This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2025.3591096

DepGuard: Depression Recognition and Episode Monitoring System with A Ubiquitous Wrist-worn Device

Yufei Zhang, Shuo Jin, Wenting Kuang, Yuda Zheng, Qifeng Song, Changhe Fan, Yongpan Zou, *Member, IEEE*, Victor C. M. Leung, *Life Fellow, IEEE*, Kaishun Wu, *Fellow, IEEE*

***Abstract*—Depression significantly impacts mental health, severely disrupting patients' daily lives. During depressive episodes, individuals may experience symptoms such as excessive guilt, self-harm, and suicidal ideation. Compared to proprietary devices like brain electrode caps, wearable technologies for depression detection have gained attention due to their affordability and portability—enabling real-time monitoring of depressive states. However, challenges such as low-quality data from ubiquitous devices, individual variability, and the complexity of multimodal physiological signal analysis limit model generalizability. To address these issues, we present *DepGuard*, a novel ubiquitous wearable system for depression assessment based on multimodal physiological signals. *DepGuard* performs a two-stage detection process: depression recognition and real-time episode monitoring. For depression recognition, we propose an unsupervised domain adaptation method to reduce the domain gap between source and target subjects. For episode monitoring, we employ a few-shot learning strategy to enable personalized modeling. Both approaches enhance cross-subject generalization. Our system achieves 90.75% accuracy in cross-subject depression recognition using 30 unlabeled samples per target subject, and 93.52% accuracy in episode monitoring using 15 labeled samples per class.**

***Index Terms*—Wearable devices; Depression recognition; Unsupervised domain adaptation; Depressive episode monitoring**

Yufei Zhang, Shuo Jin, Wenting kuang, Yuda Zheng, Qifeng Song and Yongpan Zou are with the College of Computer Science and Software Engineering, Shenzhen University, 3688 Nanhai Ave, Shenzhen, Guangdong, China. Email: yfyaya@foxmail.com; {jinshuo2022, kuangwenting2023, songqifeng2021, 2021150056}@email.szu.edu.cn; yongpan@szu.edu.cn.

Changhe Fan is with Department of Psychology, Guangdong Second Provincial General Hospital. Email: changhefan@yahoo.com.

Kaishun Wu is with the Information Hub, Hong Kong University of Science and Technology (Guangzhou), 1 Duxue Road, Guangzhou, Guangdong, China. Email: wuks@hkust-gz.edu.cn.

Victor C. M. Leung is with the Artificial Intelligence Research Institute, Shenzhen MSU-BIT University, Shenzhen 518172, Guangdong, China. Email: vleung@ieee.org

This work was supported in part by China NSFC under Grant 62172286 and Grant U2001207, in part by Key Project of Department of Education of Guangdong Province under Grant 2024ZDZX1011, in part by Guangdong NSF under Grant 2022A1515011509, in part by the Guangdong Provincial Key Lab of Integrated Communication, Sensing and Computation for Ubiquitous Internet of Things under Grant 2023B1212010007, in part by the Project of DEGP under Grant 2023KCXTD042, in part by Guangdong "Pearl River Talent Recruitment Program" under Grant 2019ZT08X603, in part by Guangdong "Pearl River Talent Plan" under Grant 2019JC01X235, in part by 111 Center under grant D25008, in part by Shenzhen Science and Technology Foundation (ZDSYS20190902092853047), and in part by Tencent "Rhinoceros Birds" - Scientific Research Fund for Young Researchers of Shenzhen University. Yongpan Zou is the corresponding author.

Manuscript received December 31, 2024; revised April, 14, 2025.

I. Introduction

Depression is a common mental disorder affecting approximately 280 million people worldwide [1]. It is characterized by persistent sadness, insomnia, loss of interest, and reduced ability to carry out daily activities for at least two weeks [2]. Without timely treatment, individuals are at increased risk of suicidal behavior. The Lancet Psychiatry, a leading journal in the field, emphasizes that early recognition and intervention targeting lifestyle habits and risk factors are critical for effective prevention [3]. However, continuous supervision by doctors or family members is not feasible, making early detection of depressive episodes essential to improving patient outcomes and reducing the risk of self-harm or suicide. Real-time monitoring, coupled with timely communication to healthcare professionals, plays a vital role in ensuring prompt and appropriate care. This proactive approach not only helps prevent adverse events but also enhances overall mental health management.

In recent years, depression recognition research has increasingly combined individual-level insights with advanced technologies to enable objective, passive assessment of depressive states. Specifically, given the characteristic patterns of negative affect and social withdrawal in individuals with depression, mental states can be inferred through the analysis of lifestyle data, including sleep patterns [4], [5], [6], social connectivity [7], [8], [9], behavioral records [10], [11], [12], [13]. However, behavior-based methods for depression detection are often hindered by individuals intentionally or unintentionally concealing their symptoms. Similarly, traditional self-report scales suffer from subjectivity and typically lack robust quantitative analysis. Physiological signal monitoring offers a promising alternative. Depression-related symptoms such as emotional instability and physical fatigue often manifest as measurable physiological changes that are difficult to consciously control, making them less susceptible to self-report bias [14]. High-end commercial wearable devices can capture rich physiological data-such as electroencephalography (EEG), electrocardiography (ECG), and magnetoencephalography (MEG) [15], but their high cost, limited accessibility, and operational complexity restrict widespread application. In contrast, ubiquitous wearable devices offer low-cost and portable solutions, but they tend to capture only shallow physiological patterns, which weakens their effectiveness in depression recognition and episode monitoring. Consequently,

Authorized licensed use limited to: SHENZHEN UNIVERSITY. Downloaded on July 29,2025 at 15:21:18 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2025.3591096

proprietary equipment remains the standard for high-accuracy detection [16], [17], [18], despite its impracticality for everyday use [19], [20].

Recent research has begun to explore the potential of ubiquitous devices for mental health monitoring [21], [22], [23], [24], though their application in depression detection remains in its early stages. For instance, Hassantabar et al. [21] introduce MHDeep, a framework that integrates wearable sensor data with artificial neural networks to classify mental health disorders. Their model utilizes multimodal inputs such as galvanic skin response, skin temperature, inter-beat interval, and tri-axial acceleration, collected from smartwatches and smartphones. Similarly, Hu et al. [23] develop an ensemble machine learning model that classifies users as depressed or non-depressed based on sleep data from wearable devices. Ahmed et al. [25] propose a multimodal framework that combines convolutional neural networks, attention mechanisms, and random forests to assess affective states and depression. Despite these advancements, several key challenges remain unresolved.

C1: How can depressive episodes be detected quickly to provide timely interventions? Much of the existing research focuses solely on depression recognition at the model level, using either professional or consumer-grade equipment, with limited attention to real-time monitoring of depressive episodes. However, the progression from diagnosis to ongoing monitoring is a continuous process, inherently linking detection with treatment. These tasks are causally and practically connected in real-world settings. To alleviate the substantial burden placed on healthcare providers and family members during inpatient care, a comprehensive systems approach is needed—one that enables timely intervention and mitigates the psychological risks associated with depressive episodes.

