CCF Transactions on Pervasive Computing and Interaction
https://doi.org/10.1007/s42486-026-00249-9

REVIEW PAPER

A tutorial review of AI-powered physiological computing for mental health assessment: application scenarios, open issues, and current solutions

Yufei Zhang[1] · Wenting Kuang[1] · Yulan Li[1] · Yuqi Cheng[1] · Kaishun Wu[2] · Yongpan Zou[1]

Received: 19 March 2026 / Accepted: 5 June 2026
© China Computer Federation (CCF) 2026

Abstract

Mental health is a critical component of overall well-being, with nearly one billion individuals worldwide affected by mental disorders. Recent advancements in artificial intelligence (AI), coupled with the growing accessibility of physiological sensing technologies, have opened new avenues for continuous and data-driven mental health monitoring. Physiological signals such as electroencephalography (EEG), electrocardiography (ECG), and galvanic skin response (GSR) serve as rich, nonverbal indicators of emotional and cognitive states. These signals can be acquired using both specialized medical-grade equipment and ubiquitous wearable devices, making them valuable for detecting emotions, stress levels, and mental health disorders in diverse settings. This review provides an extensive overview of AI approaches for mental health assessment based on physiological signals. It first outlines common affective contexts, widely used signal modalities, benchmark datasets, and foundational AI models, offering a structured entry point for researchers in this field. Subsequently, we propose the critical challenges of domain shift, model interpretability, and multimodal fusion to ensure both the generalizability and the practical applicability in this field. To be concrete, we systematically analyze representative methodologies and recent advances for each challenge and discuss promising future directions. We believe this work can serve as a bridge between foundational knowledge and emerging research, fostering further advancements in affective computing.

Keywords Mental health assessment · Physiological signal analysis · Affective computing · Wearable computing · Ubiquitous computing

1 Introduction

Mental health constitutes a fundamental aspect of overall well-being and quality of life. In recent years, the prevalence of mental disorders has risen sharply. As reported by the World Health Organization, the global prevalence of mental disorders, particularly depression and anxiety, increased by approximately 25% in 2020 alone, culminating in nearly one billion individuals affected worldwide (Organization et al. 2023). This alarming trend highlights the need to investigate mechanisms underlying mental health deterioration. A growing body of research has pointed to emotional dysregulation as a central mechanism in the onset of mental disorders. Chronic stress has been also shown to disrupt emotional regulation processes and increases vulnerability to psychological conditions (Khansari et al. 1990; O'leary 1990; Maddock and Pariante 2001). And mental disorders can further exacerbate stress reactivity by

✉ Yongpan Zou
yongpan@szu.edu.cn

Yufei Zhang
yfyaya@foxmail.com

Wenting Kuang
kuangwenting2023@email.szu.edu.cn

Yulan Li
yulanlee2024@gmail.com

Yuqi Cheng
cyuqi2090@gmail.com

Kaishun Wu
wuks@hkust-gz.edu.cn

[1] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518062, Guangdong, China

[2] Information Hub, Hong Kong University of Science and Technology, Guangzhou 510317, Guangdong, China

Published online: 26 October 2026

undermining cognitive and emotional resilience, thereby forming a self-reinforcing loop (Hammen 2005). This bidirectional relationship reveals the dynamic interplay among stress, emotional instability, and mental health outcomes.

Given this interdependence, the ability to identify emotional disturbances at an early stage is of great clinical importance. Yet, conventional psychiatric assessments often depend on self-reports and clinician interpretation, which may lack objectivity, consistency, and scalability. Therefore, there is an urgent need to introduce more accurate and reliable technologies to assist in psychiatric diagnosis. Researchers have found that emotional fluctuations, particularly those associated with suppressed emotions or distress, can trigger measurable changes in the autonomic nervous system (ANS), thereby governing involuntary physiological responses (Ekman et al. 1983; Taylor 1991). For instance, heightened stress or anxiety typically activates the sympathetic nervous system, leading to increased heart rate, amplified skin conductance, and amygdala-driven arousal, all of which serve as real-time physiological markers of emotional arousal (Jansen et al. 1995; Doust and Schneider 1952). These responses offer a non-verbal and real-time channel for assessing emotional dynamics, making them especially valuable when individuals are unwilling or unable to articulate their feelings.

Leveraging these properties, physiological signals enable continuous, objective monitoring of emotional dynamics in the field of affective computing. Recent advances in artificial intelligence (AI), particularly in machine learning (ML) and deep learning (DL), have further amplified the potential of physiological signal analysis. These approaches can uncover complex, non-linear patterns in high-dimensional data and have demonstrated success in tasks such as emotion recognition, stress detection, and even the diagnosis of specific mental disorders (Yang et al. 2024). Some studies suggest that AI models can approach or surpass expert-level performance in certain diagnostic tasks, indicating their potential to complement traditional clinical practices (Viswan et al. 2024). Motivated by these developments, this review provides a systematic review of AI-powered approaches to mental health assessment based on physiological signals, as illustrated in Fig. 1[1]. We begin by out-

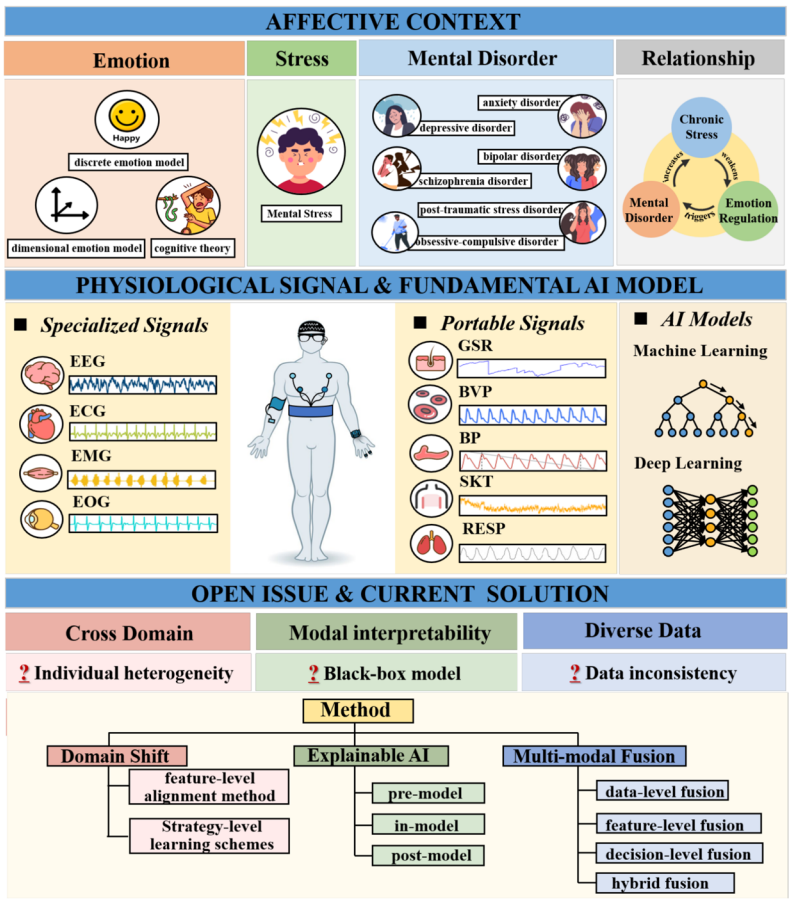


Fig. 1 Overall introduction framework

[1] The term "emotion regulation" in this context refers to the regulation of negative emotions. In other parts of the manuscript, the term "emotion" generally follows the emotion categories defined in Section 3.1.

lining key mental health assessment tasks across different application contexts, including emotion recognition, stress detection, and mental disorder diagnosis. We then review commonly used physiological modalities and public datasets, followed by a taxonomy of computational models ranging from traditional ML to recent DL-based architectures. In addition, we examine open issues such as individual heterogeneity, black-box model and signal variability. We also explore promising solutions that incorporate domain shift, model interpretability, and multi-modal fusion techniques to address the limitations of current physiological signal-based mental health assessments. Together, these discussions aim to provide insights that contribute to apprehending the continued evolution of AI-driven approaches applied in this area.

Our main contributions are summarized as follows:

- We provide an up-to-date review of the literature on the application of artificial intelligence in emotion recognition, stress detection, and mental disorder diagnosis, highlighting a range of promising advancements in mental health assessment.
- We offer a holistic tutorial that introduces the core components of this research area, including affective contexts, physiological signal modalities, widely used public datasets, and representative AI models. This guide aims to facilitate researchers new to the field in quickly gaining an understanding of the key resources and available data.
- We present a synthesis of insights on the open issues confronted by researchers and developers in implementing artificial intelligence models, discuss key challenges in real-world implementation and future directions from both technical and clinical perspectives.

The remainder of this paper is organized as follows: Sect. 2 reviews prior surveys and clarifies the distinct contributions of this work. Section 3 to Sect. 5 introduce affective contexts, physiological signals, public datasets, and foundational AI models relevant to mental health applications. Section 6 to Sect. 7 examine major open issues, explore emerging solutions, and outline future research directions. Finally, Sect. 8 concludes the paper and summarizes the key insights.

2 Related work

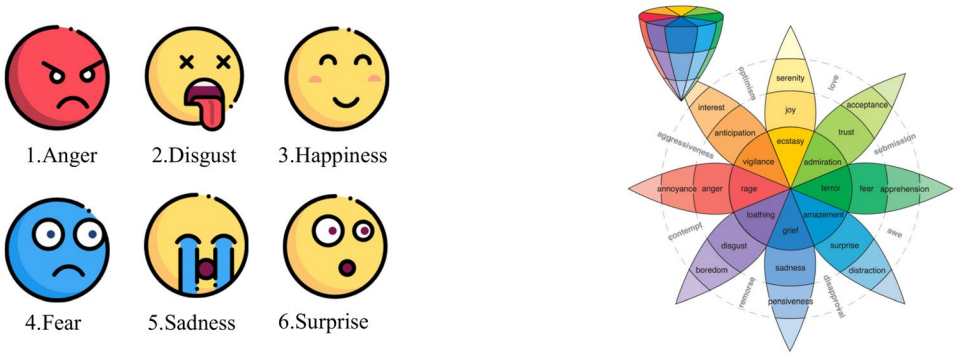
Mental health assessment based on physiological signals has attracted growing research attention, leading to several surveys summarizing progress in emotion recognition, stress and mental disorder detection, as summarized in Table 1.

Early reviews (García-Ponsoda et al. 2023; Ahmed et al. 2022; Pourmohammadi and Maleki 2020) emphasize the importance of effective feature engineering for physiological signal analysis. These studies provide intuitive insights into the feasibility of physiology-based mental health assessment but lack in-depth methodological discussion and do not sufficiently address the practical implementation challenges. In addition, some earlier surveys, such as Charles and Nixon (2019), focused on physiological measurements for mental workload assessment in ergonomics and human-factor applications. However, these studies mainly emphasize traditional physiological analysis and are less focused on AI-driven mental health assessment using modern ML and DL techniques. With the advancement of ML and DL, more recent surveys (Badr et al. 2024; Yasin et al. 2021; Garg et al. 2023; Li et al. 2022; Can et al. 2023; Saganowski et al. 2022) shift toward neural network-based affective computing, highlighting the use of ML algorithms such as Support Vector Machines (SVM), Principal Component Analysis (PCA), and k-Nearest Neighbors (KNN) for effective recognition, as well as DL architectures including CNN, RNN, transformer, and graph neural networks (GNN) for modeling complex affective dynamics. For instance, most works (Badr et al. 2024; Yasin et al. 2021; Garg et al. 2023; Li et al. 2022; Saganowski et al. 2022) primarily focus on single-task research using either EEG or peripheral physiological signals, offering detailed overviews of signal characteristics, feature processing, and classification methods. Recently, Hegde (Hegde and Jayalath 2025) reviewed affective computing techniques for emotional support,

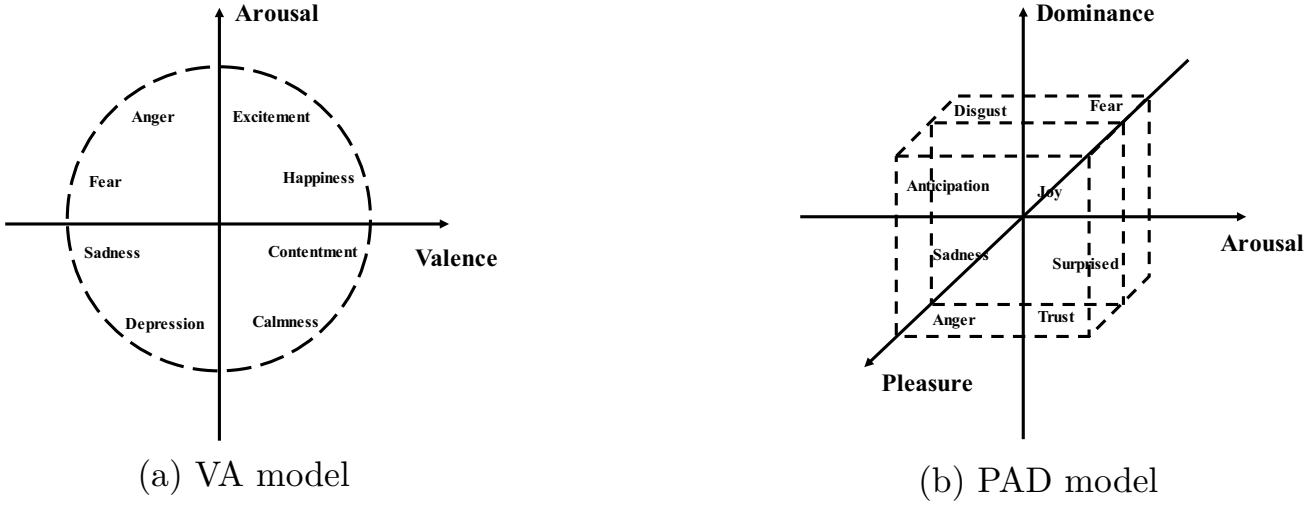
Table 1 Summary of surveys on physiological signal-based mental health assessment

References	Emotion	Stress	Mental disorder	Modality	Technique
García-Ponsoda et al. (2023)	×	×	✓	Single	Feature Engineering
Ahmed et al. (2022)	×	✓	✓	Multiple	Feature Engineering
Pourmohammadi and Maleki (2020)	×	✓	×	Multiple	Feature Engineering
Badr et al. (2024)	×	✓	×	Single	Only DL
Yasin et al. (2021)	×	×	✓	Single	Only DL
Garg et al. (2023)	✓	×	×	Multiple	Only DL
Li et al. (2022)	✓	×	×	Single	ML & DL
Can et al. (2023)	✓	×	×	Multiple	ML & DL
Saganowski et al. (2022)	✓	×	×	Multiple	ML & DL
Ours	✓	✓	✓	Multiple	ML & DL

Fig. 2 Discrete emotion



(a) Six basic emotions (b) Emotion wheel



(a) VA model (b) PAD model

Fig. 3 Dimensional emotion

primarily focusing on text-, audio-, and video-based multimodal AI systems. In contrast, our work specifically investigates physiological signal-based emotion, stress, and mental disorder assessment, with additional emphasis on domain shift, explainability, and multimodal fusion challenges in physiological computing.

3 Affective context application

3.1 Emotion taxonomy

Emotion Taxonomies serve as theoretical foundations of affective computing by providing structured representations of emotional states. Existing models can be broadly categorized into three paradigms: discrete, dimensional, and cognitive.

Discrete emotion model defines a fixed set of fundamental emotions. Ekman's six basic emotions (anger, disgust, fear, happiness, sadness, and surprise) were widely accepted as universal across cultures (Ekman 1999), as shown in Fig. 2 (a). Plutchik later extended this view by introducing an emotion wheel that incorporates hierarchical structure and intensity variation among eight emotions (2003), as illustrated in Fig. 2 (b).

Dimensional model represents emotions within a continuous affective space. The Valence-Arousal (VA) model captures emotional states based on pleasantness and intensity (Russell 1980), as shown in Fig. 3 (a), while the Pleasure-Arousal-Dominance (PAD) model introduces a third dimension, dominance, to better differentiate emotions with similar VA profiles but diverge in dominance. For example, anger is associated with high dominance, reflecting a sense of control or assertiveness, whereas fear is linked to low dominance, indicating feelings of helplessness or submission (Mehrabian and Russell 1974), as illustrated in Fig. 3 (b).

Cognitive theory posits that emotions arise from multilevel assessments and interpretations of situational contexts. In contrast to discrete and dimensional emotion models, cognitive theory emphasizes the role of thought, evaluation, and interpretation in shaping emotional experiences. According to this view, emotions are not automatic or reflexive responses, but are contingent upon the cognitive

processing of events, situations, or information (Yeo and Ong 2024). Thus, cognition plays a central role in emotional appraisal, guiding responses based on an individual's understanding and interpretation of the context (Clore and Ortony 2000; Sander et al. 2005). For example, the same event may evoke different emotional reactions in the same person depending on how it is interpreted, while individuals who perceive an event similarly are likely to experience comparable emotions.

3.2 Stress response

Mental stress arises from perceived threats or excessive demands. Prolonged stress disrupts homeostasis, impairs immune function, and increases vulnerability to both physical and psychological illness (Chrousos and Gold 1992; Maddock and Pariante 2001). Given its impact on emotional regulation and physiological signals, stress detection plays a crucial role in early mental intervention.

3.3 Mental disorder category

Mental disorders are complex conditions characterized by persistent disturbances in emotion, cognition, and behavior, making them central targets for affective computing and computational psychiatry. Guided by the DSM-5 criteria (American Psychiatric Association and Association 2013), we highlight six major categories relevant to affective and cognitive abnormalities.

