of depression [72]. These findings suggest new avenues for early screening and detection of mental illness.

3.3 Facial expression and speech recognition

The physiological approaches using specific sensors for emotion signals have the advantage of being more precise, but are generally more costly and need more effort in clinical context. Facial and speech information is more applicable in these natural environments. Chronic stress, anxiety, and depressive states are three intertwined processes which constitute the vicious circle in common affective disorders such as depression. Chronic stress may induce autonomic responses concurrent with anxiety states, and anxiety may lead to depressive states when stress continues and coping strategies are ineffective. Gavrilescu and Vizireanu for the first time proposed a neural network-based architecture for predicting levels of stress, anxiety, and depression based on FACS in a nonintrusive and real-time manner. Their method allows the experts to monitor the three emotional states in real time. Additionally, 93% accuracy was achieved discriminating between healthy individuals and those with major depressive disorder (MDD) or post-traumatic stress disorder (PTSD) [73]. This method is an attractive alternative to traditional self-report measurements based on questionnaires.

A new approach to predict the depressive symptoms with Beck Depression Inventory II (BDI-II) scores from video data is proposed based on the deep convolutional neutral networks (DCNN). The proposed framework is designed to capture both the facial appearance and dynamics in the video data by integrating two deep networks into one. The method could predict with over 80% accuracy depressive behavior, achieving a comparable performance to most methods combing video and audio data [74]. Thus their method provided a more efficient and convenient way of prediction than multimodal methods.

Harati et al. used several metrics of variability to extract unsupervised features from video recordings of patients before and after deep brain stimulation (DBS) treatment for major depressive disorder (MDD). Their goal was to quantify the treatment effects on emotion indicated with facial expression. Their preliminary results indicate that unsupervised features learned from these video recordings using dynamic latent variable model (DLVM) based on multiscale entropy (MSE) of pixel intensities can distinguish different phases of depression and recovery [75]. Therefore, their methods may provide more precise markers of treatment response.

As a relatively objective and easily available variable, speech has potential value in the diagnosis of depression. The acoustic analysis of patients with mental illness showed that there is greater than moderate correlation between speech-related variables and symptom indicators [76, 77]. Pan et al. build a speech-based depression recognition model with logical regression (LR) classification methods. The results show that the speech recognition accuracy reached 82.9% [78].

They found that four voice features (PC1, PC6, PC17, PC24, $P < 0.05$, corrected) made significant contribution to depression and that the contribution effect of the voice features alone reached 35.65%. These results demonstrate that voice features have great potential in applications such as clinical diagnosis and prediction.

3.4 Multimodal emotion recognition

Facial, video, and textual information are the most available affective information in clinical context. Therefore, recent studies explored multimodal emotion recognition methods to improve the accuracies and specificities when the multimodal emotion information was input as predictors. Haque et al. present a machine learning method for measuring the severity of depressive symptoms. Their multimodal method uses 3D facial expressions and spoken language, commonly available from modern cell phones. It demonstrates an average error of 3.67 points (15.3% relative) on the clinically validated Patient Health Questionnaire (PHQ) scale. For detecting major depressive disorder, their model demonstrates 83.3% sensitivity and 82.6% specificity [79]. Yang et al. proposed new text and video features and hybridizes deep and shallow models for depression estimation and classification from audio, video, and text descriptors. They demonstrated that the proposed hybrid framework effectively improves the accuracies of both depression estimation and depression classification [80]. SimSensei Kiosk was a virtual human interviewer which aims to automatically assess the verbal and nonverbal behaviors indicative of depression, anxiety, or post-traumatic stress disorder (PSTD). A multimodal real-time sensing system was used to simultaneously capture different modalities (e.g., smile intensity, 3D head position and orientation, intensity or lack of facial expressions like anger, disgust, and joy) to model the relation between mental states and human behavior [81].