C2: How can we improve model adaptability to achieve high-accuracy detection for new subjects? Existing research often overlooks cross-domain challenges, limiting model performance on unseen individuals. Due to the uniqueness of individual physiological traits and emotional responses, even those with similar mental health conditions may exhibit different physiological signal patterns. This challenge is compounded by limited data and scarce annotations, highlighting the need for models to learn more discriminative and generalizable patterns.

To address the aforementioned challenges, we introduce *DepGuard*, a depression assessment system that analyzes multimodal physiological signals to identify both depressive disorders and ongoing depressive episodes. To tackle the issue of real-time episode monitoring, *DepGuard* integrates a dedicated module that operates in the system's second phase. Users wear a wrist-worn device that continuously collects physiological data, which is then transmitted via Bluetooth to a paired mobile application. A lightweight and personalized model is deployed within the app to perform real-time predictions and deliver results to the user. Given the device's limited computational resources, we employ a knowledge distillation strategy to compress the model, thereby reducing its complexity while maintaining predictive performance.

To address the second challenge, we design task-specific strategies that align with the unique requirements of each objective. For depression recognition, we utilize an unsupervised domain adaptation approach to transfer knowledge from the source domain (*i.e.*, labeled data from known subjects) to the target domain (*i.e.*, unseen subject), thereby improving generalization. For depressive episode monitoring, we apply few-shot learning to enhance adaptability with minimal labeled data. By integrating transfer learning and few-shot learning, our system effectively mitigates inter-individual variability and enables robust, cross-subject depression assessment and monitoring.

Our main contributions can be summarized as follows.

- We present *DepGuard*, a wearable depression assessment system designed for both depression diagnosis and real-time monitoring of depressive episodes. Evaluated under cross-subject testing on a real-world dataset, *DepGuard* achieves 90.75% accuracy in depression recognition and 93.52% in depressive episode monitoring.
- We propose an unsupervised domain adaptation method for depression recognition that bridges the gap between source and target domains. This approach minimizes data and training overhead for new subjects while maintaining high recognition accuracy.
- We propose a few-shot learning approach combined with a knowledge distillation strategy for depressive episode monitoring, aiming to build high-performance personalized models on edge devices with minimal computational cost.
- To facilitate this line of the research[1], we collect a real-world dataset from individuals diagnosed with depression by certified psychiatrists at our partner hospital.

The remainder of the paper is conducted as follows. Sec. II reviews the related works about different types of devices and cross-subject detection of depression. Subsequently, Sec. III introduces the two-stage procedure of *DepGuard*. Then, Sec. IV conducts the implementation of our proposals. Sec. V evaluates the effectiveness of *DepGuard*. Finally, Sec. VI and Sec. VII discuss and conclude our works.

II. Related Works

A. Depression Detection with Different Devices

Proprietary devices such as EEG [16], [17], [26], ECG [18], [27], MEG [28], functional near-infrared spectroscopy (fNIRS) [29] and functional magnetic resonance imaging (fMRI) [30] are commonly used to detect physiological signals. The high resolution, precision, and stability of these devices allow for the detection of subtle physiological changes, providing researchers with reliable, high-quality data. However, these devices require complex operations when worn. For example, Huang et al. [28] use a bulky whole-head 306-channel Elekta Neuronmag Vectorview to capture MEG signals. Pange et al. [31] using EEG and ECG signals require conductive gel and power connections, making the setup inconvenient for users.

[1]Our source code is available at *https://github.com/ISAC-GROUP/AFFC-DepGuard*.

Authorized licensed use limited to: SHENZHEN UNIVERSITY. Downloaded on July 29,2025 at 15:21:18 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2025.3591096

For depression detection based on portable devices, several studies [11] have indicated that the great significance of utilizing sensor data from smartphones and wearable devices, such as Fitbit and Microsoft Band, can effectively represent the features of depression defined in the DSM-5 diagnostic criteria [2]. These sensors data encompass various dimensions including phone usage patterns, acceleration, GPS, sleep duration, and audio recordings, providing a comprehensive understanding of subjects physical activity, social interactions, and attention levels in daily life. Analyzing this data opens new possibilities for precision medicine in diagnosing and treating depression. As a result, portable devices are explored for physiological signal detection and further utilized to predict depression. Specifically, Tazawa et al. [32] apply machine learning to screen and assess depression severity using Silmee W20 wristband data, achieving 76% accuracy. Moshe et al. [33] utilize Oura Ring data on activity, sleep, and heart rate variability to predict depression, revealing significant correlations with depressive symptoms. Moreover, Hassantabar et al. [21] collect physiological signals from both the E4 smartwatch and Samsung smartphone, achieving an average accuracy of 87.3% in distinguishing healthy and depressive instances. Based on a lightweight TN1012/ST Pulse Transducer, the research [22] develops a pulse rate variability detection model to identify depression with the highest accuracy of 98.46%. However, data collected by portable devices face technical constraints, including low temporal resolution and noise susceptibility, which collectively impact prediction reliability. These limitations motivate the development of a wrist-worn device equipped with multi-source physiological sensors. Furthermore, due to inter-individual variability, research on depression recognition remains underdeveloped and requires further progress.

B. Cross-subject Depression Detection

Individual-independent detection tasks have become a central research challenge. In depression detection, physiological and psychological variations, such as differences in neural activity, mood fluctuations, and cognitive responses, pose substantial obstacles to the generalization of models trained on source domain data when applied to target domains. As a result, to enhance the accuracy and robustness of cross-subject tasks, some of the research is directed towards employing advanced techniques such as domain adaptation [34], [35] and domain generalization [36], which aim to mitigate the domain shift between source and target data distributions, improving the accuracy and robustness of cross-subject models. For instance, Zhang et al. [36] propose a novel EEG-based Graph Neural Network for depression detection. In their proposal, domain generalization based on adversarial training is adapted to the model, and a secondary subject partitioning method is proposed to group subjects with similar data distributions into the same domain with a shared domain label. Fang et al. [37] present an unsupervised cross-domain fMRI adaptation framework (UFA-Net), addressing heterogeneity in major depressive disorder detection. This approach leverages graph convolutional networks and maximum mean discrepancy for feature alignment, improving model generalizability across domains without target labels. Chen et al. [35] introduce a Graph Neural Network-based Semi-supervised Domain Adaptation (GNN-SDA) framework for major depressive disorder detection. The GNN-SDA addresses semantic misalignment and class imbalance issues by aligning domains at an individual level and optimizing pseudo-labels, enhancing the model's ability to generalize across different datasets without relying on target labels. Zhang et al. [38] introduce a multi-classification model, DepL-GCN, that employs a graph convolutional network (GCN) and domain generalization to classify depression levels from EEG data. This model addresses subject variability and sample imbalance across categories by integrating a penalty coefficient for smaller classes, enhancing its ability to generalize and accurately detect varying degrees of depression without relying on target domain labels.