Depressive and anxiety disorders represent two of the most prevalent and frequently co-occurring conditions. Depression is marked by persistent sadness, loss of interest, and impaired functioning, whereas anxiety is defined by excessive fear and anticipatory tension. Both can lead to heightened irritability, disrupted sleep, and difficulty concentrating, with chronic forms often contributing to broader cognitive dysfunction.

Bipolar disorder bridges depressive and manic states, reflecting rapid shifts in mood, energy, and activity levels. Its manic episodes are characterized by impulsivity and heightened arousal, overlapping with some symptoms of anxiety and risk-taking behaviors observed in certain schizophrenia subtypes.

Schizophrenia is distinguished by profound alterations in perception and thought—such as hallucinations and delusions—alongside social and occupational dysfunction. It shares affective instability with bipolar disorder, particularly in schizoaffective presentations.

Obsessive-compulsive disorder (OCD) and **post-traumatic stress disorder (PTSD)** both involve heightened emotional reactivity and intrusive cognitive patterns. OCD is driven by recurrent obsessions and compulsive behaviors aimed at anxiety relief, while PTSD emerges from trauma exposure and involves re-experiencing, avoidance, and hyperarousal. Both disorders underscore the role of maladaptive emotion regulation and memory processing in mental illness.

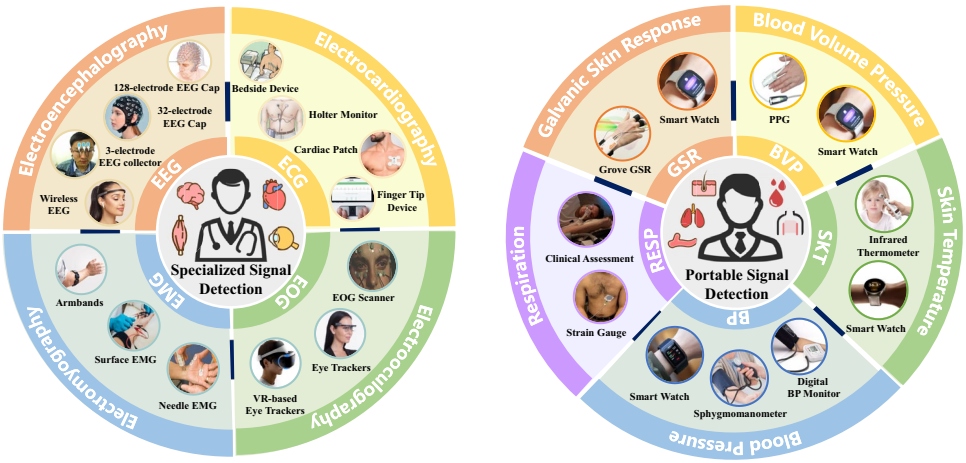
4 Physiological signal application

4.1 Physiological signal

In mental health assessment, specialized physiological signals such as electroencephalography (EEG), electrocardiography (ECG), electromyography (EMG), and electrooculography (EOG) are widely used for fine-grained analysis of mental states, owing to the high resolution of electrode-based acquisition devices in clinical settings. In contrast, physiological signals such as galvanic skin response (GSR), blood volume pulse (BVP), blood pressure (BP), skin temperature (SKT), and respiration (RESP) can be acquired via wearable sensors, allowing continuous monitoring in daily life. However, due to the fact that wearable sensors typically employ fewer channels, lower sampling rates, and minimal shielding, their recordings are far more susceptible to motion and environmental artifacts. These features tend to be much noisier and lower temporal resolution than in clinical-grade devices. The acquisition devices for both types of signals are illustrated in Fig. 4. Meanwhile, Table 2 and Table 3 summarize the key specialized physiological features and pervasive physiological features for mental health assessment, respectively. This review focuses on the distinctive characteristics of physiological signals in mental health assessment and summarizes the dynamic changes observed during emotional regulation.

4.1.1 EEG

EEG signal records the brain's electrical activity generated by the synchronized firing of large populations of neurons. These subtle voltage fluctuations are detected by electrode-based devices, such as 128-electrode elastic cap, 32 electro-gel EEG cap, wearable 3-electrode EEG collector, or commercial wireless brain-tracking devices. EEG signals are closely linked to mental health, as patterns of brain electrical activity can reflect emotional fluctuations (Ding et al. 2022; Guo et al. 2023), stress responses (Liu et al. 2020; Fang et al. 2022) and various mental disorders, including depression, anxiety, bipolar, schizophrenia disorders and OCD (Barros et al. 2021; Ma et al. 2025). After rigorous preprocessing to eliminate artifacts from eye movements, muscular activity, and other non-cerebral electrical signals,



(a) Specialized acquisition device (b) Portable acquisition device

Fig. 4 Examples of signal acquisition devices. (**a**) Specialized acquisition device. (**b**) Portable acquisition device

EEG signals can be classified by their frequency ranges on brain regions, divided into five main bands: δ, θ, α, β, and γ.

4.1.2 ECG

ECG captures cardiac electrical activity during ventricular depolarization and repolarization, initiated by the sinoatrial node (Kashou et al. 2017). They are acquired via chest or limb electrodes using Holter monitors or bedside systems. ECG is widely employed in mental health research, as heart rate (HR) and heart rate variability (HRV) exhibit alterations under stress (Pourmohammadi and Maleki 2020), PTSD (Thiagarajan 2023), anxiety (Baygin et al. 2024), depression (Geng et al. 2023; Khare et al. 2023), and bipolar disorder (Polcwiartek et al. 2021). HRV is analyzed in both time and frequency domains. Time-domain metrics, such as the standard deviation of normal-to-normal intervals (SDNN) and the root mean square of successive differences (RMSSD), are derived from RR intervals between consecutive R-peaks. Frequency-domain analysis decomposes HRV into low-frequency (LF, $0.04 - 0.15$ Hz) and high-frequency (HF, $0.15 - 0.4$ Hz) bands, reflecting sympathetic and parasympathetic activity, respectively.

4.1.3 EMG

EMG measures the electrical signals produced by skeletal muscle contractions, providing information about muscle function and neural activity. These signals are captured via surface or needle electrodes using conventional EMG systems or wearable armbands. EMG has been widely applied to monitor emotional states (Koelstra et al. 2011; Sharma et al. 2019), stress (Schmidt et al. 2018), and disorders such as PTSD and schizophrenia (Chan et al. 2022; Khare et al. 2023) through biofeedback. As shown in Table 2, relevant features include Root Mean Square (RMS), Integrated EMG (IEMG), and Power Spectral Density (PSD), as well as median frequency, which together reflect muscle response under emotional fluctuations.

4.1.4 EOG

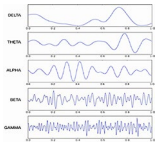
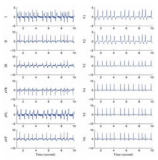
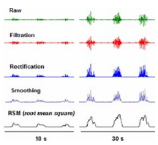
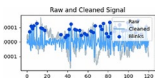
EOG records the voltage difference between the cornea and retina, which changes with eye movements and blinks. These signals are typically collected using electrodes placed around the eyes, or through integrated sensors in smart glasses and virtual reality (VR) devices. EOG patterns, such as saccades, blinks, and fixations are associated with emotional regulation (Kose et al. 2021; Cai et al. 2021), psychological stress (Wu et al. 2022), and mental disorders including depression and anxiety (Zhu et al. 2019; Guo et al. 2020).

4.1.5 GSR

GSR, or electrodermal activity (EDA), measures sweat-induced changes in skin conductivity, governed by the sympathetic nervous system.

This signal is commonly acquired using electrodes attached to the fingers or wrist. EDA signals are generally decomposed into two components: Skin Conductance Level (SCL) and Skin Conductance Response (SCR). Specifically, SCL represents the slowly varying tonic component associated with sustained autonomic arousal, whereas SCR reflects rapid phasic fluctuations triggered by transient emotional or stress-related stimuli. During emotional arousal or stress responses, sympathetic nervous system activation increases sweat gland activity, thereby altering skin conductivity. Due to its high sensitivity to autonomic nervous activity, GSR has emerged as an effective physiological indicator for detecting emotions (Domínguez-Jiménez et al. 2020; Shukla et al. 2019), stress (Schmidt et al. 2018; Greco

Table 2 Key Specialized Physiological Features for Mental Health Assessment

Feature	Description	Function	Data Example
EEG			
δ band	0.5-3 Hz	Deep, dreamless sleep; unconscious mind	
θ band	4-7 Hz	Subconscious mind; sleep and dreaming	
α band	8-13 Hz	Relaxed alertness; parietal/occipital lobes; brain inactivation	
β band	14-30 Hz	Active mental states; focused tasks; frontal cortex	
γ band	31-50 Hz	Hyper brain activity; higher cognition; complex processing	
ECG			
HR	The number of heartbeats per minute.	Increases during arousal and decreases during relaxation	
SDNN	Standard deviation of RR intervals	Overall heart rate variability; High value indicates relaxation and low value indicates stress	
RMSSD	Root mean square of RR intervals	Reflects parasympathetic activity	
LF	0.04-0.15 Hz	Reflects sympathetic activity	
HF	0.15-0.4 Hz	Reflects parasympathetic activity	
EMG			
RMS	Root Mean Square intensity	Overall muscle activation level; Indicates emotional response	
IEMG	Integrated muscle activity	Total muscle activity over time; Indicates prolonged emotional states	
Median Freq.	Middle value in the frequency	Detects muscle fatigue; associates with ongoing emotional stress	
PSD	Power spectral density	Captures dynamic changes in muscle activity; Reflects emotional fluctuations	
EOG			
Saccade	Rapid eyes movement	Reflects high arousal emotions	
Blink	Closing and opening of the eyelids	Reacts to anxiety or stress	
Fixation	Eyeballs stay in one position	Reflects emotional attention and interest	

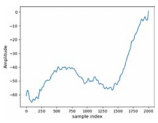
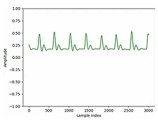
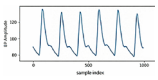
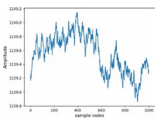
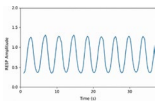
et al. 2021), and mental disorders (Hassantabar et al. 2022; Šalkevicius et al. 2019).

4.1.6 BVP

BVP measures blood volume changes in tissue microvasculature. As the heart pumps blood, peripheral vessels experience volume fluctuations, altering light absorption or reflection at specific wavelengths. Photoplethysmography (PPG) sensor detect these changes, converting them into electrical signals, and are commonly integrated into smartwatch or clip-on device.

Several representative physiological features can be derived from BVP signals, including Heart Rate (HR), Pulse Amplitude, and Inter-Beat Interval (IBI). Specifically, HR reflects emotional arousal and stress-related cardiovascular activity, Pulse Amplitude indicates peripheral blood perfusion and autonomic regulation, while IBI characterizes beat-to-beat cardiac rhythm variations associated with stress and emotional states. These features have been widely applied

Table 3 Key pervasive physiological features for mental health assessment

Feature	Description	Function	Data Example
GSR/EDA			
SCL	Skin Conductance Level	Slowly varying tonic component reflecting sustained sympathetic nervous system arousal and long-term emotional stress	
SCR	Skin Conductance Response	Rapid phasic conductance fluctuation associated with transient emotional activation and stress response	
BVP			
HR	Heart rate derived from pulse signals	Reflects emotional arousal and autonomic nervous system activity	
Pulse Amplitude	Magnitude of blood volume fluctuation	Indicates cardiovascular activation and emotional variation	
IBI	Inter-beat interval between consecutive pulse peaks	Reflects heart rhythm variability related to stress and relaxation states	
BP			
SBP	Systolic Blood Pressure	Reflects stress-induced cardiovascular activation and sympathetic response	
DBP	Diastolic Blood Pressure	Indicates vascular resistance and emotional stress regulation	
SKT			
Peripheral Temp	Peripheral skin temperature variation	Reflects thermoregulation changes associated with stress, anxiety, and emotional arousal	
RESP			
Respiration Rate	Breathing frequency per minute	Reflects emotional stress, anxiety level, and autonomic nervous activity	
Respiration Amplitude	Depth variation of respiratory signals	Indicates emotional regulation and mental workload changes	

in stress detection (Zhou et al. 2022), emotion recognition (Domínguez-Jiménez et al. 2020), and depressive disorder assessment (Casado et al. 2023).

4.1.7 BP

BP represents the hemodynamic force exerted by circulating blood on arterial walls, manifesting as measurable pressure waves that propagate through the vascular system. Current measurement devices encompass sphygmomanometers, digital BP monitors, and wearable smartwatches that support continuous or periodic cardiovascular tracking. BP signals are commonly characterized by systolic blood pressure (SBP) and diastolic blood pressure (DBP), which respectively reflect arterial pressure during cardiac contraction and relaxation phases. Variations in SBP and DBP are closely associated with autonomic nervous system activity and emotional regulation, with notable responses observed under different emotional states (Dai et al. 2017), stress conditions (Tomitani et al. 2021), and mental disorders such as depression, anxiety, and bipolar disorder (Lim et al. 2023).

4.1.8 SKT

SKT reflects heat dissipated from the skin due to metabolic activity and is modulated by autonomic nervous system responses. Sympathetic activation can trigger vasodilation, elevating skin temperature during stress or emotional arousal. SKT can be measured using infrared thermometers and smartwatches. Fluctuations in SKT are associated with emotional states (Zhang et al. 2023; Khan et al. 2021), stress (Anusha et al. 2018; Schmidt et al. 2018), and mental disorders such as anxiety and schizophrenia (Šalkevicius et al. 2019; Bufano et al. 2023).

4.1.9 RESP

RESP indicates the rhythmic patterns of inhalation and exhalation, with measurements typically obtained using spirometers for clinical assessments or strain gauges to detect chest expansion. Generated by the movement of diaphragmatic and intercostal muscles, RESP signal facilitates gas exchange and capture both physiological and psychological responses to external stimuli. For example, variations in RESP rate and rhythm can signal emotional states such as tension or anger, and their fluctuations serve as objective markers of emotional responses (Koelstra et al. 2011; Zhang et al. 2023) and mental disorders (Harrison et al. 2021).

4.2 Public physiological dataset based on mental health

This section provides an overview of publicly available physiological signal datasets commonly used for the detection of emotional states, stress, and mental health disorders. Table 4 presents a comprehensive summary of key characteristics, including the number of participants, types of physiological signals, stimulus materials, and assessment methods. Additionally, we briefly describe the most widely used datasets across different application domains.

4.2.1 DEAP

DEAP dataset (Koelstra et al. 2011) comprises EEG and peripheral physiological signals, including GSR, RESP, SKT, ECG, BVP, EMG, and EOG, collected from 32 subjects exposed to 40 different music video stimuli lasting 1 min each. During the evaluation phase, subjects provide self-reports on dimensions such as arousal, valence, dominance, liking, and familiarity. The first four dimensions are rated on a scale from 1 to 9, while familiarity is rated on a scale from 1 to 5.

Table 4 Public dataset based on physiological signals

Dataset	Subjects	Stimuli	Physiological Signals	Assessment	Website
Emotion Recognition					
DEAP (Koelstra et al. 2011)	32	music videos	EEG, GSR, RESP, SKT, ECG, BVP, EMG, EOG	Arousal, Valence, Dominance Liking, Familiarity	link
EMDB (Carvalho et al. 2012)	32	film clips	GSR, ECG	Arousal, Valance, Dominance	link
SEED (Zheng and Lu 2015)	15	film clips	EEG	Positive, Neutral, Negative	link
DECAF (Abadi et al. 2015)	30	music videos, film clips	EMG, EOG, ECG	Arousal, Valence	link
DREAMER (Katsigiannis and Ramzan 2017)	23	audio-visuals	EEG, ECG	Arousal, Valence, Dominance	link
MAHNOB-HCI (Lichtenauer and Soleymani 2011)	27	videos, images	EEG, ECG SKT, GSR, RESP	Arousal, Valence	link
RECOLA (Ringeval et al. 2013)	46	interaction	ECG, GSR	Arousal, Valance	link
CASE (Sharma et al. 2019)	30	videos	ECG, BVP, EMG, GSR, RESP, SKT	Arousal, Valence	link
MPED (Song et al. 2019)	23	videos	EEG, GSR, RESP, ECG	Arousal, Valence, Different Emotion Scale	link
ASCERTAIN (Subramanian et al. 2016)	58	film clips	EEG, ECG, GSR	Valance, Arousal	link
AMIGOS (Miranda-Correa et al. 2018)	40	videos	EEG, ECG, GSR	Valance, Arousal, Dominance, Liking, Familiarity, Emotion Keywords	link
DAPPER (Shui et al. 2021)	88	not mentioned	HR, GSR	TIPI-C, PANAS, Arousal, Valance	link
K-EmoCon (Park et al. 2020)	21	videos, debates	ECG, BVP, EMG, GSR, HR, SKT	Arousal, Valence Emotional Keywords	link
Stress Detection					
WESAD (Schmidt et al. 2018)	15	book, video, TSST	ECG, BVP, EMG, GSR, RESP, SKT	PANAS, STAI, SAM, SSSQ	link
TILE (Mundnich et al. 2020)	212	working	HR, ECG, EEG	PANAS, STAI, PSS	link
SWELL (Koldijk et al. 2014)	25	working	ECG, GSR	Arousal, Valance, Stress Scale	link
DRIVEDB (Healey and Picard 2005)	17	driving	ECG, EMG, GSR, RSP	Scale, Ranking of Events	link
Mental Diagnosis					
MODMA (Cai et al. 2022)	53/55	images	EEG	GAD-7, PHQ-9	link
HBN (Alexander et al. 2017)	10000	film clips, visuals	EEG, EOG	Clinician and Self administration	link

4.2.2 SEED

The SEED dataset (Zheng and Lu 2015) consists of 62-channel EEG and eye-tracking data collected from 12 subjects, along with additional 62-channel EEG recordings from 3 more subjects. These participants view 15 Chinese film clips, each lasting 4 min. After each clip, they complete a questionnaire to provide emotional feedback, with ratings for positive, neutral, and negative emotions. The SEED dataset has since been expanded into multiple versions, including SEED-IV, SEED-V and SEED-VII etc. These versions reflect variations in participant numbers, emotional categories, test scenarios, stimulus clips, and participant nationalities.