4. Conclusions

This chapter summarized the recognition of human affect based on internal and external signals of emotion, which has gained intensive attention in research fields such as artificial intelligence, psychology, cognitive neuroscience, and physiology. The reviewed empirical researches rarely deal with "social emotions" such as guilt, shame, and embarrassment. Instead of the more traditional cognitive and biological perspectives of emotion, the sociological perspective focused on functions of emotions to control social interactions and sustain the social order. Future studies need to deal with its extension to social emotions and the relevant theoretical foundations.

Emotion recognition is based on discrete and dimensional views of emotion, with underlying CNS and PNS dynamics. Single modal as well as multimodal emotion recognition rely on facial, speech, posture, physiological, and textual emotional information, which could function separately or concurrently. Integrating multimodal emotion information for emotion recognition remains challenging, and much research is needed about the way they relate to human affect.

Furthermore, the application of emotion recognition in depressive disorder may pave an avenue for more precise diagnosis of the syndrome and prediction of its disease course. Identifying specific physiological substrates of depressive disorder, combined with emotion classification technique such as machine learning, may help identify the dimensional constructs of RDoC, which are implicit in the clinical phenomena of depressive disorder.

Conflict of interest

The authors declare no conflict of interest.

Author details

Xiaotong Cheng[1†], Xiaoxia Wang[2*†], Tante Ouyang[1] and Zhengzhi Feng[1]

1 College of Psychology, Army Medical University, Chongqing, China

2 Department of Basic Psychology, College of Psychology, Army Medical University, Chongqing, China

*Address all correspondence to: lemonowang@gmail.com

† Cheng and Wang contributed equally to this article and should be considered co-first authors.

References

[1] Minsky M. The Society of Mind. New York, NY: Simon & Schuster Imprint; 1985

[2] Picard RW. Affective Computing. Cambridge, MA: The MIT Press; 1997

[3] Poria S, Cambria E, Bajpai R, Hussain A. A review of affective computing: From unimodal analysis to multimodal fusion. Information Fusion. 2017;**37**:98-125. DOI: 10.1016/j.inffus.2017.02.003

[4] Ekman P, Levenson RW, Friesen WV. Autonomic nervous system activity distinguishes among emotions. Science. 1983;**221**(4616):1208-1210. DOI: 10.1126/science.6612338

[5] Singh D. Human emotion recognition system. International Journal of Image, Graphics and Signal Processing. 2012; **4**(8):50-56. DOI: 10.5815/ijigsp.2012.08.07

[6] Quigley KS, Barrett LF. Is there consistency and specificity of autonomic changes during emotional episodes? Guidance from the conceptual act theory and psychophysiology. Biological Psychology. 2014;**98**:82-94. DOI: 10.1016/j.biopsycho.2013.12.013

[7] Ekman P, Friesen WV. Facial Action Coding System: A Technique for the Measurement of Facial Movement. Palo Alto, CA: Consulting Psychologists Press; 1978

[8] Yang C, Li H. Validity study on facereader's images recognition from Chinese facial expression database. Ergonomics. 2015;**21**:38-41. DOI: 10.13837/j.issn.1006-8309.2015.01.0008

[9] Gong X et al. Revision of the Chinese facial affective picture system. Chinese Journal of Mental Health. 2011;**250**: 40-46. DOI: 10.3969/j.issn.1000-6729.2011.01.011

[10] Liu J et al. Establishment of the Chinese facial emotion images database with intensity classification. Chinese Journal of Mental Health. 2019;**33**: 120-125. DOI: 10.3969/j.issn.1000-6729.2019.02.009

[11] Ballano S, Hupont I, Cerezo E, Baldassarri S. Recognizing Emotions from Video in a Continuous 2D Space. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. pp. 600-603

[12] Polikovsky S, Kameda Y, Ohta Y. Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor. In: Proceedings of the 3rd International Conference on Crime Detection and Prevention; 3 December 2009; London. London, UK: IEEE; 2009. pp. 1-6. DOI: 10.1049/ic.2009.0244

[13] Shreve M, Godavarthy S, Manohar V, Goldgof D, Sarkar S. Towards macro- and micro-expression spotting in video using strain patterns. In: Proceedings of the IEEE Workshop on Applications of Computer Vision; 2009/12; 7-8 December 2009; Snowbird, UT. Snowbird, UT: IEEE; 2009. pp. 1-6