As for our design, we adopt targeted cross-domain techniques tailored to the distinct characteristics of the two task modules: depression recognition and depressive episode monitoring. The first phase focuses on depression recognition. Given the absence of labeled data in the target domain, we employ an unsupervised domain adaptation module, UDA-DR, which transfers knowledge from the source domain while effectively minimizing the distribution gap without relying on target domain labels. In the second phase, a small amount of labeled data is available due to the subjects diagnosed with depression. To maximize the utility of this data, we propose the FSL-DEM module based on few-shot learning, which enables rapid adaptation to new individuals and supports personalized monitoring of depressive episodes. By combining unsupervised and few-shot learning, *DepGuard* enhances model generalization and adapts to varying levels of data availability. This dual approach is a key advantage of the *DepGuard* system in addressing cross-individual depression detection challenges.

III. System Design

In this section, we introduce *DepGuard* system for recognizing depression and monitoring depressive episodes. As shown in Fig. 1, the system pipeline begins with the collection of multimodal physiological signal data via a wrist-worn device, which is transmitted to software for specialized preprocessing. The processed data is then used for initial-stage depression detection. If depression is detected, the system proceeds to monitor depressive episodes in the second stage. Otherwise, the results are directly displayed. The depressive episode monitoring task involves real-time analysis of the multimodal physiological signals collected from the patient, providing an assessment of whether the patient is currently in a depressive state.

A. Data Preprocessing

Due to the prohibitive costs of off-the-shelf commercial devices for collecting physiological signals and the lack of available data interfaces, we develop our hardware and software solutions to facilitate data collection and other tasks.

Authorized licensed use limited to: SHENZHEN UNIVERSITY. Downloaded on July 29,2025 at 15:21:18 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2025.3591096

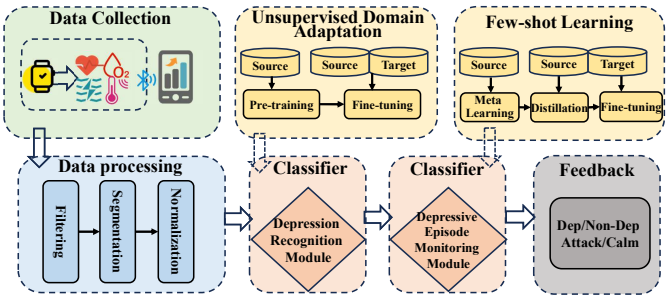


Fig. 1. The overview of the data-processing pipeline

Heart rate signal: To mitigate PPG signal drift caused by varying ambient light, we apply a Detrended Fluctuation Analysis (DFA) with a time window length of 20 to remove trend components. The PPG signal comprises two main frequency bands: low-frequency components (0.05 to 0.5 Hz), which reflect sympathetic nerve activity, fluid regulation, and changes in respiration and blood pressure; and high-frequency components (0.5 to 4 Hz), which are primarily associated with parasympathetic nerve activity and short-term heart rate variability (HRV). To remove noise and other interferences, we apply a Butterworth band-pass filter with a frequency range of 0.05 to 4 Hz.

Blood oxygen saturation signal: On the collected infrared and red light signal data, we first apply a Short-time Fourier Transform (STFT) to analyze the signal's frequency components. Then, we use a Butterworth band-pass filter ($0.1 \sim 4$ Hz) to remove noise and improve signal quality.

Galvanic skin response signal: GSR signals primarily show low-frequency conductivity changes. To focus on physiological arousal and sweat gland activity, we apply STFT and a second-order Butterworth filter with a cutoff frequency set at 0.3 Hz.

Skin temperature signal: The trend in skin temperature signal (SKT) is relatively subtle. We apply an STFT and find that the primary components are predominantly within the effective frequency range. Additionally, power frequency interference can be introduced into the sensor signal, affecting the accuracy of the skin temperature sensor. To mitigate power frequency interference, we use a Butterworth filter to truncate the frequency of the skin temperature signal.

After processing the physiological signals, we use sliding processing with a window size of 1 and a stride size of 1. And then, Eq. (1) is used to normalize each data channel to a range of -1 to 1.

$$y_{norm} = 2\frac{x - x_{min}}{x_{max} - x_{min}} - 1 \tag{1}$$

B. Unsupervised Domain Adaptation for Depression Recognition

Poor performance in cross-subject depression recognition is primarily attributed to the lack of data from new subjects during the training phase, which fails to capture the heterogeneity of individual physiological signals. This diversity is crucial for accurate model training. Without incorporating this range of individual differences, the model struggles to adapt to the unique physiological patterns of new subjects, thus reducing its effectiveness in recognizing depression across individuals. Additionally, depression is a complex mental illness typically diagnosed through clinical evaluations, making the acquisition of labeled data challenging. The nature of depression recognition prevents us from acquiring labeled target-domain (*i.e.*, target subject) data in the fine-tuning stage, leaving us to depend solely on unlabeled data. Therefore, we propose an unsupervised domain adaptation for depression recognition (UDA-DR) to enhance cross-subject performance. Fig. 2 shows the structure of UDA-DR, which adopts a two-phase training process of pre-training followed by fine-tuning. The method primarily consists of three modules: encoder, class classifier, and domain classifier. The encoder ensures that the class classifier performs well during pre-training and that the domain classifier cannot distinguish between source and target domain data during fine-tuning. The class classifier aims to correctly categorize the data based on features extracted by the encoder, while the domain classifier distinguishes whether the data originates from the source or target domain.

1) Model building: By processing data from five channels, we extract time-domain feature samples. Fig. 2 presents a multi-channel temporal network based on LSTM, CNN and attention mechanism (LCMTN), serving as the backbone network for Sec. III-B and Sec. III-C. The LCMTN mainly consists of an encoder and classifier, where the classifier uses a Multilayer Perceptron (MLP). We first apply a Long Short-Term Memory (LSTM) network for the feature encoder to extract local temporal features from each channel.