4.2.3 CASE

The dataset (Sharma et al. 2019) includes 32-channel physiological signals obtained from ECG, BVP, EMG, GSR, RESP, and SKT sensors. Simultaneously, 30 subjects are required to represent their emotional experiences as they manipulate a red pointer within suitable regions on valence and arousal axes, to achieve continuous affect annotations. The experiment incorporate 5 short videos, each lasting about one minute, interspersed with 10-second transitional blue screens. These videos aim to elicit emotions of excitement, sadness, relaxation, fear, and joy.

4.2.4 K-EmoCon

The dataset (Park et al. 2020) includes multiple modalities, such as audio-visual recordings, EEG, and peripheral physiological signals like BVP, GSR, SKT, and HR. These signals are collected during 16 paired debates on social issues, with each session lasting approximately 10 min. Distinguishing itself from earlier datasets, it incorporates emotion annotations from multiple sources: self-reports, debate partners, and external observers. These annotations, provided at 5-second intervals during the review of debate recordings, cover dimensions such as arousal and valence, rated on a scale of 1 to 5, along with 18 discrete emotional states.

4.2.5 WESAD

The dataset (Schmidt et al. 2018) leverages chest-worn and wrist-worn devices to capture BVP, GSR, SKT, and acceleration data. These signals are collected from 17 participants across three distinct emotional states: neutral, stress, and amusement. During data collection, participants engage in a series of activities, including reading, watching 11 video clips, and completing the Trier Social Stress Test (TSST). Subsequently, participants are assessed with scoring criteria to evaluate their emotions states. These include the Positive and Negative Affect Schedule (PANAS), rated on a five-point scale, State-Trait Anxiety Inventory (STAI), scored on a four-point scale, Self-Assessment Manikins (SAM), and Short Stress State Questionnaire (SSSQ). These scoring methods are designed to capture subjective emotional experiences under different conditions.

4.2.6 MODMA

This dataset (Cai et al. 2022) focuses on depression recognition, collecting multi-modal data from depressed patients and healthy controls in both resting and stimulated states. EEG recordings include 53 subjects (24 depressed, 29 controls) using a 128-channel cap in both conditions, and 55 subjects (26 depressed, 29 controls) using a wearable 3-electrode device in the resting state. During rest, subjects close their eyes; during stimulation, they view emotional facial images (fearful, sad, happy, neutral). All patients are diagnosed based on DSM-IV criteria, with PHQ-9 scores greater than or equal to 5.

5 Fundamental AI model

Advancements in information technology have propelled ML and DL as pivotal tools in mental health assessment. Traditional ML methods, such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Decision Trees (DT), typically operate on interpretable, low-dimensional handcrafted features, allowing for direct visualization and clearer understanding of model decisions. In contrast, DL techniques, including Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Graph Neural Network (GNN), automatically learn high-dimensional representations from raw data, offering superior performance but reduced transparency. Figure 5 illustrates representative AI model architectures commonly employed in mental health assessment based on physiological signals.

5.1 Machine learning technique

SVM tackles nonlinear problems by implicitly projecting input features into high-dimensional spaces via kernel functions, thereby enabling their effective linear separation. For example, Liao et al. (2017) apply SVM for depression diagnosis by transforming EEG data to achieve classification through a hyperplane. KNN classifier samples based on the majority class among their nearest neighbors in the feature space. Wang et al. (2013) utilize KNN to evaluate driver stress by analyzing EEG features, leveraging distance-based proximity to capture local distributions. DT segments the

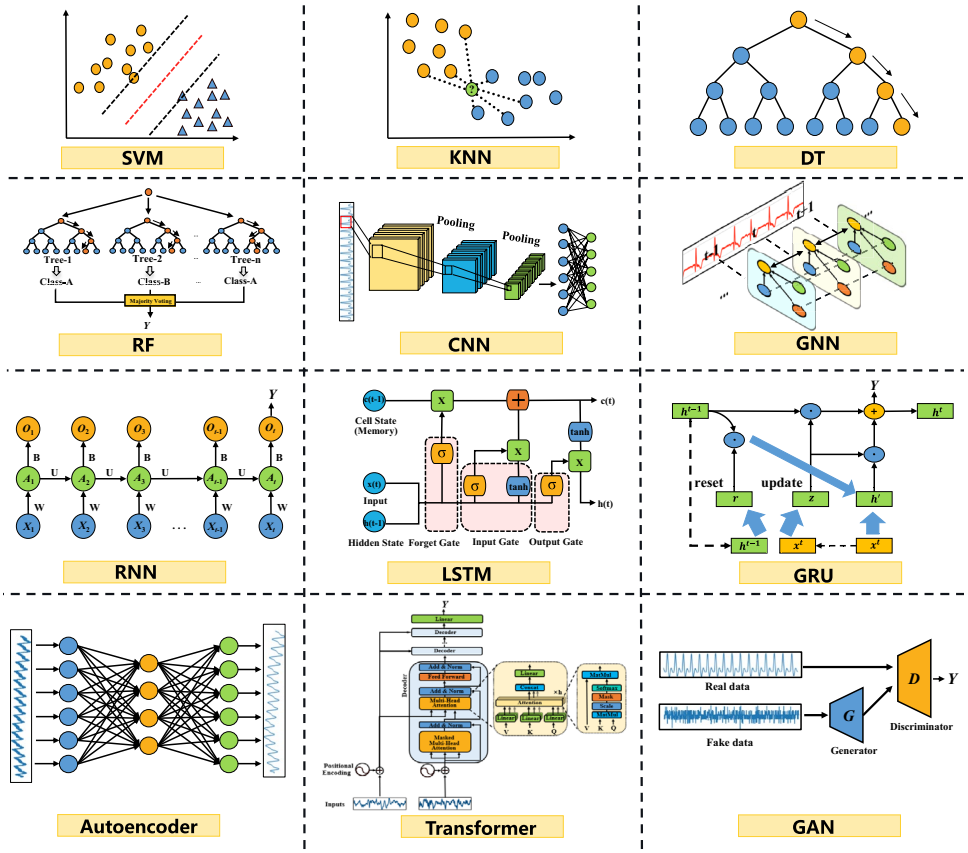


Fig. 5 Fundamental artificial intelligence models

feature space through a hierarchical structure, yielding interpretable rules with high computational efficiency. Their transparency makes them suitable for mental diagnosis, such as Alzheimer (Lehmann et al. 2007) and depression (Acharya et al. 2015). However, DT is susceptible to overfitting and limited generalization. To mitigate these issues, ensemble methods are introduced. Random Forest (RF) aggregates results from multiple trees trained on bootstrapped samples. Boosting-based methods, including XGBoost and LightGBM (Gayathri et al. 2022), construct decision trees sequentially by focusing on the residual errors of previous iterations, to enhance predictive performance.

5.2 Deep learning technique

DL technique has become a dominant paradigm in physiological signal-based mental health assessment due to its strong capability in modeling complex spatial, temporal, and cross-modal patterns from high-dimensional data. Existing DL approaches can generally be divided into two complementary streams: discriminative architectures for mental state classification and generative or representation-learning architectures for data enhancement and latent feature modeling. The former focuses on learning discriminative spatiotemporal representations for prediction tasks, whereas the latter aims to improve data robustness, alleviate modality discrepancies, and enhance representation quality.

5.2.1 Discriminative architectures for classification

Discriminative DL models, including CNN, RNN, GNN, and Transformer-based architectures, are primarily designed to map physiological signals to specific emotional or mental health states by capturing informative spatial-temporal dependencies.

Among these approaches, CNN and RNN are the most widely adopted architectures in physiological signal analysis.

CNNs are particularly effective in extracting local spatial patterns and have demonstrated strong performance in EEG-based mental health assessment tasks (Afify et al. 2025). In contrast, RNNs, especially Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), are capable of modeling temporal dependencies by preserving sequential information across time steps. For example, Kuang et al. (2024) utilize LSTM gating mechanisms to capture long-term physiological dynamics for emotion recognition, while Khan et al. (2021) employ GRU to reduce computational complexity through simplified gating operations. To jointly exploit spatial and temporal characteristics,

hybrid CNN-RNN frameworks have also been proposed (Song et al. 2022).

Beyond Euclidean feature modeling, Graph Convolutional Networks (GCNs) extend convolution operations to graph-structured data, enabling the extraction of topologically informed representations from physiological signals with intrinsic spatial relationships. For instance, Chen et al. (2022) model EEG electrodes as graph nodes and construct spatial-temporal connections between channels for depression recognition.

More recently, Transformer-based architectures have attracted increasing attention due to their self-attention mechanisms, which enable dynamic modeling of long-range dependencies and modality importance. Zheng et al. (2023) employ Transformer networks to capture cross-modal interactions and salient features, thereby improving both prediction performance and interpretability in sentiment analysis tasks.

5.2.2 Generative architectures for data optimization

Unlike discriminative models that primarily focus on decision-making, generative and representation learning architectures aim to address intrinsic challenges in physiological data, including noise contamination, limited labeled samples, and cross-domain distribution discrepancies.

Autoencoder (AE) learns compact and informative latent representations by encoding high-dimensional inputs into low-dimensional feature spaces and reconstructing the original signals through a decoder. Various AE variants have been proposed to improve robustness and representation quality under different scenarios. For example, Wang et al. (2022) propose a Multi-modal Domain Adaptive Variational Autoencoder (MMDA-VAE) to learn shared latent representations from EEG and eye-movement signals, thereby mitigating cross-subject discrepancies. Other variants, including denoising AE, sparse AE, and convolutional AE, have also been explored to enhance noise robustness and feature sparsity (Li et al. 2015; Qiu et al. 2018; Wen and Zhang 2018).

Similarly, Generative Adversarial Networks (GANs) provide an effective framework for data augmentation and domain adaptation. GAN training is formulated as a min-max adversarial process, in which the generator attempts to synthesize realistic samples while the discriminator distinguishes generated data from real observations (Goodfellow et al. 2014). Chen et al. (2019) propose Emotional GAN to generate realistic EEG signals, thereby improving model generalization under limited labeled data conditions.

In summary, discriminative architectures mainly contribute to downstream mental state prediction, whereas generative and representation-learning models enhance data quality and latent feature robustness. The combination of these paradigms has significantly advanced the accuracy, robustness, and interpretability of DL-based mental health assessment systems.

6 Open issue and current solution

This section discusses three pivotal challenges in physiological signal-based mental health assessment: domain shift, model interpretability, and multi-modal fusion, together with the representative technical solutions developed to address them.

Although these challenges have been extensively studied in the broader AI literature, they become substantially more complicated in physiological mental health assessment due to the strong inter-subject variability of physiological responses, the abstract and non-intuitive nature of physiological signals, and the heterogeneity across multimodal biological systems. Furthermore, while existing methodologies have demonstrated promising performance in conventional affective computing tasks such as emotion recognition, their applicability to more clinically complex tasks, including mental disorder assessment and long-term psychological monitoring, remains insufficiently explored.

6.1 Domain shift

In physiological signal-based mental health assessment, domain shift poses a particularly significant challenge to model generalization due to the strong inter-subject variability and non-stationary characteristics of physiological responses. Unlike conventional vision or language tasks, physiological reactions to the same mental state may vary substantially across individuals because of intrinsic biological differences, emotional sensitivity, and environmental influences. Moreover, mental states such as stress, anxiety, and depression often exhibit overlapping and ambiguous physiological patterns, further increasing the difficulty of learning robust cross-domain representations. These factors lead to considerable distribution discrepancies across subjects and datasets, resulting in reduced accuracy and reliability on unseen data.

This issue is generally categorized into two scenarios: cross-subject (Li et al. 2024; Guo et al. 2023; Liu et al. 2024; Gu et al. 2023; Pan et al. 2024; She et al. 2023) and cross-dataset (Liao et al. 2024; Zhou et al. 2023; Jiang et al. 2023; Yuan et al. 2023). Cross-subject testing involves training on data from one group (e.g., Subject Group A) and testing on another group (e.g., Subject Group B) within the same dataset. In contrast, cross-dataset testing involves training on one dataset (e.g., Source Dataset) and evaluating on another

(e.g., Target Dataset). Most existing studies focus primarily on cross-subject scenarios, since real-world mental health applications commonly encounter unseen users with diverse physiological characteristics.

The domain shift problem mainly arises from two aspects: individual differences and testing differences (Hong et al. 2024). Individual differences originate from subject-specific physiological characteristics, leading to substantial variations in signal patterns even among individuals with the same mental condition. Testing differences refer to signal variability caused by differences in acquisition devices, sampling rates, noise levels, and preprocessing strategies, which are particularly prominent in cross-dataset settings.

To tackle these challenges, domain adaptation (DA) and domain generalization (DG) are promising approaches.

Domain Adaptation: focuses on adapting a model trained on a labeled source domain to perform well on a target domain with limited or no labeled data. Formally, this involves labeled source domain $\mathcal{D}_s = \{x_i^s, y_i^s\}_{i=1}^{n_s}$ and target domain with either limited labeled data $\mathcal{D}_t = \{x_j^t, y_j^t\}_{j=1}^{n_t}$ or unlabeled data $\mathcal{D}_t = \{x_j^t\}_{j=1}^{n_t}$, both with identical feature and label spaces but differing marginal distributions ($p(x^s) \neq p(x^t)$). The goal of DA is to minimize the domain discrepancy, enabling reliable performance in target domains with scarce labeled data.

Domain Generalization: Distinct with DA, DG focuses on learning a model from multiple source domains that generalizes well to unseen target domains without accessing any target domain data during training. Formally, given multiple labeled source domains $\mathcal{D}_s = \{x_j^s, y_j^s\}_{j=1}^{n_s}$ drawn from different distributions ($p(x_s) \neq p(x_s')$ for $s \neq s'$), the objective is to learn domain-invariant representations that perform well on an unseen domain $\mathcal{D}_t$. DG aims to address the generalization challenge caused by the inter-subject variability and heterogeneity in affective signals, enabling robust model performance in real-world mental health assessment scenarios where the target environment or user is unknown during training.

In the remainder of this section, considering the versatility of various techniques in addressing both DA and DG, the following discussion provides a detailed introduction organized from the perspectives of **Feature-level alignment methods** and **Strategy-level learning schemes**, followed by an analysis of their specific implementations within these two cross-domain scenarios. An overview of the domain shift methods is presented in Fig. 6.

6.1.1 Feature-level alignment methods

The primary objective of feature-level alignment methods is to map features from diverse domains into a shared latent space to eliminate distribution discrepancies. This goal is realized through adversarial-based feature alignment and metric-based feature alignment, which synchronize data distributions via implicit adversarial competition or explicit statistical distance minimization.

Adversarial-based feature alignment: This method leverages the principles of generative adversarial networks to match feature distributions without explicitly defining a distance metric. GAN provides a foundational adversarial framework. Building upon this idea, the Domain Adversarial Neural Network (DANN) is designed to learn domain-invariant features through a min-max optimization game. DANN consists of feature extractor $G_f(\cdot;\theta_f)$, label predictor $G_y(\cdot;\theta_y)$, and domain classifier $G_d(\cdot;\theta_d)$, where θ_f, θ_y, and θ_d are trainable parameters. The training objective is formulated as:

$$\min_{\theta_f,\theta_y}\max_{\theta_d}\left[\frac{1}{n_s}\sum_{i=1}^{n_s}\mathcal{L}_{CE}(G_y(G_f(x_i^s)),y_i^s)-\lambda\left(\frac{1}{n_s}\sum_{i=1}^{n_s}\mathcal{L}_{CE}(G_d(G_f(x_i^s)),0)+\frac{1}{n_t}\sum_{j=1}^{n_t}\mathcal{L}_{CE}(G_d(G_f(x_j^t)),1)\right)\right] \tag{1}$$

where $\mathcal{L}_{CE}$ denotes the cross-entropy loss and λ is a balancing hyperparameter. The objective encourages the extraction of features that are predictive for the main task while being indistinguishable across domains.

Unlike the shared encoder architecture in DANN, ADDA employs separate feature extractors to implement asymmetric mappings, allowing for more flexible, tailored representations for each domain. The training objective also follows min-max game, formulated as:

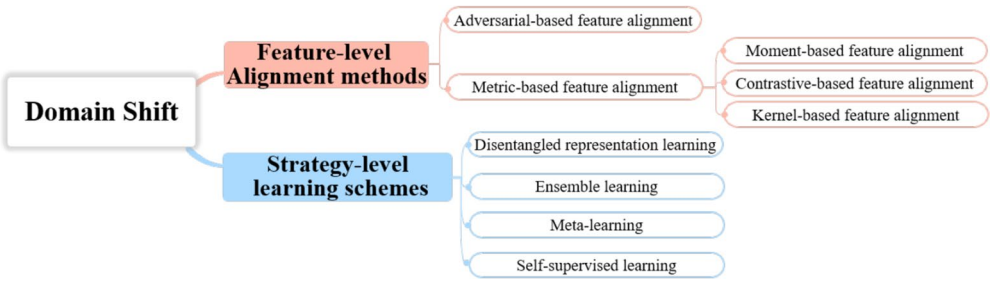


Fig. 6 The overview of domain shift

$$\min_{F_t}\max_{D}\ \mathbb{E}_{x_s\sim\mathcal{D}_s}[\log D(F_s(x_s))]+\mathbb{E}_{x_t\sim\mathcal{D}_t}[\log(1-D(F_t(x_t)))] \tag{2}$$

where F_s and F_t denote the feature extractors for the source and target domain. By employing separate feature extractors for each domain, ADDA introduces asymmetric mappings that allow each domain to learn tailored representations. For instance, He et al. (2022a, 2022b) further explore Adversarial Discriminative Domain Adaptation (ADDA) in cross-subject emotion recognition. Whether for DANN, ADDA, or other variants, the core of adversarial alignment remains the deployment of the Gradient Reversal Layer (GRL), which reverses gradients from the domain classifier to encourage domain confusion in the feature space.