[14] Yan W-J, Li X, Wang S-J, Zhao G, Liu Y-J, Chen Y-H, et al. CASME II: An improved spontaneous micro-expression database and the baseline evaluation. PLoS One. 2014;**9**:e86041. DOI: 10.1371/journal.pone.0086041

[15] Lee CM, Narayanan SS. Toward detecting emotions in spoken dialogs. IEEE Transactions on Speech and Audio Processing. 2005;**13**:293-303. DOI: 10.1109/tsa.2004.838534

[16] Schuller B, Rigoll G, Lang M. Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture. In: Proceedings of the 2004 IEEE International Conference

on Acoustics, Speech, and Signal Processing; 17-21 May 2004; Montreal. Montreal, Quebec: IEEE; 2004. pp. I-577

[17] Ververidis D, Kotropoulos C. Emotional speech recognition: Resources, features, and methods. Speech Communication. 2006;**48**: 1162-1181. DOI: 10.1016/j.specom.2006.04.003

[18] Rickheit G, Strohner H. Handbook of Communication Competence. Germany: Mouton de Gruyter; 2008

[19] Gharavian D, Sheikhan M, Nazerieh A, Garoucy S. Speech emotion recognition using FCBF feature selection method and GA-optimized fuzzy ARTMAP neural network. Neural Computing and Applications. 2012;**21**: 2115-2126. DOI: 10.1007/s00521-011-0643-1

[20] Devi JS, Yarramalle S, Prasad NS. Speaker emotion recognition based on speech features and classification techniques. International Journal of Image, Graphics and Signal Processing. 2014;**6**:61-77. DOI: 10.5815/ijigsp.2014.07.08

[21] Zhang L et al. Speech emotion recognition based on deep belief network. Journal of Taiyuan University of Technology. 2019;**50**:101-107. DOI: 10.16355/j.cnki.isn1007-9432tyut.2019.01.016

[22] Zhu Y, Bodong S, Lichen Z. Temporal and spatial spectral feature speech emotion recognition algorithm based on trajectory. Computer System Application. 2019;**28**:146-151. DOI: 10.15888/j.cnki.csa.006794

[23] Liu Y, Qin H. The application of speech emotion recognition technology in the field of manned space. In: People's Forum Academic Frontiers. 2018. pp. 69-73. DOI: 10.16619/j.cnki.rmltxsqy.2018.17.008

[24] Tao J, Tan T. Affective Computing: A Review. Berlin, Heidelberg: Springer Berlin Heidelberg; 2005

[25] Castellano G, Villalba SD, Camurri A. Recognising human emotions from body movement and gesture dynamics. In: Paiva ACR, Prada R, Picard RW, editors. Affective Computing and Intelligent Interaction. ACII 2007. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer; 2007. pp. 71-82

[26] Bernhardt D, Robinson P. Detecting Emotions from Everyday Body Movements [EB/OL]. 2007. Available from: https://www.cl.cam.ac.uk/

[27] Liu Y, Liu D, Han Z. Research on emotion extraction method based on motion recognition. Computer Engineering. 2015;**41**:300-305

[28] Shao J, Wang W. Emotion recognition based on three-dimensional texture features of body movement sequences. Journal of Computer Applications. 2018;**35**:3497-3499. DOI: 10.3969 /j.issn.1001-3695.2018.11.071

[29] Hao Y et al. Method of microblog emotion recognition based on word fusion features. Computer Science. 2018;**45**:105-109. DOI: 10.11896/j.issn.1002-137X.2018.11A.018

[30] Hao M et al. Emotion classification and prediction algorithm based on Chinese microblog. Computer Application. 2018;**38**:89-96

[31] Huang F-L et al. Weibo themed emotion mining based on multi-feature fusion. Chinese Journal of Computers. 2017;**40**:872-888. DOI: 10.11897/SP.J.1016.2017.00872