Subsequently, to obtain the coupling of temporal dependencies, we deploy a multi-head self-attention mechanism [39] in LCMTN to enhance the derived representation of temporal features. The multi-head self-attention mechanism enables the model to simultaneously attend to different segments of the input sequence, capturing diverse temporal dependencies and enhancing the global representation of physiological signals over time. Concretely, as shown in Eq. (2), given a feature representation $S \in \mathcal{R}^{L\times C}$ fed into linear layers to generate query space, key space and value space, represents learnable weighted matrices, L and C denote the time series and the number of channels, respectively. Next, one head of attention output is formulated as shown in Eq. (3), where the softmax function is applied to the scaled dot product of the query and key matrices to produce attention scores. These scores are then used to weigh the values in the value matrix $V(S)$, resulting in an attention head output that captures the importance of different temporal components across the sequence. In Eq. (4), the output of multi-head self-attention $MHA(S)$ is a repeated process for multiple heads to jointly enable the model to focus on various aspects of the input sequence, where $W^O \in R^{C\times C}$ denotes a learnable weighted matrix, $Concat$ is a concatenated operation, and k is the number of attention heads.

$$Q(S) = S \cdot W_Q, K(S) = S \cdot W_K, V(S) = S \cdot W_V \tag{2}$$

$$h_j = \text{Softmax}\left(\frac{Q(S)K^T(S)}{\sqrt{d_K}}\right)V(S) \tag{3}$$

Authorized licensed use limited to: SHENZHEN UNIVERSITY. Downloaded on July 29,2025 at 15:21:18 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2025.3591096

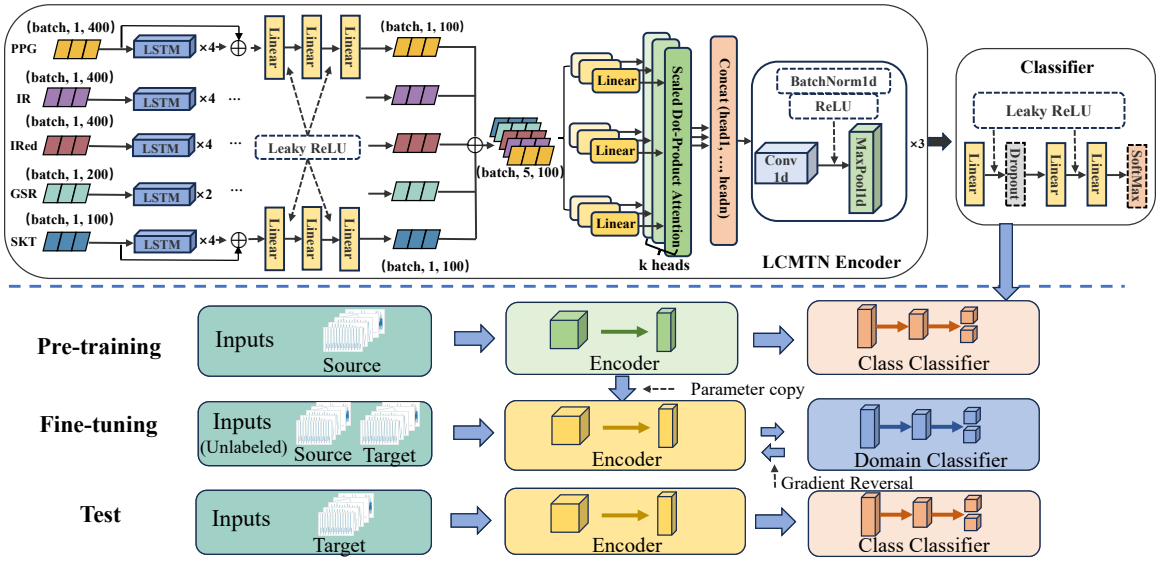


Fig. 2. The architecture of UDA-DR network

$$\mathrm{MHA}(\mathrm{Q(S)}, \mathrm{K(S)}, \mathrm{V(S)}) = \mathrm{Concat}(h_1, h_2, ..., h_k)W^O \quad (4)$$

Finally, we need to integrate features from various channels effectively, ensuring they complement each other for more comprehensive feature information. Thus, we employ a Convolutional Neural Network to precisely capture the correlations between different channels, providing a richer feature representation for the fusion process.

2) Pre-training phase: The primary objective of this phase is to train LCMTN on large-scale datasets, aiming to obtain a high-performance encoder and class classifier for the task of depression recognition. In the pre-training phase, we only utilize the data of source domain subjects for supervised training. During the training of the network, the data of source domain subjects is partitioned into a training set, validation set and test set. The model parameters are updated through backpropagation using the cross-entropy loss.

3) Fine-tuning phase: This phase is crucial for enhancing cross-subject performance in depression recognition. During Phase I, we train a high-performance encoder and class classifier optimized for depression recognition. The class classifier demonstrates strong classification capabilities for this task. However, the encoder is not exposed to the data features of the target subject, resulting in poor performance when used directly by the target subject. To address this, we employ unsupervised domain adaptation, utilizing unlabeled data from both source and target domain subjects to fine-tune the encoder. By confusing source and target domain features, the encoder learns domain-invariant features, enabling the class classifier to perform more effectively on target subject data.

Specifically, we mark the source domain data as one category of labels and the target domain data as another category of labels before starting fine-tuning. Then, we introduce a domain classifier using an MLP and integrate it with the encoder obtained in phase I to form an integrated network. The primary task of this network is to determine whether the input data originates from the source or target domain. A Gradient Reversal Layer (GRL) is placed between the encoder and domain classifier to align the encoder with the target domain data. The GRL inverts the gradient flow during backpropagation, effectively confusing the distinction between source and target domains. This design mitigates domain differences, improving the encoder's adaptability to the target subject data. During the forward pass, the network applies an identity transformation (Eq. (5)), while in the backward pass, it reverses the gradient direction (Eq. (6)). Here, I denotes the identity matrix, $-I$ indicates reversed gradient flow and λ is a hyperparameter that controls the gradient magnitude during backpropagation.

$$R_\lambda(x) = x \quad (5)$$

$$\frac{dR_\lambda(x)}{dx} = -I \quad (6)$$

At the beginning of the fine-tuning, we input both the source domain and target domain data into the network at a certain ratio in each iteration. When passing through the GRL, it undergoes an identity transformation as usual. The loss function of the entire network is then formulated as:

$$Loss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log(p_{i,c}) - \frac{1}{M}\sum_{j=1}^{M}\sum_{d=1}^{D} z_{j,d}\log(q_{j,d}) \quad (7)$$

where the first term $-\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log(p_{i,c})$ denotes domain classifier loss L_{source} and the second term $\frac{1}{M}\sum_{j=1}^{M}\sum_{d=1}^{D} z_{j,d}\log(q_{j,d})$ denotes the target domain loss L_{target}. N and M denote the number of source domain samples and target domain samples, respectively. C represents the number of classes (depression or non-depression). D represents the number of domains (source domain or target domain). $y_{i,c}$ and $p_{i,c}$ are the true domain label indicator and predicted probability of sample i belonging to class c, respectively. $z_{j,d}$ and $q_{j,d}$ are the true domain label indicator and predicted probability of the sample j belonging to domain

Authorized licensed use limited to: SHENZHEN UNIVERSITY. Downloaded on July 29,2025 at 15:21:18 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2025.3591096