Depending on the availability of domain labels, adversarial-based alignment methods pursues different learning objectives. The most representative application occurs in DA, where domain labels are often defined based on subject identities. Neural network simultaneously processes labeled source data and unlabeled target data during the training phase. The adversarial struggle between the discriminator and the feature extractor enables the model to discard domain-specific variances, effectively mapping physiological features into a unified manifold space to extract robust, domain-invariant representations (Li et al. 2024). Moreover, the definition of domain labels is not restricted to subject identities. It can be extended to broader dimensions, such as dataset and center domain (Wang et al. 2024). Beyond traditional global alignment methods, An et al. (2025) designs dual alignment of local feature patterns, specifically spatial-temporal and spatial-spectral modes, and global fused features through the deployment of multiple GRLs. This hierarchical adversarial strategy facilitates the learning of more discriminative and complementary information.

Metric-based feature alignment: This approach offers an alternative paradigm for distribution matching through the minimization of explicit statistical distances between domains. This category of approach focuses on the direct quantification of domain discrepancy, which enables the alignment of multiple domain distributions through mathematical measures. Specific implementations of these mathematical measures encompass moment-based, contrastive-based, and kernel-based alignment approaches.

(1) **Moment-based feature alignment**: Moment-based feature alignment methods reduce distribution discrepancies by matching statistical moments across domains. The general objective can be formulated as:

$$\mathcal{L}_{moment}=\sum_{k=1}^{K}\left|\frac{1}{n_s}\sum_{i=1}^{n_s}(G_f(x_i^s))^k-\frac{1}{n_t}\sum_{j=1}^{n_t}(G_f(x_j^t))^k\right| \tag{3}$$

here $G_f(\cdot)$ denotes the feature extractor, k represents the order of the statistical moment, K is the highest order considered for alignment, and n_s and n_t denote the numbers of samples in the source and target domains, respectively.

In the work of Gao et al. (2025), a covariance-based alignment method is proposed to minimizes the Frobenius norm between covariance matrices. By enforcing this constraint, the feature extractor is encouraged to learn subject-invariant representations, which effectively reduce the statistical discrepancy between the source domain and the target subject. Beyond second-order statistics, some studies further align higher-order moments to capture more complex distribution characteristics. For example, Central Moment Discrepancy (CMD) (Jia et al. 2024) aligns multiple orders of central moments between domains, enabling more comprehensive distribution alignment. However, such methods often introduce additional computational complexity and may become unstable when the feature dimensionality is high.

(2) **Contrastive-based feature alignment**: The core mechanism of contrastive learning centers on pulling similar samples together while pushing dissimilar ones apart in the representation space. This paradigm has found extensive application in general representation learning and has recently emerged as a significant strategy for the extraction of domain-invariant features.

For instance, Shen et al. (2022) present a DG-based method called CLISA. The method includes a contrastive learning phase followed by a prediction phase. Specifically, a CNN encoder learns invariant representations by comparing positive pairs, which are EEG signals from different subjects exposed to the same stimulus, with negative pairs that originate from different stimuli. Furthermore, Cai and Pan (2023) design a DG approach for EEG-based emotion recognition across subjects. They introduce a two-phase framework centered on prototypical contrastive learning. Initially, the model learns a matching function through contrastive loss. This function is refined using an iterative process that updates matches based on the closest same-class prototypes. This strategy aims to learn domain-invariant features by minimizing the distance between inputs and prototypes of the same class, in contrast to those from different classes.

(3) **Kernel-based feature alignment**: This method measures domain discrepancies by embedding feature distributions into a reproducing kernel Hilbert space (RKHS) and computing distances between the embedded distributions. A representative approach is Maximum Mean Discrepancy (MMD), whose empirical formulation can be written as:

$$\mathcal{L}_{MMD}=\left\|\frac{1}{n_s}\sum_{i=1}^{n_s}\phi(G_f(x_i^s))-\frac{1}{n_t}\sum_{j=1}^{n_t}\phi(G_f(x_j^t))\right\|_{\mathcal{H}}^2 \tag{4}$$

where $G_f(\cdot)$ denotes the feature extractor, $\phi(\cdot)$ represents the implicit feature mapping induced by the kernel function, and $\|\cdot\|_{\mathcal{H}}$ denotes the norm in the RKHS $\mathcal{H}$. n_s and n_t are the numbers of samples in the source and target domains, respectively.

To further enhance its effectiveness, several variants have been proposed, such as multi-kernel MMD (Liu et al. 2025), joint MMD (Tang et al. 2025), class-wise MMD (Xu et al. 2025), and dynamic MMD (Shi et al. 2023), which aim to capture more complex distribution discrepancies across domains. However, their performance is sensitive to kernel selection and may incur higher computational costs when applied to high-dimensional data.

Overall, feature alignment methods possess broad applicability, yet most of them require domain labels that are not always available. Furthermore, to achieve effective alignment between different domain distributions, certain methods require substantial data to accurately estimate the underlying distributions.

6.1.2 Strategy-level learning schemes

This category provides a holistic framework to coordinate training protocols and facilitate cross-domain knowledge transfer. It encompasses disentangled representation, ensemble, and meta-learning, all of which leverage high-level paradigms to extract domain-invariant information.

Disentangled representation learning: This approach aims to explicitly decompose learned representations into domain-shared and domain-specific components, enabling the model to isolate invariant information from domain-dependent variations (Kuang et al. 2025). By separating transferable knowledge from domain-specific factors, this paradigm effectively mitigates negative transfer caused by distribution discrepancies across domains, thereby improving the robustness of learned representations.

Chen et al. (2024) propose the Graph Domain Disentangled Network (GDDN), which addresses the significant challenge of inter-subject variability in EEG-based emotion recognition by decomposing graph connectivity patterns and corresponding representations into common and subject-specific attributes. Moreover, Jia et al. (2024) introduce the multi-level disentangling network called MDNet for multi-modal emotion recognition. MDNet employs a hierarchical feature disentanglement mechanism to simultaneously model modality-specific and subject-specific characteristics, wherein shared encoders are utilized to extract invariant representations that capture cross-modal and cross-subject coherence, while private encoders are designed to learn modality-dependent and subject-dependent features. This disentanglement process is implemented at both the modality level and the subject level, aiming to improve corss-modal feature interaction and cross-subject recognition capabilities.

Although feature disentanglement methods possess inherent advantages in addressing domain shift, the design process requires careful consideration to ensure the model captures pure disentangled features. This objective often necessitates the integration of the aforementioned adversarial-based or metric-based methods to further constrain the disentangled features, so that the resulting representations achieve enhanced cross-domain robustness.

Ensemble learning: Ensemble strategies enhance a model's generalization by aggregating outputs from a diverse set of networks. By employing a consensus-based voting mechanism, this approach effectively elevates predictive precision and overall system robustness.

For example, Pan et al. (2023) propose RAVE framework integrates an internal AdaBoost ensemble to strengthen individual Riemannian-based classifiers against non-linear session shifts. It then executes a majority voting protocol across 441 unique time-frequency windows to stabilize performance and enable cross-session transfer with minimal additional training. Li et al. (2023) expand their research into multi-source domain adaptation by treating each subject as an independent source domain within a parallel ensemble framework. During training, the framework employs a Random Matching Operation (RMO) to dynamically shuffle the pairings between domain-specific feature extractors and classifiers, which prevents over-reliance on individual data sources and minimizes overall classification variance. In the inference stage, a Top-h selection module is utilized to aggregate only the highest-ranked classifiers, achieving an optimal bias-variance trade-off and significantly enhancing the model's generalization across diverse subjects. In DG setting, Zhang et al. (2024) propose the EPNNE framework, which leverages evolutionary programming to automatically evolve a population of neural networks with diverse topologies to capture complex individual variability in cross-subject EEG signals. The method employs an associative reinforcement loss to collaboratively align the domain distributions of ensemble members and integrates their predictions through a collective voting mechanism, ensuring robust generalization of emotion recognition under heterogeneous data.

Meta-learning: The essence of meta-learning lies in its " learning to learn" capability, where a model distills fundamental learning strategies by observing the metadata and behavioral patterns of various ML techniques. Kuang et al. (2024) propose the EmoTracer system, which adopts a few-shot meta-learning strategy based on prototype networks, utilizing a task generation mechanism to partition multi-source physiological signals into support and query sets to simulate and learn emotional embedding representations

for new subjects. By establishing prototype centers for each emotion category within the feature space of a multi-source fusion network and employing Euclidean distance with Inverse Multiquadric weighting to measure sample-to-prototype similarity, the model achieves rapid cross-subject adaptation with minimal samples and no retraining. Similarity, Zhang et al. (2026) propose the FSL-DEM module, which utilizes a subject-based task generator to transform individuals from diverse source domains into independent meta-tasks for training, thereby acquiring well-initialized parameters for rapid adaptation to new subjects. The system further integrates a feature-based knowledge distillation strategy to compress the representational capacity derived from meta-learning into a lightweight student model. Through two-stage fine-tuning on limited target samples, the system achieves personalized, real-time monitoring of depressive episodes for new users.

Meta-learning significantly enhance model adaptability while ensuring computational economy by minimizing the reliance on extensive training data. Despite these advantages, the framework can suffer from heuristic gaps in its rule sets. Furthermore, certain implementations face representational bottlenecks, where the meta-features lack the capacity to encapsulate comprehensive informational depth.

Self-supervised learning: Self-supervised learning (SSL) leverages unlabeled data to capture intrinsic spatio-temporal dynamics by designing pretext tasks, which is particularly effective for enhancing cross-domain generalization in physiological sensing where high-quality labels are scarce (Duan et al. 2023). Based on the nature of pretext tasks, SSL strategies in this field can be categorized into predictive-based, generative-based, and contrastive-based methods.

Predictive-based SSL methods learn robust representations by predicting discrete pseudo-labels derived from signal transformations. For instance, Sarkar and Etemad (2022) apply six distinct transformations to ECG signals and train a network to recognize these modifications, enabling the extraction of spatiotemporal features that generalize well across different emotion categories and datasets. Similarly, Wu et al. (2023) propose a Transformer-based framework for wearable emotion recognition, utilizing signal transformation recognition as a pretext task to learn generalized multimodal representations from extensive unlabeled physiological data.

Generative-based SSL focuses on capturing contextual correlations through reconstruction tasks. Wang et al. (2022) introduce the Multi-modal Domain Adaptive Variational Autoencoder (MMDA-VAE), which employs a multi-modal VAE to project EEG and eye-movement data into a common latent space, learning shared cross-domain representations through reconstruction and cycle-consistency. Furthermore, Zhang et al. (2025) propose the MV-SSTMA-DA model, which utilizes a masked autoencoder architecture. By masking portions of the input EEG and reconstructing the masked segments, the model effectively extracts structured, topologically informed features suitable for cross-subject adaptation.

Contrastive-based SSL encourages similar data points to cluster together while pushing dissimilar ones apart in the representation space. Can et al. (2025) adapt contrastive architectures like TS-TCC for daily life affect recognition, leveraging unlabeled multimodal signals to learn meaningful representations that mitigate the noise and subjectivity of real-world labels. In the context of clinical diagnosis, Cai et al. (2025) propose a self-supervised contrastive domain generalization framework (SS-CDG). By contrasting EEG responses induced by different stimuli in healthy subjects, the model acquires unsupervised representations that can be successfully transferred to decode the musical neural activity of patients with disorders of consciousness.

By capturing the underlying distribution of physiological signals through these varied pretext tasks, SSL effectively bridges the gap between different domains, providing a powerful initialized state for downstream mental health assessment tasks even with limited labeled samples.

6.2 Model interpretability

The rapid advancement of deep learning has significantly improved the performance of intelligent mental health assessment systems. However, the black-box nature of many DL models, caused by their high dimensionality and non-linearity, limits the transparency of the decision-making process and has consequently driven increasing interest in eXplainable AI (XAI) research.

In physiological signal-based mental health assessment, interpretability is particularly critical due to the clinical and subjective nature of mental state analysis. Unlike conventional vision- or language-based AI tasks, where decision evidence can often be intuitively verified through perceptible visual or textual patterns, physiological signals are inherently abstract and non-intuitive. Variations in heart rate, skin conductance, or brainwave activity cannot be directly understood without specialized physiological knowledge, making it difficult for clinicians and users to interpret the rationale behind model predictions. Moreover, mental states such as stress, anxiety, and depression often exhibit subtle and overlapping physiological responses, while physiological signals themselves are highly dynamic and continuously changing over time. These characteristics substantially increase the difficulty of establishing transparent and trustworthy decision mechanisms for real-world mental health applications.

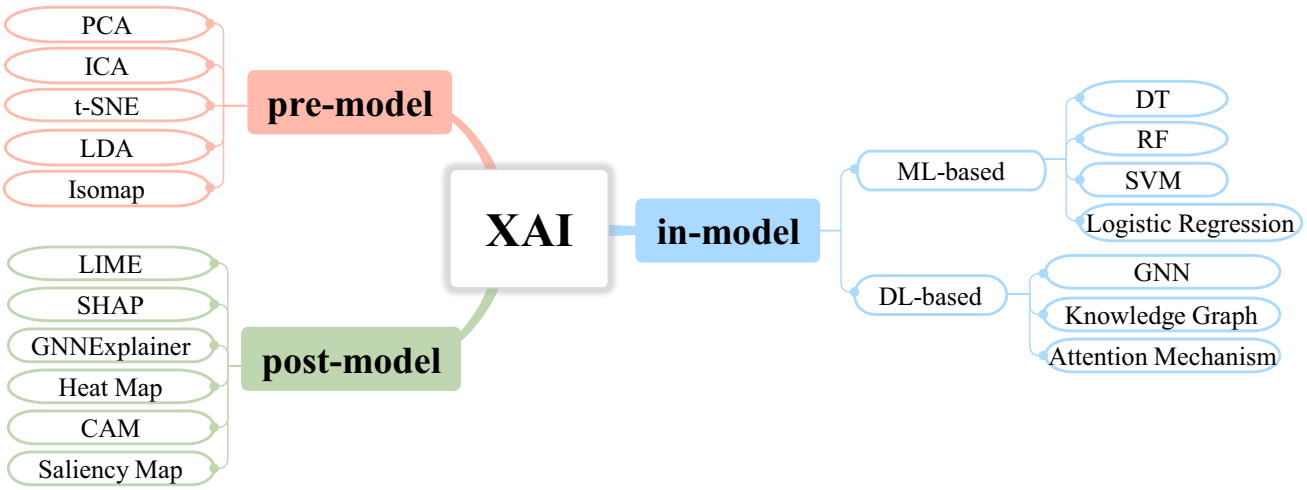
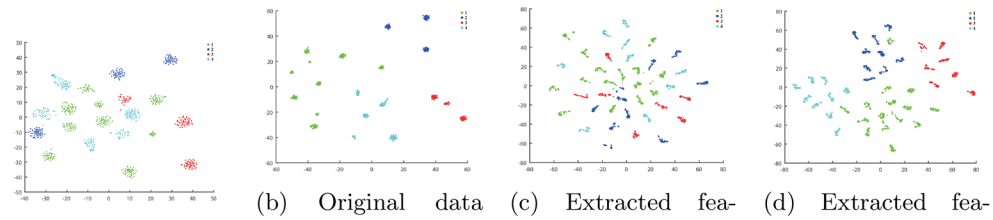


Fig. 7 The overview of XAI



(a) Original data distribution in DEAP. (b) Original data distribution in MAHNOB-HCI. (c) Extracted feature distribution in DAEP. (d) Extracted feature distribution in MAHNOB-HCI.

Fig. 8 t-SNE Visualizations of feature space in different datasets (Li et al. 2024)

Inspired by the previous work of Cortiñas-Lorenzo and Lacey (2024), we categorize explainability techniques in mental health assessment into three stages: pre-model, in-model, and post-model, and summarize recent advances for each, as illustrated in Fig. 7.

6.2.1 Pre-model phase

In the pre-model phase, exploratory data analysis is employed to reveal the intrinsic properties of raw signals before model training. Some specialized physiological signals such as EEG and ECG, characterized by high temporal resolution and complex dynamics, result in high-dimensional datasets. Consequently, dimensionality reduction techniques are applied to project the data into lower-dimensional spaces. These techniques not only enable intuitive visualization of data distribution, but also uncover underlying structure of emotional states.

Dimensionality Reduction Techniques: Common dimensionality reduction methods, including Principal Component Analysis (PCA), Independent Component Analysis (ICA), Isometric Mapping (Isomap), t-Distributed Stochastic Neighbor Embedding (t-SNE), and Linear Discriminant Analysis (LDA), have been widely adopted in mental health assessment to enhance data visualization and interpretability. Specifically, PCA (Zhang et al. 2018) transforms correlated features into orthogonal components, facilitating variance-focused analysis; ICA (Valenza et al. 2011) separates independent sources within mixed signals, mitigating the interference from multiple sources; Isomap and t-SNE (Bodaghi et al. 2024) preserve global and local structures based on similarity, capturing the intrinsic geometry of the original signal space; and LDA (Kim and André 2008) maximizes class separability, crucial for distinguishing closely related affective states. These methods are suited to physiological signals due to their ability to uncover key latent patterns obscured by noise and redundancy, enable more transparent understanding. In addition, visual representations derived from these methods offer intuitive insights for evaluating model performance before and after training. For instance, Li et al. (2024) employ t-SNE to visualize feature space from both both original brain network and proposed MESNP model on the MAHNOB-HCI dataset and DEAP dataset. As shown in Fig. 8, the improved clustering of emotional states after feature extraction demonstrates the model's ability to learn discriminative representations, reinforcing its interpretability.