[32] Zhang C, Qian T, Ji D. A joint model of microblogging emotion recognition and incentive extraction based on neural network. Computer Application. 2018;

38:2464-2468+2476. DOI: 10.11772/j.issn.1001-9081.2018020481

[33] James W. The physical basis of emotion. Psychological Review. 1994; **101**(2): 205-210. DOI: 10.1037/0033-295X.101.2.205

[34] Kreibig SD. Autonomic nervous system activity in emotion: A review. Biological Psychology. 2010;**84**: 394-421. DOI: 10.1016/j.biopsycho.2010.03.010

[35] Picard RW, Vyzas E, Healey J. Toward machine emotional intelligence: Analysis of affective physiological state. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2001;**23**: 1175-1191. DOI: 10.1109/34.954607

[36] Kim J, Andre E. Emotion recognition based on physiological changes in music listening. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2008;**30**: 2067-2083. DOI: 10.1109/tpami.2008.26

[37] Yan F, Liu GY, Lai XW. The research on material selection algorithm design with improved OWA in affective regulation system based on human-computer interaction. Journal of Information and Computational Science. 2013;**10**: 4477-4486. DOI: 10.12733/jics20102223

[38] Li C-L, Ye N, Huang H-P, Wang R-C. Physiological signal emotion recognition based on recursive quantitative analysis. Computer Technology and Development. 2018;**28**: 94-98. +102

[39] Jin C, Chen G. Multi-modal physiological signal emotion recognition based on optimized LSTSVM. Application of Electronic Technology. 2018;**44**:112-116. DOI: 10.16157/j.issn.0258-7998.171839

[40] Yong P, Jia-Yi Z, Wei-Long Z, Bao-Liang L. EEG-based emotion recognition with manifold regularized extreme learning machine. In: Proceedings of the 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 26-30 August 2014; Chicago. Chicago, IL: IEEE; 2014. pp. 974-977

[41] Jie X, Cao R, Li L. Emotion recognition based on the sample entropy of EEG. Bio-Medical Materials and Engineering. 2014;**24**:1185-1192. DOI: 10.3233/bme-130919

[42] Duan R-N, Zhu J-Y, Lu B-L. Differential entropy feature for EEG-based emotion classification. In: Proceedings of the 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER); November 2013; San Diego. San Diego, CA: IEEE; 2013. pp. 81-84

[43] Soleymani M, Lichtenauer J, Pun T, Pantic M. A multimodal database for affect recognition and implicit tagging. IEEE Transactions on Affective Computing. 2012;**3**:42-55. DOI: 10.1109/t-affc.2011.25

[44] Koelstra S, Muhl C, Soleymani M, Jong-Seok L, Yazdani A, Ebrahimi T, et al. Deap: A database for emotion analysis using physiological signals. IEEE Transactions on Affective Computing. 2012;**3**:18-31. DOI: 10.1109/t-affc.2011.15

[45] Tai K, Chau T. Single-trial classification of NIRS signals during emotional induction tasks: Towards a corporeal machine interface. Journal of NeuroEngineering and Rehabilitation. 2009;**6**:1-14. DOI: 10.1186/1743-0003-6-39

[46] Mandryk RL, Atkins MS. A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. International Journal of Human-Computer Studies. 2007;**65**:329-347. DOI: 10.1016/j.ijhcs.2006.11.011

[47] Zhang D, Wan B, Ming D. Research progress on emotion recognition based on physiological signals. Journal of Biomedical Engineering. 2015;**32**: 229-234

[48] Konstantinidis EI, Frantzidis CA, Pappas C, Bamidis PD. Real time emotion aware applications: A case study employing emotion evocative pictures and neuro-physiological sensing enhanced by graphic processor units. Computer Methods and Programs in Biomedicine. 2012;**107**:16-27. DOI: 10.1016/j.cmpb.2012.03.008

[49] Liu Y, Sourina O, Nguyen MK. Real-time EEG-based emotion recognition and its applications. Transactions on Computational Science XII. 2011;**6670**: 256-277. DOI: 10.1007/978-3-642-22336-5_13