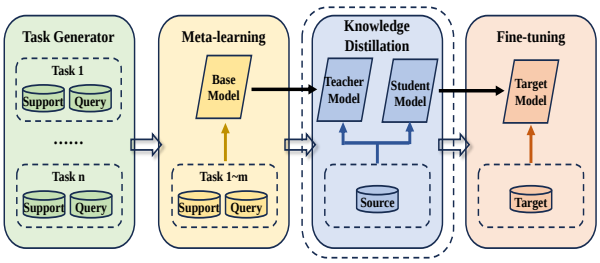


Fig. 3. The architecture of FSL-DEM

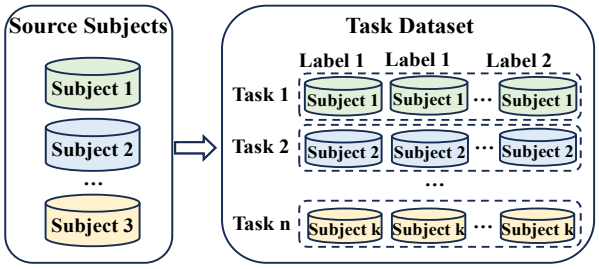


Fig. 4. Subject-based task generator

d, respectively. During backpropagation, the GRL causes the parameter gradients of the encoder to be reversed. The gradient update formulas for the domain classifier and the encoder are as follows:

$$\theta_y = \theta_y - \mu \frac{\partial L_y}{\partial \theta_y} \tag{8}$$

$$\theta_f = \theta_f - (-\mu \frac{\partial L_y}{\partial \theta_f}) \tag{9}$$

Here, θ_y represents the parameters of the domain classifier and θ_f represents the parameters of the encoder. Eqs. (8) and (9) show that θ_y updates normally, but θ_f gradients are reversed.

4) Testing phase: After pre-training and fine-tuning phases, the classifier and the encoder are obtained. We combine these to form a personalized depression recognition model for a new subject. This approach not only adapts well to individual differences but also facilitates the creation of personalized models that accurately fit new subjects with minimal data, thereby promoting personalized depression detection.

C. Few-shot Learning for Depressive Episode Monitoring

Depressive episode monitoring also presents a key challenge: individual differences among subjects reduce cross-subject performance. To address this, strategies such as unsupervised domain adaptation and few-shot learning are commonly used. Although unsupervised domain adaptation is flexible for handling distribution differences between source and target domains, it may be less effective than few-shot learning methods in some cases. In contrast, by leveraging the limited labeled samples in the target domain, few-shot learning can directly guide the model to learn specific patterns in the target domain. Meanwhile, the monitoring task differs from depression recognition in that it can yield limited labeled target domain data for enhancing cross-domain performance. To this end, we propose a few-shot learning method for depressive episode monitoring (FSL-DEM), as shown in Fig. 3. The FSL-DEM consists of four modules: task generator, meta-learning, knowledge distillation, and fine-tuning.

1) Task generator: Given that this task employs the meta-learning method in few-shot learning, it is worth noting that the basic unit of meta-learning training is the meta-task. The primary goal of the task generator is to produce representative and well-defined meta-tasks, ensuring excellent model performance during training. A common strategy for generating meta-tasks involves random selection. This approach aggregates data from all source domain subjects, shuffles it, and then selects a subset to form each meta-task. It is important to note that although we use data from source domain subjects in meta-learning, there still are significant differences in the data of the same categories among different subjects. In fact, the data from each individual can be considered as a unique classification task dataset. This means that in the source domain data, each individual has its unique characteristic and pattern, forming multiple independent classification tasks. As shown in Fig. 4, we treat each source domain subject as an independent dataset and randomly select a certain number of data samples for each category to form a meta-task. In this manner, each source domain subject generates multiple meta-tasks, with an equal number of samples for each category within each meta-task. Meanwhile, we further divide each meta-task into a support set and a query set. The support set is used to train a temporary model for the corresponding meta-task, while the query set is used with the temporary model to compute the loss for updating the base model.

2) Meta learning: The goal of this phase is to obtain well-initialized parameters so that the model can quickly adjust its parameters when facing new subjects, enabling it to adapt more quickly and effectively. Each meta-task consists of a support set and a query set, of which the support set comprises only a small portion. Additionally, the number of samples for each category remains consistent within both the support set and the query set. Fig. 5 illustrates the meta-learning training process, which can be divided into two stages: the inner training stage and the outer training stage.

For the inner training stage, we extract a subset of meta-tasks as an inner training batch and only use the support sets of sampled tasks for model training. At the start of training, each meta-task independently copies the base model to use as its temporary model, allowing the temporary model to fit some extent on the support set of each task. Subsequently, each meta-task updates its temporary model parameters using the cross-entropy loss function. Each meta-task has an independent loss calculation process that does not interfere with each other, and its parameter gradient update formula is:

$$\theta_i^{'} = \theta - \alpha \nabla_\theta L_{T_i}(f_\theta) \tag{10}$$

where θ represents the model parameters, α is the learning rate during inner training, $L_{T_i}(f_\theta)$ is the loss function on task T_i, and $\nabla_\theta L_{T_i}(f_\theta)$ is the gradient of the loss function with respect to the model parameters. In the gradient update, second-order derivatives are generally involved. However, the calculation of the second-order derivatives increases the computational complexity and cost. Therefore, to reduce computational cost, the second-order partial derivative in the formula is approximated to zero, and only the first-order derivative is calculated in practical applications.

Authorized licensed use limited to: SHENZHEN UNIVERSITY. Downloaded on July 29,2025 at 15:21:18 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2025.3591096

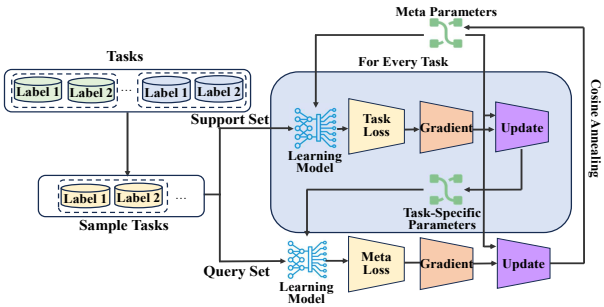


Fig. 5. The Meta-learning training process

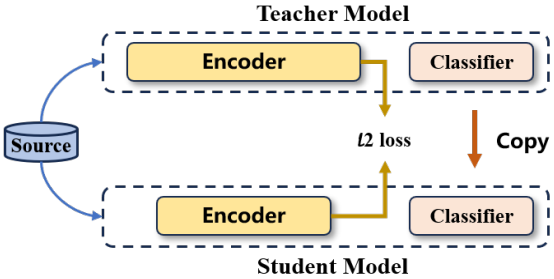


Fig. 6. Feature-based knowledge distillation

The outer training stage is primarily evaluated using a query set of specific tasks and updates the base model parameters. In the inner training stage, we obtain the updated temporary model parameters corresponding to each meta-task. We use the query set of each meta-task for forward propagation in the corresponding temporary model, obtaining the loss for each meta-task. As shown in Eq. (11), we sum up these loss to obtain a total loss, which is used to update the parameters of the base model. Subsequently, cosine annealing is employed to continuously adjust the learning rate, which is used to update the parameters of the base model during the outer training stage. Since the goal of meta-learning is to find well-initialized parameters that enable rapid adaptation to new tasks, it is essential to dynamically adjust the learning rate during training to ensure the model converges quickly on different tasks. Cosine annealing precisely offers this dynamic adjustment mechanism, helping the model better escape local optima during the learning process and improving generalization ability. As shown in Eq. (12), cosine annealing reduces the learning rate by using a cosine function.