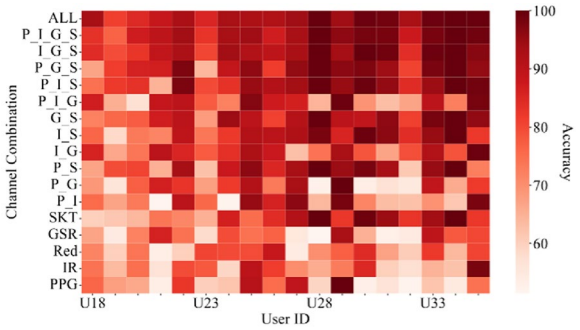


Fig. 9 Model and feature analysis with Heat Map (Zhang et al. 2026)

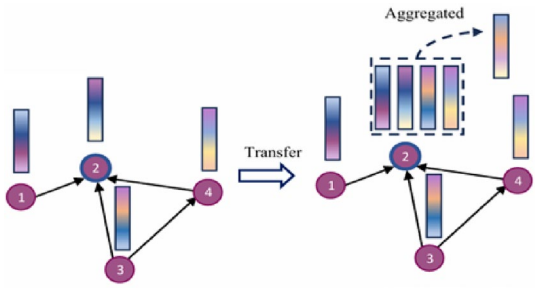


Fig. 10 Feature analysis with GNNExplainer (Shen et al. 2023)

6.2.2 In-model phase

In-model interpretability is achieved by embedding interpretive mechanisms directly into the model architecture, enabling the model to produce inherently explainable outputs throughout training and inference. These techniques participate actively in the learning process, allowing the model to not only make predictions but also reveal the underlying relationships between input features and outputs.

ML-based Method: Previous studies primarily rely on fundamental machine learning methods, such as DT, RF, SVM and logistic regression (Lin et al. 2019; Marey et al. 2024; Giannakakis et al. 2019), reveal the relationships between physiological signals and mental health diagnoses through explicit feature weights or rule-based structures. For example, Lin et al. (2019) apply RF to assess the contribution of each physiological channel in emotion recognition, enabling intuitive interpretability at the sensor level.

Knowledge Graph: Knowledge graphs encode structured domain knowledge through entities and relations, enabling semantic reasoning that links low-level features to high-level emotional or clinical concepts. EmotionO+ (Su et al. 2014) constructs an ontology-based graph from EEG, fNIRS, and contextual data, supporting concept-level reasoning and transparent decision tracing. Knowledge-Embedded Temporal Convolutional Transformer model (KETCT) (Zheng et al. 2023) embed medical and affective knowledge into the model's latent space, aligning signal representations with psychological constructs for domain-informed outputs. Obviously, integrating structured knowledge into the model facilitates semantically grounded and explainable inference.

Attention Mechanism: The attention mechanism enables the model to focus on the most significant parts of the input data, allowing it to identify which features of physiological signals are crucial for mental health diagnosis (Yang et al. 2021; Guo et al. 2023; Li et al. 2024). Specifically, Guo et al. (2023) compare model performance before and after introducing multi-head attention module, showing that attention significantly increases the spatial separability between emotional categories in the embedding space by adaptively weighting informative features across modalities. Moreover, Li et al. (2024) analyze temporal frame attention and observe consistent activation patterns across trials for the same time intervals, while highlighting variation across different time intervals, demonstrating the module's ability to localize task-relevant temporal segments. Although attention-based outputs may not always fully align with established physiological interpretations due to the data-driven nature of model learning, they nonetheless offer insights into which implicit patterns and modality preferences that the model prioritizes during prediction. Visualization results are typically represented as Heat Map or Grad-CAM, as shown in Fig. 9 and Fig. 14.

Graph Neural Network: GNN enable interpretable learning by leveraging their internal message-passing mechanisms to capture both node-level features and topological structures within the graph (Ye et al. 2025; Shen et al. 2023). Ye et al. (2025) employ degree centrality metrics on the dynamic adjacency matrix learned by the HD-GCN model, demonstrating that connection strength between EEG channels increases with frequency, highlighting the model's capacity to localize spatially relevant brain regions. Shen et al. (2023) employ the biased weighting parameter within the GCN to assess the associations among brain channel features under depressive stimuli. Similarly, Yu et al. (2023) analysis the most discriminate brain region and functional sub-networks based on attention coefficient maps by utilizing standard t-test, while Song et al. (2020) apply node-level centrality measures to interpret each channel's contribution to emotion classification. Collectively, these methods underscore the capacity of GNNs to deliver built-in spatial interpretability by structurally capturing interdependency and characteristic among physiological signal channels during model learning. Visualization results are typically generated using GNN interpretability toolkits, as shown in Fig. 10.

6.2.3 Post-model phase

The goal of post-model XAI techniques is to understand how model outputs relate to inputs after model training.

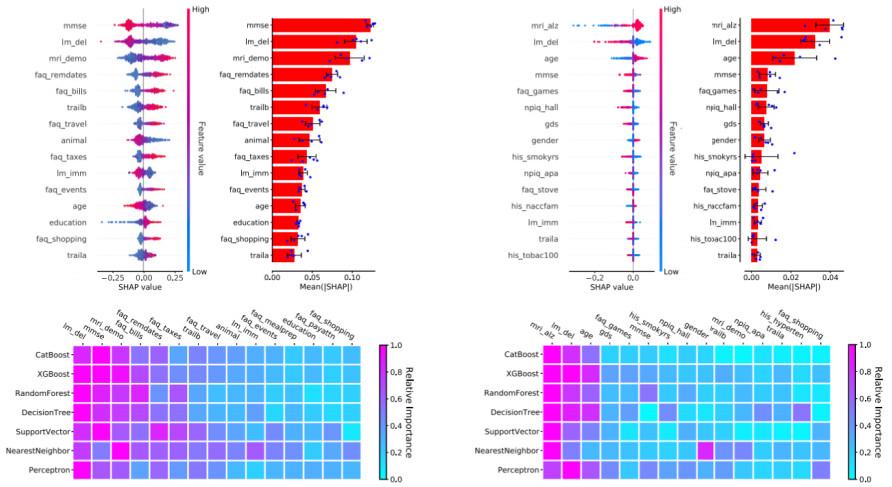


Fig. 11 Feature importance analysis with SHAP (Qiu et al. 2022)

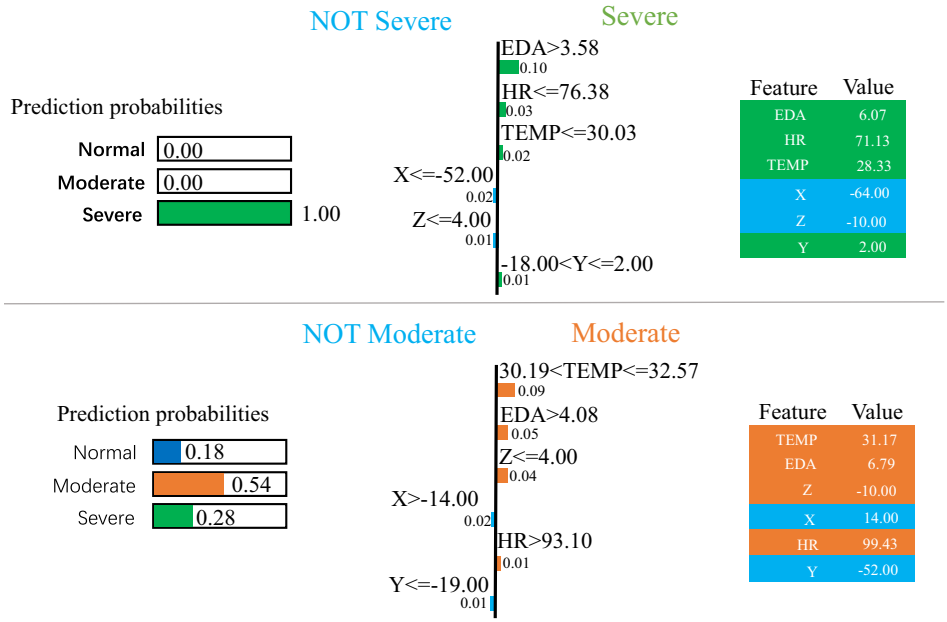


Fig. 12 Feature importance analysis with LIME (Boulanouar et al. 2024)

This subsection introduces commonly used auxiliary interpretation tools that analyze the correlation between model predictions and inputs at either the model or instance level. It is important to note that some of the post-model methods are also applicable to the pre-model phase, such as the t-SNE method described earlier.

SHAP: It (Lundberg and Lee 2017) is a model-agnostic interpretability method based on Shapley values from cooperative game theory. It enables instance-level explanations by quantifying how each physiological feature contributes to emotional predictions across all possible feature subsets. Specifically, it identifies key features driving predictions, reveals features with positive or negative contributions for class separation, and uncovers consistently important features across different model architectures, as illustrated in Fig. 11. This is particularly valuable for interpreting complex multi-modal models, highlighting dominant signals, and assessing model robustness across individuals.

LIME: It (Ribeiro et al. 2016) is a local interpretability method that explains individual predictions by approximating the complex model with a simple surrogate model in the neighborhood of a given instance. As shown in Fig. 12, LIME is used to interpret predictions from classifier. For a severe stress case, features like HR, EDA, and TEMP positively influenced the prediction. In a moderate stress case, a different feature set including TEMP, EDA, and Z-axis movement exhibit positive connections. These examples demonstrate LIME's ability to reveal localized feature importance patterns across emotional states, which is crucial for understanding individual-specific model behavior.

Saliency Map: This method (Simonyan et al. 2014) estimates feature region importance by perturbing inputs and

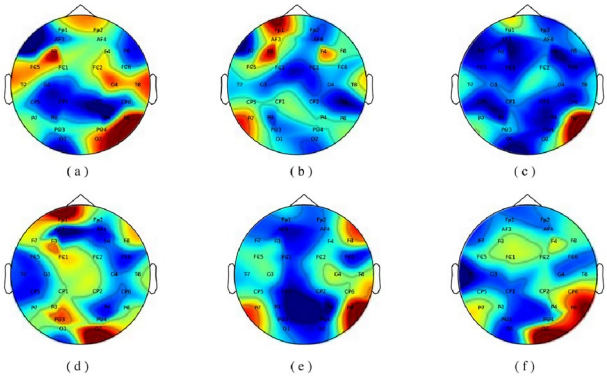


Fig. 13 Region importance analysis with Saliency Map (Ding et al. 2022)

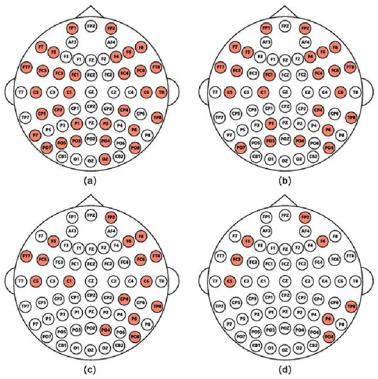


Fig. 14 Channel importance analysis with Grad-CAM (Xu et al. 2023)

observing the resulting changes in prediction probability. It offers coarse-grained visualization of which neural signal regions drive emotion-related decisions. For example, as shown in Fig. 13, Ding et al. (2022) apply Saliency Maps on the DEAP dataset to identify that arousal-related responses are predominantly linked to the frontal, temporal, and parietal lobes.

Grad-CAM: This method (Selvaraju et al. 2016) generates visual explanations by computing the gradients of target outputs with respect to the final convolutional layer, producing spatial activation maps that highlight regions most influential to the model's decision. Grad-CAM serves as an effective tool for intuitively localizing critical neural signal channels in mental health assessment. For instance, Grad-CAM has been used to rank 62 EEG channels on the SEED dataset, identifying top contributors such as P6, C5, TP8, and F6, as shown in Fig. 14, which are critical for emotion arousal.

Heat Map: This method (Gu 2022) visualizes high-dimensional features or model outputs as color gradients in a two-dimensional space, highlighting the model attention, such as feature importance, activation levels, or classification weights. As shown in Fig. 9, the Heat Map illustrates the performance of various physiological channel combinations in detecting depressive disorders across users. The PPG + SKT combination demonstrates the strongest complementarity, and supports the effectiveness of multi-channel fusion strategy. The heatmap also highlights inter-individual variability in model responses, which is an important consideration in depression recognition. Additionally, Zhu et al. (2022) utilize Heat Map to visualize the performance of EEG frequency bands to depression diagnosis models, not only clarifying the contribution of specific bands but also showcasing the superiority of proposed models over traditional baselines.

GNNExplainer: As a graph-wise method, GNNExplainer (Ying et al. 2019) can interpret the predictions of GNN by identifying crucial subgraph structures and node features, without requiring architectural changes or retraining. It formulates explanation as an optimization problem that maximizes the mutual information between a model's prediction and the distribution of potential explanatory subgraphs. For instance, Shen et al. (2023) apply GNNExplainer to EEG-based depression detection, as shown in Fig. 10, assigning importance scores to nodes to reveal key channels and spatial relevance.

6.3 Multi-modal fusion

The ambiguous and overlapping nature of mental states makes it difficult for a single modality to capture subtle affective variations, particularly under noisy or incomplete conditions. This limitation highlights the importance of multi-modal fusion in physiological signal-based mental health assessment.

However, multi-modal fusion in physiological computing presents challenges that differ substantially from those in conventional vision-language or audio-visual tasks. Unlike visual or textual modalities with relatively explicit semantic structures, physiological signals are generated from heterogeneous biological systems and exhibit distinct temporal dynamics, response delays, and physiological meanings. For example, EEG reflects neural activity, ECG captures cardiac dynamics, while GSR represents autonomic nervous system responses. These modalities are often asynchronous, subject to sensor-specific acquisition conditions, and characterized by varying sampling frequencies and signal distributions, which significantly complicate cross-modal representation learning and alignment.

Moreover, physiological signals collected in real-world wearable scenarios are highly susceptible to motion artifacts, sensor failures, and environmental noise, making incomplete or corrupted modalities a common practical issue. Against this backdrop, existing multi-modal fusion strategies can generally be categorized into data-level, feature-level, decision-level, and hybrid approaches, as

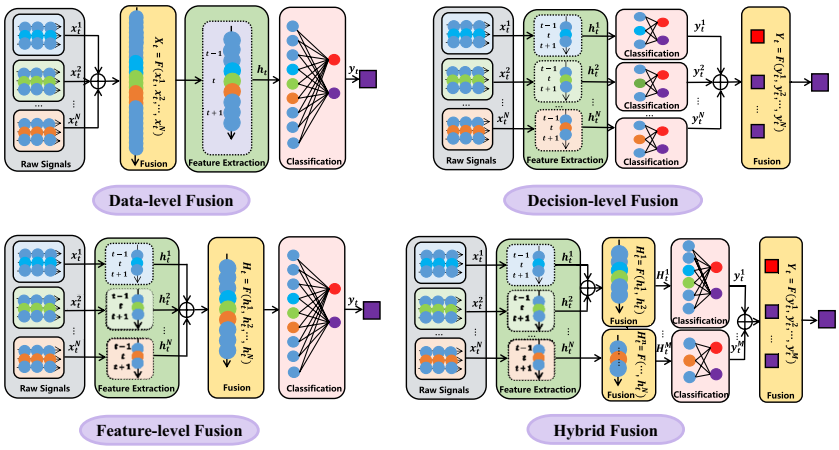


Fig. 15 Type of fusion level

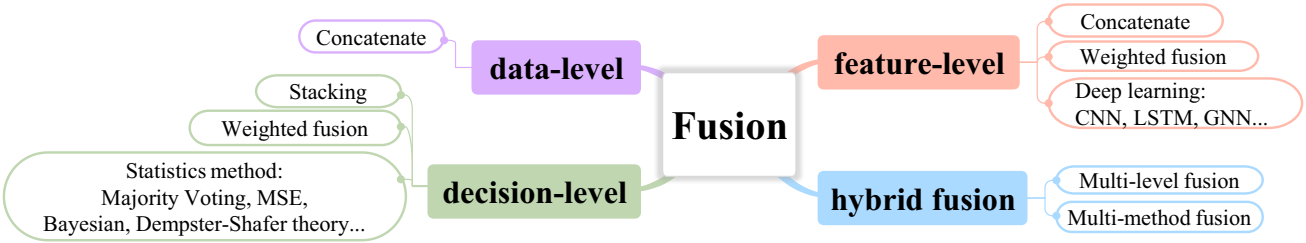


Fig. 16 The overview of multi-modal fusion

illustrated in Fig. 15. In the following sections, we systematically analyze these fusion paradigms and their corresponding techniques for addressing the aforementioned challenges, as shown in Fig. 16.

6.3.1 Data-level fusion

Data-level fusion operates at the earliest stage of multimodal fusion by directly concatenating raw data streams without feature transformation. It aims to preserve the full fidelity of each modality, but must address challenges such as temporal misalignment and data incompleteness. For instance, Kanjo et al. (2018) fuse physiological, environmental, and self-reported data through timestamp alignment and interpolation, to construct a unified dataset for emotion prediction. While this method maintains data integrity, it often results in data redundancy, high computational cost, and overlooks complex cross-modal dependencies. Consequently, data-level fusion is less prevalent in recent physiological signal-based mental health assessment studies, which increasingly favor more flexible fusion strategies.