[50] D'Mello SK, Kory J. A review and meta-analysis of multimodal affect detection systems. ACM Computing Surveys. 2015;**47**:1-36. DOI: 10.1145/2682899

[51] McGurk H, Macdonald J. Hearing lips and seeing voices. Nature. 1976;**264**: 746-748. DOI: 10.1038/264746a0

[52] Douglas-Cowie E, Cowie R, Sneddon I, Cox C, Lowry O, McRorie M, et al. The HUMAINE data-base: Addressing the collection and annotation of naturalistic and induced emotional data. In: Proceedings of the 2nd International Conference on Affective Computing and Intelligent Interaction. Berlin: Springer Berlin Heidelberg; 2007. pp. 488-500

[53] Douglas-Cowie E, Cowie R, Campbell N. A new emotion database: Considerations, sources and scope. In: Proceedings of the ISCA Workshop on Speech and Emotion; April 2003. Belfast: Textflow; 2000. pp. 39-44

[54] McKeown G, Valstar M, Cowie R, Pantic M, Schroder M. The SEMAINE database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent. IEEE Transactions on Affective Computing. 2012;**3**:5-17. DOI: 10.1109/t-affc.2011.20

[55] Busso C, Bulut M, Lee C-C, Kazemzadeh A, Mower E, Kim S, et al. IEMOCAP: Interactive emotional dyadic motion capture database. Language Resources and Evaluation. 2008;**42**: 335-359. DOI: 10.1007/s10579-008-9076-6

[56] Martin O, Kotsia I, Macq B, Pitas I. The eNTERFACE 05 audio-visual emotion database. In: Proceedings of the 22nd International Conference on Data Engineering Workshops (ICDEW'06); 3-7 April 2006; Atlanta. Atlanta, GA, USA: IEEE; 2006. p. 8

[57] Dhall A, Goecke R, Joshi J, Hoey J, EmotiW GT. Video and group-level emotion recognition challenges. In: Proceedings of the 18th ACM International Conference on Multimodal Interaction—ICMI 2016. ACM Press; 2016. pp. 427-432

[58] Li Y, Tao J, Schuller B, Shan S, Jiang D, MEC JJ. The multimodal emotion recognition challenge of CCPR 2016. In: Tan T, Li X, Chen X, Zhou J, Yang J, Cheng H, editors. Pattern Recognition. CCPR 2016. Communications in Computer and Information Science. Vol. 2016. Singapore: Springer; 2016. pp. 667-678

[59] Gunes H, Piccardi M. A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior. In: Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06); 20-24 August 2006. Hong Kong: IEEE; 2006. pp. 1148-1153

[60] Gunes H, Piccardi M. Bi-modal emotion recognition from expressive face and body gestures. Journal of

Network and Computer Applications. 2007;**30**:1334-1345. DOI: 10.1016/j.jnca.2006.09.007

[61] Yan J, Zheng W, Xin M, Qiu W. Bimodal emotion recognition based on body gesture and facial expression. Journal of Image and Graphics. 2013;**18**: 1101-1106

[62] Wang W, Shao J. Emotion recognition combining facial expressions and body movement characteristics. Television Technology. 2018;**42**:73-76.+83. DOI: 10.16280/j.videoe.2018.01:014

[63] Wang X, Hou D, Hu M, Ren F. Bimodal emotion recognition of composite spatiotemporal features. Journal of Image and Graphics. 2017;**22**: 39-48. DOI: 10.11834/jig.20170105

[64] Jiang M et al. Bimodal emotion recognition of expressions and postures in video sequences. Progress in Laser and Optoelectronics. 2018;**55**:167-174. DOI: 10.3788/LOP55.071004

[65] Yee CM, Javitt DC, Miller GA. Replacing DSM categorical analyses with dimensional analyses in psychiatry research: The research domain criteria initiative. JAMA Psychiatry. 2015: 1159-1160. DOI: 10.1001/jamapsychiatry.2015.1900