$$L_{total} = \sum_{i=1}^{m} L_{T_i}(f_\theta) \tag{11}$$

$$\beta_t = \beta_{min} + \frac{1}{2}(\beta_{max} - \beta_{min})(1 + \cos(\frac{T_{cur}}{T_{max}}\pi)) \tag{12}$$

where β_t is the learning rate at step t, β_{max} and β_{min} are the maximum and minimum values of the learning rate, defining the range of the learning rate. T_{cur} refers to the current iteration count, and T_{max} refers to the total number of iterations within one cycle. This formula is designed to make the learning rate decrease slowly at the beginning of a cycle, then drop rapidly, and finally decrease slowly again.

3) Knowledge distillation: This phase of knowledge distillation serves as an optional module. The task of depressive episode monitoring requires real-time performance. If this monitoring module is embedded in devices with limited hardware resources, the knowledge distillation module becomes particularly important. Generally, there are three methods of knowledge distillation: feature-based knowledge distillation, response-based knowledge distillation, and relation-based knowledge distillation. When knowledge distillation targets the encoder, the smaller model inherits the representational capacity of the larger model while avoiding overfitting. In contrast, distilling the classifier may cause the student model to overfit to the teacher's predictive behavior, potentially disregarding the true data distribution. To address this, we adopt a feature-based knowledge distillation strategy. As illustrated in Fig. 6, we use the base model trained with meta-learning as the teacher, and a redesigned, lightweight encoder with reduced complexity and fewer parameters as the student. The input data still comes from the source domain, and the loss between the feature vectors output by the encoder of the teacher model and the encoder of the student model is calculated using Eq. (13). Subsequently, only the parameters of the encoder in the student model are updated using the obtained loss. The encoder of the trained student model is combined with the classifier of the teacher model to form a new student model.

$$Loss = \frac{1}{n}\sum_{1}^{n}(f_{teacher}(x_i) - f_{student}(x_i))^2 \tag{13}$$

where $f_{teacher}(x_i)$ is the feature vector output by the encoder of teacher model, $f_{student}(x_i)$ is the feature vector output by the encoder of student model, and n is the length of the feature vectors. The formula calculates the $\ell 2$ loss between $f_{teacher}(x_i)$ and $f_{student}(x_i)$. When updating the encoder parameters of the student model through this method, the focus is on minimizing the difference in feature representations.

In the knowledge distillation module, we introduce a parameter τ to control the degree of distillation. This parameter determines the extent to which knowledge from the teacher model is transferred to the student model. A higher τ signifies that the student model prioritizes the integration of knowledge distilled from the teacher model, whereas a lower value conversely leads the student model to rely more heavily on the original training dataset for parameter updates. The parameter τ is systematically optimized in Sec. V-F3 to balance robust generalization capability of the student model with mitigation of over-reliance on the teacher's supervisory signals.

4) Fine-tuning phase: After meta-learning or knowledge distillation training, the base model has acquired well-initialized parameters. In this phase, the new subject provides a small amount of data to fine-tune the base model, further adapting it to the characteristics of the target domain data. We divide the fine-tuning process into two stages based on the number of iterations. In the first stage, we update the model parameters with a high learning rate, but the model has not yet reached its best convergence position. In the second stage, a small learning rate is used to update the model parameters to further optimize the model towards the optimal solution and take into account the stability of optimization. In this way, each new subject obtains a customized model tailored to individual

Authorized licensed use limited to: SHENZHEN UNIVERSITY. Downloaded on July 29,2025 at 15:21:18 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2025.3591096

TABLE I
SPECIFICATIONS OF DIFFERENT SENSORS

Signal type	Sensor model	Sampling rate	Data type	Price
Heart rate	Pulse sensor	400 Hz	Raw analog data	3.425 $
Blood oxygen saturation	Max30102	400 Hz	Red light and infrared light intensity data	0.959 $
Galvanic skin response	Grove GSR	200 Hz	Voltage simulation data	8.219 $
Skin temperature	LMT70	100 Hz	Voltage simulation data	3.425 $

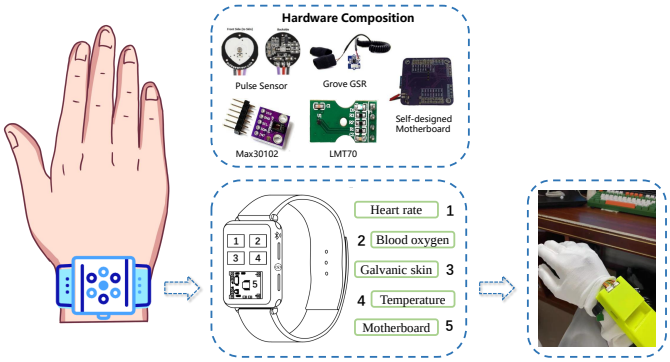


Fig. 7. The design of our wrist-worn device

characteristics, thereby better adapting to their personalized needs in the depressive episode monitoring task.

IV. SYSTEM IMPLEMENTATION

A. Hardware Design

The hardware part of the *DepGuard* is a wrist-worn device, including four parts: control module, sensor module, communication module, and power module. The control module utilizes STM32, which offers precise data processing and maintains stable performance in low-power mode. The sensor module integrates four ubiquitous biometric sensors for collecting data such as heart rate, blood oxygen saturation, galvanic skin response, and skin temperature.

Table I presents the sensors and their parameters to ensure efficient and accurate acquisition of various environmental data by the *DepGuard*. The communication module uses the Bluetooth BLE5.0 version protocol. The power module employs the SY8089DC-DC buck converter and a TP4054 charging circuit for a maximum 500 mA charge, effectively managing power for the wristband device.

Fig. 7 presents the overall hardware architecture of *DepGuard*. The system comprises a custom-designed motherboard integrated with four ubiquitous biometric sensors, forming a compact, wrist-worn device. The onboard microprocessor collects physiological signals and transmits the data to the companion app via Bluetooth.