6.3.2 Feature-level fusion

Feature-level fusion is performed after data preprocessing and feature extraction, aiming to integrate feature representations from heterogeneous modalities into a unified embedding space. The most straightforward strategy is concatenation (Zhu et al. 2022; Cheng et al. 2023), which joins aligned features along a specific dimension. Suppose EEG and GSR modalities produce feature representations of shapes (T, d_1) and (T, d_2), where T denotes the shared temporal dimension and d_1, d_2 represent the modality-specific feature dimensions. Concatenate fusion makes them into a unified representation of shape $(T, d_1 + d_2)$, avoiding explicit time or space alignment at the data level but failing to account for inter-modal relevance or intra-modal importance. To address this, recent works introduce adaptive weighting mechanisms to dynamically modulate modality contributions. For instance, Fu et al. (2024) and Liu et al. (2023) reshape concatenated EEG features into serval partitions and apply Transformer to re-weight each partition. This design allows the fusion process to reflect temporal or spatial shifts in affective salience across modalities, effectively capturing the fluctuating importance of physiological signals under varying emotional contexts. Such adaptive mechanisms are particularly valuable for mental health assessment, where the contribution of each modality often varies with emotional intensity, stress level, or mental disorder-specific traits (Guo et al. 2023; Wu et al. 2023; Ye et al. 2025).

Beyond weighted fusion, DL-based fusion provide a more expressive framework for capturing the nonlinear, hierarchical relationships between modalities. CNN-based fusion effectively captures local and global spatial dependencies in physiological signals. The spatial pattern is particularly

advantageous for multi-electrode signals like EEG (Li et al. 2023; Zhang et al. 2022), where emotion-related activations often exhibit regional specificity. For example, TSception (Ding et al. 2022) employs parallel multi-scale 1D convolutions to extract global EEG patterns and hemisphere-specific features, facilitating emotional representation related to different brain spatial regions. LSTM-based fusion incorporates temporal memory to capture long-range dependencies and intra-modal dynamics. Kuang et al. (2024) introduce an LSTM Multi-source Fusion Network (LMFN) that integrates peripheral physiological signals by selectively retaining historical patterns through gating mechanisms. This design is well-suited to modeling affective transitions reflected in gradual physiological responses, such as GSR and HRV activities. GNN-based fusion introduces structural priors by modeling inter-channel relationships through graph topologies. In EEG-based affective tasks, each electrode is treated as a graph node, with edges reflecting spatial or functional connectivity. Studies such as Chen et al. (2022); Pan et al. (2024) leverage such topologies to learn emotional patterns embedded in the spatial organization of neural signals, enabling more structure-aware fusion. Feature-level fusion can streamline data through feature selection, which not only retains essential information but also significantly reduces computational complexity. Obviously, careful design of the feature selection strategy is critical to prevent the loss of crucial information and to ensure effective feature fusion from different modalities.

6.3.3 Decision-level fusion

Decision-level fusion emphasizes the independent classification of different modalities, followed by the combination of these classifications to derive a final decision. The typical approach is statistical-based fusion, such as majority voting mechanism (Hu et al. 2024), which mitigates the impact of noisy or less reliable modalities by relying on consensus. This is especially effective in depression recognition where individual signals may respond differently to affective stimuli. Beyond simple voting, Zhao et al. (2021) apply Bayesian optimal cue integration to weigh classification results based on class separability and probability confidence. This probabilistic modeling allows the system to dynamically select the most reliable modalities under emotional ambiguity. In addition, Liu et al. (2017) propose a Weighted Fuzzy Dempster-Shafer fusion, applying intuitionistic belief assignment to address ambiguous affective information across modalities. Stacking-based fusion (Rahman et al. 2022) further advances decision-level strategies by training a meta-learner to combine outputs from heterogeneous base models. In affective tasks, this allows leveraging complementary temporal or spectral insights captured by different classifiers, thereby improving generalization across emotional states. Weighted decision fusion adds granularity by assigning adaptive weights to each prediction based on modality-specific confidence. Guo et al. (2022) train separate models on time-domain and frequency-domain features, while Li et al. (2020) predict final emotion with variable weight based on the probability of emotional output of modalities. Overall, decision-level fusion emphasizes synergy among classifiers, effectively mitigating modality inconsistencies and enhancing system performance.

6.3.4 Hybrid fusion

Hybrid fusion integrates multiple fusion strategies across levels or methods, to exploit their complementary strengths. For example, Zheng et al. (2023) address the temporal and semantic complexity of emotion-related signals by jointly employing TCN and knowledge graph networks on a unified multi-modal input. A task-driven attention mechanism and gating module are introduced to adaptively fuse features, effectively capturing both local temporal patterns and global contextual relationships. Zhu et al. (2022) tackle the subtlety of mild depressive states by combining EEG and Eye Movement features through a Deep Belief Network, followed by a meta-classifier. This layered architecture enables hierarchical learning of discriminative emotional cues while mitigating overfitting. Fu et al. (2024) enhance fine-grained affective discrimination by segmenting and reweighting eye movement features via concatenate-and-weighted fusion, improving sensitivity to subtle emotional fluctuations. Hybrid fusion provides a flexible solution for integrating complex and heterogeneous modalities, outperforming single-level or single-method approaches. By combining diverse fusion strategies, it enhances model robustness and accuracy, making it more suitable for different affective applications.

7 Discussion and future direction

7.1 Domain shift: overcoming domain boundaries

Cross-domain issues in mental health diagnostics pose substantial challenges due to distribution shifts across datasets, subjects, and physiological modalities, which impair model generalization to unseen domains. To mitigate this, numerous approaches address try to address the challenge. A deeper analysis indicates that the core objective of these approaches remains the capture or isolation of domain-invariant features. These frameworks essentially treat domain-specific traits as negative transfer factors that hinder generalization. Some research in computer vision explores how these

heterogeneous features contribute to target domain prediction (Yao and Doretto 2010). In the context of mental health assessment, we hypothesize that domain-specific features contain rich physiological dynamics rather than mere noise, yet this perspective remains largely under-explored in the field of mental health assessment. Current research typically focuses on isolating domain-invariant traits while overlooking the potential utility of domain-specific information. Consequently, the development of advanced frameworks such as disentangled representation models to leverage these specific features for cross-domain tasks represents a highly promising but untapped research opportunity.

7.2 XAI: interpreting modals for trust and transparency

Current research in mental health assessment rarely incorporates interpretability at the model design stage. Instead, most studies rely on external tools either before or after model training, including visualizing data distributions to understand emotional signal properties or analyzing correlations between predictions and inputs to align results with known psychological theories, respectively. These methods remain coarse and offer limited insight, as they typically provide only feature importance without revealing underlying decision logic. Emerging post-hoc techniques for GNNs can provide richer interpretability and can be applied to intricate sentiment analysis. For instance, SubgraphX (Yuan et al. 2021) identifies high-order predictive subgraphs via Shapley values, GraphLIME (Huang et al. 2022) offers sparse feature-level explanations, and XGNN (Yuan et al. 2020) synthesizes prototypical graphs to reveal global decision logic. These tools capture structural interactions, nonlinear dependencies, and model-wide patterns, offering deeper insights for complex affective data. However, such methods still operate externally. Some interpretable architectures in vision, such as NBDT and CLIP (Wan et al. 2020; Radford et al. 2021), integrate symbolic reasoning and semantic alignment directly into the model, offering a promising design paradigm for mental health assessment. Meanwhile, Large Language Models (LLMs) show potential in generating semantically meaningful explanations and have been applied to mental state inference from media platform (Xu et al. 2024; Hu et al. 2026). Extending their reasoning to physiological signals may offer transparent multi-modal interpretations for users and clinicians.

7.3 Multi-modal fusion: fusing diverse modalities for robustness

Multi-modal fusion has become a cornerstone in physiological signal-based mental health monitoring, enabling the integration of heterogeneous physiological modalities to derive complementary feature representations. While substantial advancements have been made, some obstacles remain. For instance, missing data and incomplete modalities constitute a pervasive and critical challenge, particularly in the field of ubiquitous computing. Real-world applications frequently encounter issues such as sensor detachment, hardware malfunctions, and transmission losses. These unpredictable factors severely degrade the performance of traditional fusion models that rigidly rely on complete multi-modal inputs. Therefore, shifting the research focus toward robust multi-modal fusion under severe modality-missing scenarios holds paramount importance for practical deployment. To address this problem, future research can explore cross-modal generative frameworks to dynamically reconstruct missing physiological signals by integrating the latent correlations among available modalities or by leveraging advanced data generation methods such as GAN and Diffusion (Yi et al. 2024). Additionally, developing modality-agnostic network architectures through techniques like knowledge distillation enables models to maintain resilient affective representations (Pan et al. 2024). These strategies ensure reliable mental health monitoring even when systems operate under dynamic and unpredictable sensor configurations.

Recently, the emergence of LLMs and foundation models has introduced a promising new direction for multimodal physiological computing. Unlike conventional fusion frameworks that mainly rely on feature concatenation or attention-based alignment, LLM-driven architectures enable semantic-level integration across heterogeneous physiological modalities and language representations. Recent studies have explored the use of multimodal LLMs for EEG-based mental health assessment and physiological signal understanding, demonstrating the potential of foundation models in cross-modal reasoning, signal reconstruction, and multimodal affective analysis (Hu et al. 2024; Thapa et al. 2025; Shen et al. 2026). In particular, LLM-based generative frameworks may provide a unified solution for missing modality reconstruction, semantic alignment, and dynamic multimodal interaction by leveraging large-scale prior knowledge and contextual modeling capabilities (Hu et al. 2025). Therefore, integrating foundation models with physiological signal fusion is expected to become an important future research direction for robust and intelligent mental health assessment systems.

8 Conclusion

AI holds transformative potential in physiological signal-based mental health assessment, but its interdisciplinary nature presents both theoretical and practical challenges. In this review, we provide an extensive review of the

application framework for physiological computing and AI in mental health assessment, focusing on three core affective contexts: emotion recognition, stress response, and mental disorder diagnosis. We discuss the representation of related features for multi-modal physiological signals, incorporating both specialized and portable signals, and introduce key public datasets relevant to these contexts. Meanwhile, we also examine the application of fundamental AI models, encompassing ML and DL techniques. Building on this foundation, the paper addresses current challenges, such as domain shift, model interpretability, and multi-modal fusion, offering corresponding solutions. It is worth noting that the broader field also faces other critical challenges, such as AI ethics and data privacy and security. Although these issues are equally significant for clinical translation and real-world deployment, a detailed discussion of them falls outside the scope of this review. Moving forward, future research should prioritize overcoming the generalization bottleneck caused by domain dependency, advancing interpretable decision-making systems, and exploring dynamic and effective mechanisms for coupling multi-modal data. Ultimately, this review provides valuable theoretical insights and methodological guidance, laying the groundwork for the development of reliable, robust, and deployable AI-driven systems for mental health assessment.

Acknowledgements On behalf of all authors, the corresponding author states that there is no conflict of interest.

Funding This work was supported in part by China NSFC (62172286), in part by key program of DEGP (2024ZDZX1011), in part by Shenzhen Science and Technology Program (JCYJ20250604181429038), in part by Shenzhen Science and Technology Foundation (ZDSYS20190902092853047), in part by Guangdong Provincial Key Lab of Integrated Communication, Sensing and Computation for Ubiquitous Internet of Things (No.2023B1212010007).

References

Organization, W.H., et al.: Depressive disorder (depression) world health organization. Switzerland, Geneva (2023)

Khansari, D.N., Murgo, A.J., Faith, R.E.: Effects of stress on the immune system. Immunol. Today **11**, 170–175 (1990)

O'leary, A.: Stress, emotion, and human immune function. Psychol. Bull. **108**(3), 363 (1990)

Maddock, C., Pariante, C.M.: How does stress affect you? An overview of stress, immunity, depression and disease. Epidemiol. Psychiatr. Sci. **10**(3), 153–162 (2001)

Hammen, C.: Stress and depression. Annu. Rev. Clin. Psychol. **1**(1), 293–319 (2005)

Ekman, P., Levenson, R.W., Friesen, W.V.: Autonomic nervous system activity distinguishes among emotions. Science **221**(4616), 1208–1210 (1983)

Taylor, S.E.: Asymmetrical effects of positive and negative events: the mobilization-minimization hypothesis. Psychol. Bull. **110**(1), 67 (1991)

Jansen, A.S., Van Nguyen, X., Karpitskiy, V., Mettenleiter, T.C., Loewy, A.D.: Central command neurons of the sympathetic nervous system: basis of the fight-or-flight response. Science **270**(5236), 644–646 (1995)

Doust, J.W.L., Schneider, R.A.: Studies on the physiology of awareness: the effect of rhythmic sensory bombardment on emotions, blood oxygen saturation and the levels of consciousness. J. Meteorol. Soc. Jpn. **98**(413), 640–653 (1952)

Yang, L., Amin, O., Shihada, B.: Intelligent wearable systems: opportunities and challenges in health and sports. ACM CSUR **56**(7), 1–42 (2024)

Viswan, V., Shaffi, N., Mahmud, M., Subramanian, K., Hajamohideen, F.: Explainable artificial intelligence in Alzheimer's disease classification: a systematic review. Cogn. Comput. **16**(1), 1–44 (2024)

García-Ponsoda, S., García-Carrasco, J., Teruel, M.A., Maté, A., Trujillo, J.: Feature engineering of EEG applied to mental disorders: a systematic mapping study. Appl. Intell. **53**(20), 23203–23243 (2023)

Ahmed, T., Qassem, M., Kyriacou, P.A.: Physiological monitoring of stress and major depression: a review of the current monitoring techniques and considerations for the future. BSPC **75**, 103591 (2022)

Pourmohammadi, S., Maleki, A.: Stress detection using ECG and EMG signals: a comprehensive study. Comput. Methods Programs Biomed. **193**, 105482 (2020)

Badr, Y., Tariq, U., Al-Shargie, F., Babiloni, F., Al Mughairbi, F., Al-Nashash, H.: A review on evaluating mental stress by deep learning using eeg signals. Neural Comput. Appl. , 1–26 (2024)

Yasin, S., Hussain, S.A., Aslan, S., Raza, I., Muzammel, M., Othmani, A.: EEG based major depressive disorder and bipolar disorder detection using neural networks: a review. Comput. Methods Programs Biomed. **202**, 106007 (2021)

Garg, D., Verma, G.K., Singh, A.K.: A review of deep learning based methods for affect analysis using physiological signals. Multimed. Tools Appl. **82**(17), 26089–26134 (2023)

Li, X., Zhang, Y., Tiwari, P., Song, D., Hu, B., Yang, M., Zhao, Z., Kumar, N., Marttinen, P.: EEG based emotion recognition: a tutorial and review. ACM CSUR **55**(4), 1–57 (2022)

Can, Y.S., Mahesh, B., André, E.: Approaches, applications, and challenges in physiological emotion recognition—a tutorial overview. Proc. IEEE **111**(10), 1287–1313 (2023)

Saganowski, S., Perz, B., Polak, A.G., Kazienko, P.: Emotion recognition for everyday life using physiological signals from wearables: a systematic literature review. IEEE TAFFC **14**(3), 1876–1897 (2022)

Charles, R.L., Nixon, J.: Measuring mental workload using physiological measures: a systematic review. Appl. Ergon. **74**, 221–232 (2019)

Hegde, K., Jayalath, H.: Emotions in the loop: A survey of affective computing for emotional support, (2025). arXiv:2505.01542 arXiv preprint

Ekman, P.: etal Basic emotions. Handbook of cognition and emotion **98**(45–60), 16 (1999)

Plutchik, R.: Emotions and Life: Perspectives from Psychology. Biology, and Evolution , (2003)

Russell, J.A.: A circumplex model of affect. J. Pers. Soc. Psychol. **39**(6), 1161 (1980)

Mehrabian, A., Russell, J.A.: An Approach to Environmental Psychology , (1974)

Yeo, G.C., Ong, D.C.: Associations between cognitive appraisals and emotions: a meta-analytic review. Psychol. Bull. , (2024)

Clore, G.L., Ortony, A.: Cognition in emotion: Always, sometimes, or never. Cognitive neuroscience of emotion , 24–61 (2000)

Sander, D., Grandjean, D., Scherer, K.R.: A systems approach to appraisal mechanisms in emotion. Neural Netw. **18**(4), 317–352 (2005)

Chrousos, G.P., Gold, P.W.: The concepts of stress and stress system disorders: overview of physical and behavioral homeostasis. JAMA **267**(9), 1244–1252 (1992)

American Psychiatric Association, D., Association, A.P.: etal Diagnostic and Statistical Manual of Mental Disorders: DSM-5, vol. 5, (2013)

Ding, Y., Robinson, N., Zhang, S., Zeng, Q., Guan, C.: TSception: capturing temporal dynamics and spatial asymmetry from EEG for emotion recognition. IEEE TAFFC **14**(3), 2238–2250 (2022)

Guo, Y., Zhang, T., Huang, W.: Emotion recognition based on physiological signals multi-head attention contrastive learning. In: IEEE BIBM, pp. 1929–1934. IEEE (2023)

Liu, Y., Lan, Z., Cui, J., Sourina, O., Müller-Wittig, W.: Inter-subject transfer learning for EEG-based mental fatigue recognition. Adv. Eng. Inform. **46**, 101157 (2020)

Fang, R., Zhang, R., Hosseini, E., Parenteau, A.M., Hang, S., Rafatirad, S., Hostinar, C.E., Orooji, M., Homayoun, H.: Prevent overfitting and redundancy in physiological signal analyses for stress detection. In: IEEE BIBM, pp. 2585–2588. IEEE (2022)

Barros, C., Silva, C.A., Pinheiro, A.P.: Advanced EEG-based learning approaches to predict schizophrenia: promises and pitfalls. Artif. Intell. Med. **114**, 102039 (2021)

Ma, M., Lu, B., Gong, Y., Xiao, C., Yang, Y., Ju, Y., Xi, Z., Gao, Y., Ning, X., Zhang, Y.: EEG microstate analysis and machine learning classification in patients with obsessive-compulsive disorder. J. Psychiatr. Res. **182**, 186–194 (2025)