[66] Kraemer HC. Research domain criteria (RDoC) and the DSM-two methodological approaches to mental health diagnosis. JAMA Psychiatry. 2015:1163-1164. DOI: 10.1001/jamapsychiatry.2015.2134

[67] Lindquist KA, Wager TD, Kober H, Bliss-Moreau E, Barrett LF. The brain basis of emotion: A meta-analytic review. The Behavioral and Brain Sciences. 2012;**35**(3):121-143. DOI: 10.1017/S0140525X11000446

[68] Liu Y, Admon R, Belleau EL, Kaiser RH, Clegg R, Beltzer M, et al. Machine learning identifies large-scale reward-related activity modulated by dopaminergic enhancement in major depression. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging. 2020;**5**(2):163-172. DOI: 10.1016/j.bpsc.2019.10.002

[69] Sun YT, Chen T, He D, Dong Z, Cheng B, Wang S, et al. Research progress of biological markers for depression based on psychoradiology and artificial intelligence. Progress in Biochemistry and Biophysics. 2019;**46**: 879-899

[70] Al-Mosaiwi M, Johnstone T. In an absolute state: Elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. Clinical Psychological Science. 2019;**7**: 636-637. DOI: 10.1177/2167702619843297

[71] Liu X, Liu X, Sun J, Yu NX, Sun B, Li Q, et al. Proactive suicide prevention online (PSPO): Machine identification and crisis management for Chinese social media users with suicidal thoughts and behaviors. Journal of Medical Internet Research. 2019;**21**: e11705. DOI: 10.2196/11705

[72] Reece AG, Danforth CM. Instagram photos reveal predictive markers of depression. EPJ Data Science. 2017;**6**:15. DOI: 10.1140/epjds/s13688-017-0118-4

[73] Gavrilescu M, Vizireanu N. Predicting depression, anxiety, and stress levels from videos using the facial action coding system. Sensors (Basel). 2019;**19**(17). DOI: 10.3390/s19173693

[74] Zhu Y, Shang Y, Shao Z, Guo G. Automated depression diagnosis based on deep networks to encode facial appearance and dynamics. IEEE Transactions on Affective Computing. 2018;**9**(4):578-584. DOI: 10.1109/TAFFC.2017.2650899

[75] Harati S, Crowell A, Mayberg H, Kong J, Nemati S. Discriminating

clinical phases of recovery from major depressive disorder using the dynamics of facial expression. In: 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); January 01, 2016; United States: IEEE. 2016

[76] Cohen AS, Najolia GM, Kim Y, Dinzeo TJ. On the boundaries of blunt affect/alogia across severe mental illness: Implications for research domain criteria. Schizophrenia Research. 2012; **140**(1–3):41-45. DOI: 10.1016/j.schres.2012.07.001

[77] Covington MA, Lunden SLA, Cristofaro SL, Wan CR, Bailey CT, Broussard B, et al. Phonetic measures of reduced tongue movement correlate with negative symptom severity in hospitalized patients with first-episode schizophrenia-spectrum disorders. Schizophrenia Research. 2012;**142**(1): 93-95. DOI: 10.1016/j.schres.2012.10.005

[78] Pan W, Flint J, Shenhav L, Liu T, Liu M, Hu B, et al. Re-examining the robustness of voice features in predicting depression: Compared with baseline of confounders. PLoS One. 2019;**14**:e0218172. DOI: 10.1371/journal.pone.0218172

[79] Haque A, Guo M, Miner AS, Fei-Fei L. Measuring depression symptom severity from spoken language and 3D facial expressions. Sound. 2018;**2**:1-7

[80] Yang L, Jiang D, Sahli H. Integrating deep and shallow models for multi-modal depression analysis—Hybrid architectures. IEEE Transactions on Affective Computing. 2018;**1**:1-16. DOI: 10.1109/TAFFC.2018.2870398

[81] David DeVault RAGB, Georgila K, Gratch J, Hartholt A, Lhommet M, Lucas G, et al. SimSensei kiosk: A virtual human interviewer for healthcare decision support. In: Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014). 2014