B. Software Development

We develop a mobile app for smartphones to complement the *DepGuard* hardware, offering features like physiological monitoring, self-assessment, and depression tracking, making up the full *DepGuard* system. Fig. 8(a)-(e) shows the main interface of the app, highlighting its three core functions. Fig. 8(a) shows the main interface for monitoring physiological data with Bluetooth and real-time display. The app connects to the hardware via Bluetooth, receives real-time data and stores it for later analysis. When the subjects want to know more about specific physiological data, they can click the icon on the interface to access the corresponding display page. Fig. 8(b) and Fig. 8(c) visualize the heart rate and blood oxygen saturation, plotting their trends and displaying live data. Fig. 8(d) shows the self-assessment interface with personal info and questionnaires including PHQ-9, GAD-7, BDI-II and so on. Users can complete these questionnaires, and their scores, along with severity information for some scales, will be shown on their personal page. Fig. 8(e) is the depressive episode recording interface, which aims to collect the time periods when individuals with depression experience the symptoms described in their medical records. It is important to note that the functionality of this page is only used when we are collecting data on depressive episodes of patients.

The aforementioned interface is integrated into a mobile app and tested on the Huawei Mate 40 Pro smartphone, powered by HiSilicon Kirin 9000 chipset, 4400 mAh battery, and 8 GB RAM. Integrating software and hardware enables *DepGuard* to systematically record the subject's physiological and psychological conditions.

V. EXPERIMENT AND EVALUATION

A. Data Collection

We have undertaken a 3-month data collection and preprocessing process to facilitate the execution of this study. The experimental participants consist of two distinct groups: individuals without depression and those with depression. Before the experiment, all participants agree to sign the Institutional Review Board (IRB) agreement [40]. The data collection process for these two groups diverges; For the group without depression, potential participants are recruited from Shenzhen University. Each participant fills out PHQ-9, GAD-7, BDI-II, and other questionnaires on the software before the experiment. Afterward, participants are selected based on questionnaire scores and criteria to confirm they are "healthy." Then, we choose a quiet conference room for data collection from non-depressed individuals, keeping them free from distractions. Before starting, participants should get used to the *DepGuard* system to avoid any impact on their physical or mental condition. Subsequently, they are given a 1-minute buffer to regulate their emotions to ensure reaching a calm state. On the other side, we collect physiological data from clinically diagnosed depressed patients in Guangdong Second Provincial General Hospital to ensure data authenticity. Patients are informed about the study and sign consent forms. They learn about the *DepGuard* system and are briefed on collection precautions.

Depressive episode data are collected only for patients with diagnosed depression. In the formal phase of data collection, we collect physiological data when the patient is in a state

Authorized licensed use limited to: SHENZHEN UNIVERSITY. Downloaded on July 29,2025 at 15:21:18 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2025.3591096

(a) Main interface (b) Heart rate signal (c) Blood oxygen signal (d) Personal information (e) Episode recording

Fig. 8. The main interfaces of our mobile application

of both physical and mental tranquility. When the patient experiences an episode, they click the 'Attack' button on the corresponding interface of the software. When the patient returns to a calm state from the episode, they click the 'Calm' button. In this way, the segment of the depressive episode is recorded. Notably, the process of collecting data on depressive episodes described above is used only for the training phase of the model. In real scenarios, the *DepGuard* system collects physiological signals in real-time without clicking on the 'Attack' or 'Calm' button. Besides, during all of the above data collection procedures, we ask the subjects to remain in a resting state in order to prevent body movement artifacts from being generated. Data on depressive episodes is especially collected in the presence of a healthcare professional. Finally, we recruit 35 participants (20 males, 15 females) with 17 individuals from the healthy control group and 18 patients diagnosed with depression. The ages of the participants range from 14 to 53 years old, with the mean and standard deviation of 23.14 and 7.06, respectively. As for their occupations, twenty-eight participants are still in education, four are currently workers, technicians, or self-employed, and one is retired. The remaining participant does not provide occupational information.

The length of data for subjects with depression recognition and depressive episode detection is inconsistent, ranging from 594 to 7444 seconds. Concretely, the total amount of data for depression recognition is 42206 seconds. And the total amount of data for the depressive episodes is 38659 seconds, of which 48.7% (18839 seconds) are depressive samples.

B. Evaluation Metrics

For depression recognition, we assess performance using accuracy, detection rate, and misdiagnosis rate. Accuracy, as represented in Equation 14, is the ratio of correctly predicted samples to the total number of samples. Detection rate, as represented in Equation 15, is the ratio of correctly predicted depression samples to the total number of depression samples. Misdiagnosis rate, as represented in Equation 16, is the ratio of incorrectly predicted non-depression samples to the total number of non-depression samples.

$$Accuracy = \frac{n_{pre}}{n_{true}} \cdot 100\% \tag{14}$$

$$Detection\ rate = \frac{n_{pre=dep}}{n_{true=dep}} \cdot 100\% \tag{15}$$

$$Misdiagnosis\ rate = (1 - \frac{n_{pre=non-dep}}{n_{true=non-dep}}) \cdot 100\% \tag{16}$$

For depressive episode monitoring, we use precision, recall, F1 score, and accuracy as performance evaluation metrics. Among them, precision, recall, and F1 score are all calculated using the macro-average method, which means taking the average of evaluation metrics for each category.

C. Details of Evaluation

In both tasks, the single-subject setting involves randomly partitioning each subject's data into training (60%), validation (20%), and testing (20%) sets. For cross-subject evaluation, we employ Leave-One-Subject-Out Cross-Validation (LOSO-CV), a specific form of K-Fold cross-validation where K equals the number of subjects ($K = 35$ for depression recognition; $K = 17$ for depressive episode monitoring). In each iteration, one subject is designated as the test set, while the remaining $K - 1$ subjects constitute the training set. This exhaustive approach ensures that each subject is used once as the test set, providing a comprehensive assessment of the model's generalization across individuals.

All evaluations are implemented with the environment of Intel(R) Xeon(R) CPU E5-2686 v4 @ 2.30GHz, NVIDIA Tesla A6000 GPU. We harness the computational prowess of the cuDNN library to expedite GPU-accelerated processing.

Authorized licensed use limited to: SHENZHEN UNIVERSITY. Downloaded on July 29,2025 at 15:21:18 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2025.3591096

The training process involves the Adam optimizer [41], with a learning rate of $1e-4$, a batch size of 128, and 200 epochs to ensure robust convergence. The attention head is set to 4 in the MHA module. Random dropping ratio of 0.3 in the Dropout layer.

D. Baseline Methods

We compare the performance of *DepGuard* with the following baseline methods.