Kashou, A.H., Basit, H., Chhabra, L.: Physiology, sinoatrial node (2017)

Thiagarajan, V.: Bio-signal based detection of hyperarousal as onset of post-traumatic stress disorder (PTSD) episodes. In: International Conference on e-Health and Bioengineering, pp. 147–153. Springer (2023)

Baygin, M., Barua, P.D., Dogan, S., Tuncer, T., Hong, T.J., March, S., Tan, R.-S., Molinari, F., Acharya, U.R.: Automated anxiety detection using probabilistic binary pattern with ECG signals. Comput. Methods Programs Biomed. **247**, 108076 (2024)

Geng, D., An, Q., Fu, Z., Wang, C., An, H.: Identification of major depression patients using machine learning models based on heart rate variability during sleep stages for pre-hospital screening. Comput. Biol. Med. **162**, 107060 (2023)

Khare, S.K., March, S., Barua, P.D., Gadre, V.M., Acharya, U.R.: Application of data fusion for automated detection of children with developmental and mental disorders: a systematic review of the last decade. Information Fusion , 101898 (2023)

Polcwiartek, C., Atwater, B.D., Kragholm, K., Friedman, D.J., Barcella, C.A., Attar, R., Graff, C., Nielsen, J.B., Pietersen, A., Søgaard, P.: Association between ECG abnormalities and fatal cardiovascular disease among patients with and without severe mental illness. J. Am. Heart Assoc. **10**(2), 019416 (2021)

Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., Patras, I.: DEAP: a database for emotion analysis; using physiological signals. IEEE TAFFC **3**(1), 18–31 (2011)

Sharma, K., Castellini, C., Broek, E.L., Albu-Schaeffer, A., Schwenker, F.: A dataset of continuous affect annotations and physiological signals for emotion analysis. Sci. Data **6**(1), 196 (2019)

Schmidt, P., Reiss, A., Duerichen, R., Marberger, C., Van Laerhoven, K.: Introducing wesad, a multimodal dataset for wearable stress and affect detection. In: Proceedings of the 20th ACM International Conference on Multimodal Interaction, pp. 400–408. (2018)

Chan, S.J., Thiel, F., Kaimal, A.J., Pitman, R.K., Orr, S.P., Dekel, S.: Validation of childbirth-related posttraumatic stress disorder using psychophysiological assessment. Am. J. Obstet. Gynecol. **227**(4), 656–659 (2022)

Kose, M.R., Ahirwal, M.K., Kumar, A.: A new approach for emotions recognition through EOG and EMG signals. SIViP **15**(8), 1863–1871 (2021)

Cai, H., Liu, X., Jiang, A., Ni, R., Zhou, X., Cangelosi, A.: Combination of EOG and EEG for emotion recognition over different window sizes. In: 2021 IEEE 2nd International Conference on Human-Machine Systems, pp. 1–6. IEEE (2021)

Wu, J.-Y., Ching, C.T.-S., Wang, H.-M.D., Liao, L.-D.: Emerging wearable biosensor technologies for stress monitoring and their real-world applications. Biosensors **12**(12), 1097 (2022)

Zhu, J., Wang, Z., Zeng, S., Li, X., Hu, B., Zhang, X., Xia, C., Zhang, L., Ding, Z.: Toward depression recognition using EEG and eye tracking: an ensemble classification model CBEM. In: IEEE BIBM, 782–786 (2019). IEEE

Guo, Z., Fu, E., Pan, J., Zhang, X., Hu, B.: Anxiety detection with nonlinear group correlation fusion of electroencephalogram and eye movement. In: IEEE BIBM, pp. 2596–2602. IEEE (2020)

Domínguez-Jiménez, J.A., Campo-Landines, K.C., Martínez-Santos, J.C., Delahoz, E.J., Contreras-Ortiz, S.H.: A machine learning model for emotion recognition from physiological signals. BSPC **55**, 101646 (2020)

Shukla, J., Barreda-Angeles, M., Oliver, J., Nandi, G.C., Puig, D.: Feature extraction and selection for emotion recognition from electrodermal activity. IEEE TAFFC **12**(4), 857–869 (2019)

Greco, A., Valenza, G., Lazaro, J., Garzon-Rey, J.M., Aguilo, J., Cámara, C., Bailón, R., Scilingo, E.P.: Acute stress state classification based on electrodermal activity modeling. IEEE TAFFC **14**(1), 788–799 (2021)

Hassantabar, S., Zhang, J., Yin, H., Jha, N.K.: MHDeep: mental health disorder detection system based on wearable sensors and artificial neural networks. ACM TECS **21**(6), 1–22 (2022)

Šalkevicius, J., Damaševičius, R., Maskeliunas, R., Laukienė, I.: Anxiety level recognition for virtual reality therapy system using physiological signals. Electronics **8**(9), 1039 (2019)

Zhou, K., Schinle, M., Weimar, S., Gerdes, M., Stock, S., Stork, W.: End-to-end deep learning for stress recognition using remote photoplethysmography. In: IEEE BIBM, pp. 1435–1442. IEEE (2022)

Casado, C.Á., Cañellas, M.L., López, M.B.: Depression recognition using remote photoplethysmography from facial videos. IEEE TAFFC **14**(4), 3305–3316 (2023)

Dai, Y., Wang, X., Zhang, P., Zhang, W.: Wearable biosensor network enabled multimodal daily-life emotion recognition employing reputation-driven imbalanced fuzzy classification. Measurement **109**, 408–424 (2017)

Tomitani, N., Kanegae, H., Suzuki, Y., Kuwabara, M., Kario, K.: Stress-induced blood pressure elevation self-measured by a wearable watch-type device. Am. J. Hypertens. **34**(4), 377–382 (2021)

Lim, Y., Kim, B.C., Yoon, S.S., Kim, H.J., Lee, S.J., Lee, M.H., Kim, J.H., Park, S.J., Jeong, S., Han, H.W.: Inverse association between changes in systolic and diastolic blood pressure and risk of depression: a nationally representative cohort study. J. Affect. Disord. **335**, 49–56 (2023)

Carvalho, S., Leite, J., Galdo-Álvarez, S., Gonçalves, O.F.: The emotional movie database (EMDB): a self-report and psychophysiological study. Appl. Psychophysiol. Biofeedback **37**, 279–294 (2012)

Zheng, W.-L., Lu, B.-L.: Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. IEEE Trans. Auton. Ment. Dev. **7**(3), 162–175 (2015)

Abadi, M.K., Subramanian, R., Kia, S.M., Avesani, P., Patras, I., Sebe, N.: DECAF: MEG-based multimodal database for decoding affective physiological responses. IEEE TAFFC **6**(3), 209–222 (2015)

Katsigiannis, S., Ramzan, N.: Dreamer: a database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices. IEEE JBHI **22**(1), 98–107 (2017)

Lichtenauer, J., Soleymani, M.: MAHNOB-HCI-tagging database, London (2011)

Ringeval, F., Sonderegger, A., Sauer, J., Lalanne, D.: Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions. In: IEEE FG, pp. 1–8 (2013). IEEE

Song, T., Zheng, W., Lu, C., Zong, Y., Zhang, X., Cui, Z.: MPED: a multi-modal physiological emotion database for discrete emotion recognition. IEEE Access **7**, 12177–12191 (2019)

Subramanian, R., Wache, J., Abadi, M.K., Vieriu, R.L., Winkler, S., Sebe, N.: ASCERTAIN: emotion and personality recognition using commercial sensors. IEEE TAFFC **9**(2), 147–160 (2016)

Miranda-Correa, J.A., Abadi, M.K., Sebe, N., Patras, I.: AMIGOS: a dataset for affect, personality and mood research on individuals and groups. IEEE TAFFC **12**(2), 479–493 (2018)

Shui, X., Zhang, M., Li, Z., Hu, X., Wang, F., Zhang, D.: A dataset of daily ambulatory psychological and physiological recording for emotion research. Scientific data **8**(1), 161 (2021)

Park, C.Y., Cha, N., Kang, S., Kim, A., Khandoker, A.H., Hadjileontiadis, L., Oh, A., Jeong, Y., Lee, U.: K-EmoCon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations. Sci. Data **7**(1), 293 (2020)

Mundnich, K., Booth, B.M., l'Hommedieu, M., Feng, T., Girault, B., L'hommedieu, J., Wildman, M., Skaaden, S., Nadarajan, A., Villatte, J.L., etal Tiles-2018, a longitudinal physiologic and behavioral data set of hospital workers. Scientific Data 7(1), 354 (2020)

Koldijk, S., Sappelli, M., Verberne, S., Neerincx, M.A., Kraaij, W.: The swell knowledge work dataset for stress and user modeling research. In: Proceedings of the 16th International Conference on Multimodal Interaction, pp. 291–298. (2014)

Healey, J.A., Picard, R.W.: Detecting stress during real-world driving tasks using physiological sensors. IEEE TITS **6**(2), 156–166 (2005)

Cai, H., Yuan, Z., Gao, Y., Sun, S., Li, N., Tian, F., Xiao, H., Li, J., Yang, Z., Li, X.: A multi-modal open dataset for mental-disorder analysis. Sci. Data **9**(1), 178 (2022)

Alexander, L.M., Escalera, J., Ai, L., Andreotti, C., Febre, K., Mangone, A., Vega-Potler, N., Langer, N., Alexander, A., Kovacs, M.: An open resource for transdiagnostic research in pediatric mental health and learning disorders. Sci. data **4**(1), 1–26 (2017)

Zhang, T., El Ali, A., Wang, C., Hanjalic, A., Cesar, P.: Weakly-supervised learning for fine-grained emotion recognition using physiological signals. IEEE TAFFC **14**(3), 2304–2322 (2023)

Khan, P., Ranjan, P., Kumar, S.: AT2GRU: a human emotion recognition model with mitigated device heterogeneity. IEEE TAFFC **14**(2), 1520–1532 (2021)

Anusha, A., Jose, J., Preejith, S., Jayaraj, J., Mohanasankar, S.: Physiological signal based work stress detection using unobtrusive sensors. Biomed. Phys. Eng. Exp. **4**(6), 065001 (2018)

Bufano, P., Laurino, M., Said, S., Tognetti, A., Menicucci, D.: Digital phenotyping for monitoring mental disorders: systematic review. J. Med. Internet Res. **25**, 46778 (2023)

Harrison, O.K., Köchli, L., Marino, S., Luechinger, R., Hennel, F., Brand, K., Hess, A.J., Frässle, S., Iglesias, S., Vinckier, F.: Interoception of breathing and its relationship with anxiety. Neuron **109**(24), 4080–4093 (2021)

Liao, S.-C., Wu, C.-T., Huang, H.-C., Cheng, W.-T., Liu, Y.-H.: Major depression detection from EEG signals using kernel eigen-filter-bank common spatial patterns. Sensors **17**(6), 1385 (2017)

Wang, J.-S., Lin, C.-W., Yang, Y.-T.C.: A k-nearest-neighbor classifier with heart rate variability feature-based transformation algorithm for driving stress recognition. Neurocomputing **116**, 136–143 (2013)

Lehmann, C., Koenig, T., Jelic, V., Prichep, L., John, R.E., Wahlund, L.-O., Dodge, Y., Dierks, T.: Application and comparison of classification algorithms for recognition of Alzheimer's disease in electrical brain activity (eeg). J. Neurosci. Methods **161**(2), 342–350 (2007)

Acharya, U.R., Sudarshan, V.K., Adeli, H., Santhosh, J., Koh, J.E., Puthankatti, S.D., Adeli, A.: A novel depression diagnosis index using nonlinear features in EEG signals. Eur. Neurol. **74**(1–2), 79–83 (2015)

Gayathri, S., Anand, A., Vijayvargiya, A., Pushpalatha, M., Moorthy, V., Kumar, S., Harichandana, B.: Emosens: Emotion recognition based on sensor data analysis using lightGBM. In: 2022 IEEE International Conference on Electronics, Computing and Communication Technologies, pp. 1–6. IEEE (2022)

Afify, H.M., Mohammed, K.K., Hassanien, A.E.: Stress detection based EEG under varying cognitive tasks using convolution neural network. Neural Comput. Appl. , 1–15 (2025)

Kuang, W., Jin, S., Wang, D., Zheng, Y., Zou, Y., Wu, K.: Emotracer: A user-independent wearable emotion tracer with multi-source physiological sensors based on few-shot learning. In: IEEE BIBM, pp. 2680–2687. IEEE Computer Society (2024)

Song, X., Yan, D., Zhao, L., Yang, L.: LSDD-EEGNet: an efficient end-to-end framework for EEG-based depression detection. BSPC **75**, 103612 (2022)

Chen, T., Guo, Y., Hao, S., Hong, R.: Exploring self-attention graph pooling with EEG-based topological structure and soft label for depression detection. IEEE TAFFC **13**(4), 2106–2118 (2022)

Zheng, W., Yan, L., Wang, F.-Y.: Two birds with one stone: Knowledge-embedded temporal convolutional transformer for depression detection and emotion recognition. IEEE TAFFC **14**(4), 2595–2613 (2023)

Wang, Y., Qiu, S., Li, D., Du, C., Lu, B.-L., He, H.: Multi-modal domain adaptation variational autoencoder for EEG-based emotion recognition. IEEE/CAA J Automatica Sinica **9**(9), 1612–1626 (2022)

Li, J., Struzik, Z., Zhang, L., Cichocki, A.: Feature learning from incomplete EEG with denoising autoencoder. Neurocomputing **165**, 23–31 (2015)

Qiu, Y., Zhou, W., Yu, N., Du, P.: Denoising sparse autoencoder-based ictal EEG classification. IEEE TNSRE **26**(9), 1717–1726 (2018)

Wen, T., Zhang, Z.: Deep convolution neural network and autoencoders-based unsupervised feature learning of EEG signals. IEEE Access **6**, 25399–25410 (2018)

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. NIPS **27** (2014)

Chen, G., Zhu, Y., Hong, Z., Yang, Z.: EmotionalGAN: generating ECG to enhance emotion state classification. In: Proceedings of the 2019 International Conference on Artificial Intelligence and Computer Science, pp. 309–313. (2019)

Li, X., Chen, C.L.P., Chen, B., Zhang, T.: Gusa: graph-based unsupervised subdomain adaptation for cross-subject EEG emotion recognition. IEEE TAFFC **15**(3), 1451–1462 (2024)

Liu, C., Zhou, X., Zhu, Z., Zhai, L., Jia, Z., Liu, Y.: . In: VBH-GNN: variational Bayesian heterogeneous graph neural networks for cross-subject emotion recognition: ICLR (2024)

Gu, R.-F., Li, R., Zheng, W.-L., Lu, B.-L.: Cross-subject decision confidence estimation from EEG signals using spectral-spatial-temporal adaptive GCN with domain adaptation. In: IJCNN, pp. 1–8. IEEE (2023)

Pan, J., Liang, R., He, Z., Li, J., Liang, Y., Zhou, X., He, Y., Li, Y.: ST-SCGNN: a Spatio-temporal self-constructing graph neural network for cross-subject EEG-based emotion recognition and consciousness detection. IEEE JBHI **28**(2), 777–788 (2024)

She, Q., Zhang, C., Fang, F., Ma, Y., Zhang, Y.: Multisource associate domain adaptation for cross-subject and cross-session EEG emotion recognition. IEEE TIM **72**, 1–12 (2023)
Liao, Y., Zhang, Y., Wang, S., Zhang, X., Zhang, Y., Chen, W., Gu, Y., Huang, L.: CLDTA: contrastive learning based on diagonal transformer autoencoder for cross-dataset EEG emotion recognition. arXiv preprint arXiv:2406.08081 (2024)
Zhou, Y., Li, F., Li, Y., Ji, Y., Zhang, L., Chen, Y., Zheng, W., Shi, G.: Eeg-based emotion style transfer network for cross-dataset emotion recognition. arXiv preprint arXiv:2308.05767 (2023)
Jiang, H., Shen, F., Chen, L., Peng, Y., Guo, H., Gao, H.: Joint domain symmetry and predictive balance for cross-dataset EEG emotion recognition. J. Neurosci. Methods **400**, 109978 (2023)
Yuan, D., Yue, J., Xiong, X., Jiang, Y., Zan, P., Li, C.: A regression method for EEG-based cross-dataset fatigue detection. Front. Physiol. **14**, 1196919 (2023)
Hong, Z., Li, Z., Zhong, S., Lyu, W., Wang, H., Ding, Y., He, T., Zhang, D.: CrossHAR: generalizing cross-dataset human activity recognition via hierarchical self-supervised pretraining. IMWUT **8**(2), 1–26 (2024)
He, Z., Zhong, Y., Pan, J.: An adversarial discriminative temporal convolutional network for EEG-based cross-domain emotion recognition. Comput. Biol. Med. **141**, 105048 (2022)
He, Z., Zhong, Y., Pan, J.: Joint temporal convolutional networks and adversarial discriminative domain adaptation for EEG-based cross-subject emotion recognition. ICASSP , 3214–3218 (2022) IEEE
Li, C., Bian, N., Zhao, Z., Wang, H., Schuller, B.W.: Multi-view domain-adaptive representation learning for EEG-based emotion recognition. Information Fusion **104**, 102156 (2024)
Wang, J., Wang, A., Hu, H., Wu, K., He, D.: Multi-source domain generalization for ECG-based cognitive load estimation: Adversarial invariant and plausible uncertainty learning. In: ICASSP, pp. 1631–1635 (2024). IEEE
An, Y., Hu, S., Liu, S., Wang, X., Gu, Z., Zhang, Y.: LGDAAN-Nets: a local and global domain adversarial attention neural networks for EEG emotion recognition. KBS **318**, 113613 (2025)
Gao, C., Jia, L., Xu, J., Yang, P.: UMDA-DDSTGN: an unsupervised meta-domain adaptation method using dynamic directed spatial-temporal graph network for EEG-based emotion recognition. KBS **325**, 113938 (2025)
Jia, Z., Zhao, F., Guo, Y., Chen, H., Jiang, T., Center, B.: Multi-level disentangling network for cross-subject emotion recognition based on multimodal physiological signals. In: Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, pp. 3069–3077 (2024)
Shen, X., Liu, X., Hu, X., Zhang, D., Song, S.: Contrastive learning of subject-invariant EEG representations for cross-subject emotion recognition. IEEE TAFFC **14**(3), 2496–2511 (2022)
Cai, H., Pan, J.: Two-phase prototypical contrastive domain generalization for cross-subject EEG-based emotion recognition. In: IEEE ICASSP, pp. 1–5. IEEE (2023)
Liu, D., Li, S., Wang, Z., Li, W., Wu, D.: SDDA: spatial distillation based distribution alignment for cross-headset EEG classification. IEEE TBME, 1–12 (2025)
Tang, J., Li, Y., Su, C.-W., Fan, X., Zheng, Y., Wang, H., Naeem, H., Fang, P., Wang, J., Yao, N., et al.: Uda-dda: Unsupervised domain adaptation with dynamic distribution alignment network for emotion recognition using EEG signals. Neurocomputing, 131715 (2025)
Xu, C.-Y., Fan, F.-Y., Zhao, L.-X., Jin, L.-C., Zhang, Y.-H., Meng, Q.-H.: Major depressive disorder detection via temporal-frequency-spatial transformer with sub-domain knowledge alignment using EEG. Neural Netw. , 107965 (2025)
Shi, H., Zhao, H., Yao, W.: A transfer fusion framework for body sensor networks (BSNS): dynamic domain adaptation from distribution evaluation to domain evaluation. Information Fusion **91**, 338–351 (2023)
Kuang, W., Zhang, Y., Cheng, Y., Wu, K., Fan, C., Zou, Y.: Generalizable psychiatric screening from wearable: causal disentanglement of multi-channel biosignals with interpretability analysis. In: IEEE BIBM, pp. 3745–3750. (2025)
Chen, B., Chen, C.L.P., Zhang, T.: GDDN: graph domain disentanglement network for generalizable EEG emotion recognition. IEEE TAFFC **15**(3), 1739–1753 (2024)
Pan, L., Wang, K., Xu, L., Sun, X., Yi, W., Xu, M., Ming, D.: Riemannian geometric and ensemble learning for decoding cross-session motor imagery electroencephalography signals. J. Neural Eng. **20**(6), 066011 (2023)
Li, W., Huan, W., Shao, S., Hou, B., Song, A.: MS-FRAN: a novel multi-source domain adaptation method for EEG-based emotion recognition. IEEE JBHI **27**(11), 5302–5313 (2023)
Zhang, H., Zuo, T., Chen, Z., Wang, X., Sun, P.Z.: Evolutionary ensemble learning for EEG-based cross-subject emotion recognition. IEEE JBHI **28**(7), 3872–3881 (2024)
Zhang, Y., Jin, S., Kuang, W., Zheng, Y., Song, Q., Fan, C., Zou, Y., Leung, V.C.M., Wu, K.: DepGuard: depression recognition and episode monitoring system with a ubiquitous wrist-worn device. IEEE TMC **25**(1), 197–214 (2026)
Duan, D., Yang, H., Lan, G., Li, T., Jia, X., Xu, W.: Emgsense: A low-effort self-supervised domain adaptation framework for EMG sensing. In: IEEE PerCom, pp. 160–170. (2023)
Sarkar, P., Etemad, A.: Self-supervised ECG representation learning for emotion recognition. IEEE TAFFC **13**(3), 1541–1554 (2022)
Wu, Y., Daoudi, M., Amad, A.: Transformer-based self-supervised multimodal representation learning for wearable emotion recognition. IEEE TAFFC **15**(1), 157–172 (2023)
Zhang, L., Shi, H., Li, Z., Zheng, W.-L., Lu, B.-L.: Multi-view self-supervised domain adaptation for EEG-based emotion recognition. IEEE TAFFC **16**(4), 3055–3066 (2025)
Can, Y.S., Benouis, M., Mahesh, B., André, E.: Application of multimodal self-supervised architectures for daily life affect recognition. IEEE TAFFC **16**(3), 2454–2465 (2025)
Cai, H., Pan, J., Xiao, Q., Jin, J., Li, Y., Xie, Q.: Decoding musical neural activity in patients with disorders of consciousness through self-supervised contrastive domain generalization. IEEE TAFFC **16**(2), 726–743 (2025)
Cortiñas-Lorenzo, K., Lacey, G.: Toward explainable affective computing: a review. IEEE TNNLS **35**(10), 13101–13121 (2024)
Li, C., Li, P., Zhang, Y., Li, N., Si, Y., Li, F., Cao, Z., Chen, H., Chen, B., Yao, D., Xu, P.: Effective emotion recognition by learning discriminative graph topologies in EEG brain networks. IEEE TNNLS **35**(8), 10258–10272 (2024)
Zhang, X., Xu, C., Xue, W., Hu, J., He, Y., Gao, M.: Emotion recognition based on multichannel physiological signals with comprehensive nonlinear processing. Sensors **18**(11), 3886 (2018)
Valenza, G., Lanata, A., Scilingo, E.P.: The role of nonlinear dynamics in affective valence and arousal recognition. IEEE TAFFC **3**(2), 237–249 (2011)
Bodaghi, M., Hosseini, M., Gottumukkala, R.: A multimodal intermediate fusion network with manifold learning for stress detection. In: International Conference on Computing and Machine Intelligence, pp. 1–8. IEEE (2024)
Kim, J., André, E.: Emotion recognition based on physiological changes in music listening. IEEE TPAMI **30**(12), 2067–2083 (2008)
Lin, J., Pan, S., Lee, C.S., Oviatt, S.: An explainable deep fusion network for affect recognition using physiological signals. In: CIKM, pp. 2069–2072. (2019)
Marey, M., Salem, D., El Rashidy, N., ELBakry, H.: Enhancing the diagnosis of depression and anxiety through explainable machine

learning methods. International Journal of Advanced Computer Science & Applications **15**(4) (2024)
Giannakakis, G., Grigoriadis, D., Giannakaki, K., Simantiraki, O., Roniotis, A., Tsiknakis, M.: Review on psychological stress detection using biosignals. IEEE TAFFC **13**(1), 440–460 (2019)
Shen, J., Chen, J., Ma, Y., Cao, Z., Zhang, Y., Hu, B.: Explainable depression recognition from EEG signals via graph convolutional network. In: IEEE BIBM, pp. 1406–1412. IEEE (2023)
Su, Y., Hu, B., Xu, L., Cai, H., Moore, P., Zhang, X., Chen, J.: EmotionO+: Physiological signals knowledge representation and emotion reasoning model for mental health monitoring. In: IEEE BIBM, pp. 529–535 (2014). IEEE
Yang, K., Wang, C., Gu, Y., Sarsenbayeva, Z., Tag, B., Dingler, T., Wadley, G., Goncalves, J.: Behavioral and physiological signals-based deep multimodal approach for mobile emotion recognition. IEEE TAFFC **14**(2), 1082–1097 (2021)
Ye, M., Chen, C.L.P., Zhang, T.: Hierarchical dynamic graph convolutional network with interpretability for EEG-based emotion recognition. IEEE TNNLS **36**(11), 19489–19500 (2025)
Yu, R., Pan, C., Fei, X., Chen, M., Shen, D.: Multi-graph attention networks with bilinear convolution for diagnosis of schizophrenia. IEEE JBHI **27**(3), 1443–1454 (2023)
Song, T., Liu, S., Zheng, W., Zong, Y., Cui, Z.: Instance-adaptive graph for eeg emotion recognition, vol. 34, pp. 2701–2708. AAAI (2020)
Qiu, S., Miller, M.I., Joshi, P.S., Lee, J.C., Xue, C., Ni, Y., Wang, Y., De Anda-Duran, I., Hwang, P.H., Cramer, J.A.: Multimodal deep learning for Alzheimer's disease dementia assessment. Nat. Commun. **13**(1), 3404 (2022)
Boulanouar, F.Z., Raed, M.W., Huseyinov, I., Akdeniz, R., Kotenko, I., Fedorchenko, E.: Classification of mental stress using an explainable machine learning approach. In: International Conference on Intelligent Information Technologies for Industry, pp. 245–255. Springer (2024)
Xu, Y., Du, Y., Li, L., Lai, H., Zou, J., Zhou, T., Xiao, L., Liu, L., Ma, P.: AMDET: attention based multiple dimensions EEG transformer for emotion recognition. IEEE TAFFC **15**(3), 1067–1077 (2023)
Lundberg, S.M., Lee, S.-I.: A unified approach to interpreting model predictions. NIPS 30, (2017)
Ribeiro, M.T., Singh, S., Guestrin, C.: " why should i trust you?" explaining the predictions of any classifier. In: SIGKDD, pp. 1135–1144 (2016)
Simonyan, K., Vedaldi, A., Zisserman, A.: Visualising image classification models and saliency maps. Deep Inside Convolutional Netw. **2**, 2 (2014)
Selvaraju, R.R., Das, A., Vedantam, R., Cogswell, M., Parikh, D., Batra, D.: Grad-cam: Why did you say that? arXiv preprint, (2016). arXiv:1611.07450
Gu, Z.: Complex heatmap visualization. Imeta **1**(3), 43 (2022)
Zhu, J., Yang, C., Xie, X., Wei, S., Li, Y., Li, X., Hu, B.: Mutual information based fusion model (MIBFM): mild depression recognition using EEG and pupil area signals. IEEE TAFFC **14**(3), 2102–2115 (2022)
Ying, Z., Bourgeois, D., You, J., Zitnik, M., Leskovec, J.: Gnnexplainer: Generating explanations for graph neural networks. NIPS 32, (2019)
Kanjo, E., Younis, E.M., Sherkat, N.: Towards unravelling the relationship between on-body, environmental and emotion data using sensor information fusion approach. Inform Fusion **40**, 18–31 (2018)
Cheng, H., Yang, Z., Zhang, X., Yang, Y.: Multimodal sentiment analysis based on attentional temporal convolutional network and multi-layer feature fusion. IEEE TAFFC **14**(4), 3149–3163 (2023)
Fu, B., Chu, W., Gu, C., Liu, Y.: Cross-modal guiding neural network for multimodal emotion recognition from EEG and eye movement signals. IEEE JBHI **28**(10), 5865–5876 (2024)
Liu, B., Guo, J., Chen, C.P., Wu, X., Zhang, T.: Fine-grained interpretability for EEG emotion recognition: Concat-Aided Grad-CAM and systematic brain functional network. IEEE TAFFC **15**(2), 671–684 (2023)
Ye, W., Zhang, Z., Teng, F., Zhang, M., Wang, J., Ni, D., Li, F., Xu, P., Liang, Z.: Semi-supervised dual-stream self-attentive adversarial graph contrastive learning for cross-subject EEG-based emotion recognition. IEEE TAFFC **16**(1), 290–305 (2025)
Li, S., Zhang, T., Chen, B., Chen, C.P.: MIA-Net: multi-modal interactive attention network for multi-modal affective analysis. IEEE TAFFC **14**(4), 2796–2809 (2023)
Zhang, T., El Ali, A., Hanjalic, A., Cesar, P.: Few-shot learning for fine-grained emotion recognition using physiological signals. IEEE TMM **25**, 3773–3787 (2022)
Chen, T., Hong, R., Guo, Y., Hao, S., Hu, B.: MS^2-GNN: exploring GNN-based multimodal fusion network for depression detection. IEEE TCYB **53**(12), 7749–7759 (2022)
Hu, Y., Chen, J., Chen, J., Wang, W., Zhao, S., Hu, X.: An ensemble classification model for depression based on wearable device sleep data. IEEE JBHI **28**(5), 2602–2612 (2024)
Zhao, Y., Cao, X., Lin, J., Yu, D., Cao, X.: Multimodal affective states recognition based on multiscale CNNs and biologically inspired decision fusion model. IEEE TAFFC **14**(2), 1391–1403 (2021)
Liu, Y.-T., Pal, N.R., Marathe, A.R., Lin, C.-T.: Weighted fuzzy Dempster-Shafer framework for multimodal information integration. IEEE TFS **26**(1), 338–352 (2017)
Rahman, M.M., Sarkar, A.K., Hossain, M.A., Moni, M.A.: EEG-based emotion analysis using non-linear features and ensemble learning approaches. ESWA **207**, 118025 (2022)
Guo, G., Gao, P., Zheng, X., Ji, C.: Multimodal emotion recognition using CNN-SVM with data augmentation. In: IEEE BIBM, pp. 3008–3014. IEEE (2022)
Li, C., Bao, Z., Li, L., Zhao, Z.: Exploring temporal representations by leveraging attention-based bidirectional LSTM-RNNs for multimodal emotion recognition. Inform Proc Manag **57**(3), 102185 (2020)
Zhu, J., Xie, X., Yang, C., Wei, S., Li, X., Hu, B.: Hybrid fusion model based on DBN and secondary classifier: Multimodal mild depression recognition using EEG and eye movement. In: IEEE BIBM, pp. 1846–1850. IEEE (2022)
Yao, Y., Doretto, G.: Boosting for transfer learning with multiple sources. In: CVPR, pp. 1855–1862 (2010). IEEE
Yuan, H., Yu, H., Wang, J., Li, K., Ji, S.: On explainability of graph neural networks via subgraph explorations. In: ICML, pp. 12241–12252. (2021) . (**PMLR**)
Huang, Q., Yamada, M., Tian, Y., Singh, D., Chang, Y.: GraphLIME: local interpretable model explanations for graph neural networks. IEEE TKDE **35**(7), 6968–6972 (2022)
Yuan, H., Tang, J., Hu, X., Ji, S.: XGNN: Towards model-level explanations of graph neural networks. In: SIGKDD, pp. 430–438. (2020)
Wan, A., Dunlap, L., Ho, D., Yin, J., Lee, S., Jin, H., Petryk, S., Bargal, S.A., Gonzalez, J.E.: Nbdt: Neural-backed decision trees, (2020). arXiv:2004.00221 arXiv preprint

Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J.: etal Learning transferable visual models from natural language supervision. In: ICML, pp. 8748–8763. (2021) . (**PmLR**)

Xu, X., Yao, B., Dong, Y., Gabriel, S., Yu, H., Hendler, J., Ghassemi, M., Dey, A.K., Wang, D.: Mental-LLM: leveraging large language models for mental health prediction via online text data. IMWUT **8**(1), 1–32 (2024)

Hu, H., Ma, C., Wang, Q., Lin, L., Zhou, Y., Cui, L., Ma, F., Tian, Q.: TheraMind: a strategic and adaptive agent for longitudinal psychological counseling. In: Proceedings of the ACM Web Conference 2026, pp. 9136–9147 (2026)

Yi, Y., Xu, Y., Yang, B., Tian, Y.: A weighted co-training framework for emotion recognition based on EEG data generation using frequency-spatial diffusion transformer. IEEE TAFFC **15**(4), 2055–2069 (2024)

Pan, Y., Jiang, J., Jiang, K., Liu, X.: Disentangled-multimodal privileged knowledge distillation for depression recognition with incomplete multimodal data. In: ACM MM, pp. 5712–5721 (2024)

Hu, Y., Zhang, S., Dang, T., Jia, H., Salim, F.D., Hu, W., Quigley, A.J.: Exploring large-scale language models to evaluate EEG-based multimodal data for mental health. In: Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 412–417 (2024)

Thapa, B., Rivas, M., Griffith, H., Rathore, H.: StressLLM: Large language models for stress prediction via wearable sensor data. In: ICCE, pp. 1–6. (2025)

Shen, J., Zhu, K., Ma, R., Hu, W., Tan, X., Deng, N., Zhou, X., Liu, Q., Li, C., Xu, W., Xu, C., Zhang, Y., Hu, B.: EmoSavior: depression recognition and intervention via multimodal physiological signals and large language models. Inform. Fusion **127**, 103772 (2026)

Hu, H., You, L., Xu, H., Wang, Q., Yu, F.R., Ma, F., Cheng, Z., Lian, Z., Zhou, Y., Cui, L.: EmoBench-M: benchmarking emotional intelligence for multimodal large language models, (2025). arXiv:2502.04424 arXiv preprint

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



Yufei Zhang He is currently pursuing the PhD degree in the College of Computer Science and Software Engineering, Shenzhen University. His research interests include mobile computing, affective computing and spatial-temporal data analysis. He was honored with the Best Paper Nomination at the IEEE PerCom 2026.



Wenting Kuang She is currently working toward the master's degree with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China. Her research interests covers affective computing and intelligent sensing.



Yulan Li She is currently a master's student in the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China. She received her B.Eng. degree from the School of Computing / Software, University of South China(USC), Hengyang, China, in 2024. Her research interests include affective computing and mobile computing.



Yuqi Cheng She received the bachelor degree from the School of Mathematics and Physics, Wuhan Institute of Technology (WIT), in 2024. She is currently pursuing the master degree at the College of Computer Science and Software Engineering, Shenzhen University. Her research interests include affective computing and mobile computing.



Kaishun Wu He received the PhD degree from the Hong Kong University of Science and Technology, Hong Kong, in 2011. He is currently a professor in information hub with the Hong Kong University of Science and Technology (Guangzhou). His research interests include wireless communications and mobile computing. He won several best paper awards of international conferences, such as IEEE Globecom 2012 and IEEE MASS 2014.



Yongpan Zou He received the PhD degree from the Department of Computer Science and Engineering (CSE), Hong Kong University of Science and Technology, in 2017. He is currently an associate professor with the College of Computer Science and Software Engineering, Shenzhen University. His research interests include ubiquitous sensing, mobile computing, and human-computer interaction.