- **DANN [42]** is a classic unsupervised domain adaptation method based on adversarial learning. Its main idea is to reduce the distribution discrepancy between the source domain and the target domain by introducing domain adversarial training, thereby improving the model's generalization ability on the target domain.
- **ADDA [43]** is an unsupervised domain adaptation method that combines GAN loss. It employs a different training strategy with DANN to reduce the distribution discrepancy between the source domain and the target domain, thus improving the model's generalization performance.
- **DFA-MCD [44]** is a state-of-the-art unsupervised domain adaptation method that introduces a Gaussian-guided latent alignment approach to align the latent feature distributions of the two domains under the guidance of the prior distribution.
- **SWL-Adapt [45]** is an unsupervised domain adaptation method tailored for temporal input features, incorporating sample weight learning. It calculates sample weights according to the classification loss and domain discrimination loss of each sample with a parameterized network.
- **CDFSL-V [46]** is an effective few-shot learning method for recognizing new categories with only a few labeled examples. The method employs a masked autoencoder-based self-supervised training objective to learn from both source and target data in a self-supervised manner. Then, a progressive curriculum balances learning the discriminative information from the source dataset with the generic information learned from the target domain.
- **PN [47]** is a state-of-the-art few-shot learning method for the image dataset CUB-200-2011, where a classifier must generalize to new classes not seen in the training set, given only a small number of examples of each new class. It learns a metric space in which classification can be performed by computing distances to prototype representations of each class.
- **PN(-T)** is an ablation method for the PN based on a random task generation strategy.
- **FD(-A)** is an ablation method for the FSL-DEM based on using only easily obtainable non-episode samples.
- **FD(-F)** is an ablation method for the FSL-DEM based on not using fine-tuning module and employing meta-learning training.
- **FD(-T)** is an ablation method for the FSL-DEM based on random task generation strategy.

E. Depression Recognition Results

1) Performance of Cross-subject: We evaluate cross-subject depression recognition using LCMTN and UDA-DR, as depicted in Fig. 9. where U1 to U17 denote non-depressed subjects, and U18 - U35 denote depressed subjects. The overall average accuracy of LCMTN is 68.85%, while that of UDA-DR is 90.75%. UDA-DR enhances cross-subject recognition accuracy by 21.9% over LCMTN. Notably, the accuracy for the non-depression subject U6 and the depression subject U35 decreased. For U6, the depression scale scores are near the threshold for diagnosis. And for U35, the professional psychologist diagnoses fewer and milder symptoms. Moreover, Fig. 10 compares the cross-subject recognition performance of DANN [42], ADDA [43], DFA-MCD [44], SWL-Adapt [45], and UDA-DR under various shot conditions. UDA-DR excels in the 15-shot and 30-shot target domain scenarios due to its effectiveness with limited unlabeled data.

2) Analysis of Performance Influencing Factors: Furthermore, we test four key factors for optimizing the cross-subject recognition performance of UDA-DR: the number of target subject data samples and the number of fine-tuning iterations, the number of source domain subjects, and the batch proportion size between the source domain and the target domain.

The number of target subject data samples: Fig. 11 displays the metrics performance for target subject unlabeled data across 5, 10, 20, 30, and 60 shots. Accuracy increases with the increase in the amount of unlabeled data provided to the target subjects. With 5 shots, the accuracy is 80.89%, and the accuracy remains stable and unchanged after 30 shots. Although the detection rate at 30 and 60 shots is slightly lower than at 20 shots, the misdiagnosis rate is significantly better than at 20 shots. Thus, we set a minimum of 30 shots of unlabeled data per target subject.

The number of fine-tuning iterations: In the fine-tuning phase, we fine-tune the target subject data samples of 30 shots with 1, 3, 5, 10, 15, 20, and 25 iterations respectively. Fig. 12 shows the accuracy, detection rate, and misdiagnosis rate results for different fine-tuning numbers. With the continuous increase in the number of fine-tuning, the accuracy shows a gradual upward trend. When fine-tuning is performed 20 times, the accuracy reaches its optimal state. However, further increasing the number of fine-tuning led to a downward trend in accuracy performance. This is because the increase in the number of fine-tuning causes the model to overlearn the specific local structure in the new subject data while ignoring the overall pattern of the data. Although the detection rate of fine-tuning 20 times is slightly lower than that of fine-tuning 25 times, the misdiagnosis rate performance is significantly better than fine-tuning 25 times. Therefore, we fixed the number of fine-tunings to 20 times.

The number of source-domain subjects: The diversity of subjects is closely related to the generalization performance of the model. In the pre-training phase, apart from each individual subject, we randomly select 3, 5, 7, 10, 15 subjects respectively from the remaining subjects in each category as source domain subjects. As shown in Fig. 14, as the number of source domain

Authorized licensed use limited to: SHENZHEN UNIVERSITY. Downloaded on July 29,2025 at 15:21:18 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2025.3591096

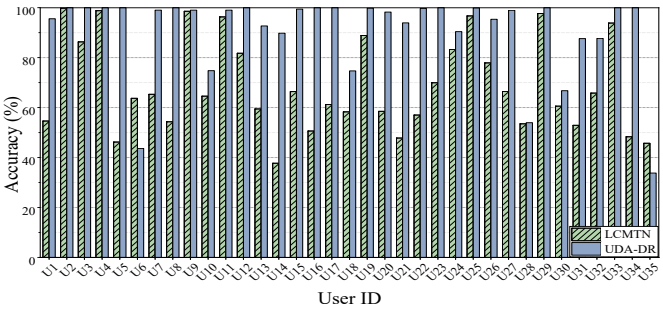


Fig. 9. The cross-subject results of depression subjects

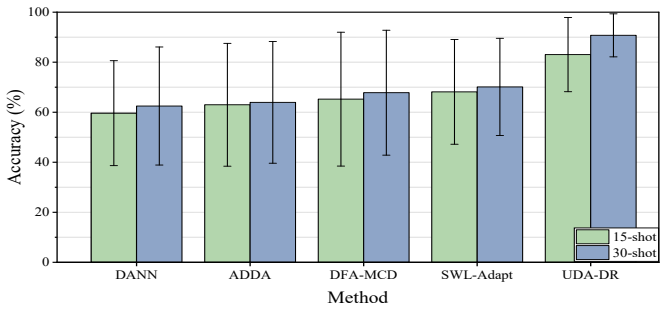


Fig. 10. The results of cross-subject for different methods

subjects is continuously increased, the accuracy is gradually improving, and the misdiagnosis rate keeps dropping. When three source domain subjects are provided for each category to all source domain subjects except the target subject, the recognition accuracy increases from 66.4% to 90.75%, and its performance gradually nears the single-subject performance under the LCMTN. It further indicates that although subject diversity is related to model generalization performance, the number of source domain subjects currently used in UDA-DR is already sufficient. Even if the number of source domain subjects is further increased, the space for improvement in recognition performance is limited while the data cost continues to increase.

The batch proportion across domains: In domain adaptation, the batch proportion refers to the ratio of source to target domain samples during fine-tuning, balancing the model's ability to learn discriminative features from the source domain with its adaptability to the target domain. In this phase, we set the batch size of the target domain to a fixed 20, and then respectively set the batch proportion size of the source domain to the target domain as 0.5, 1, 1.5, 2, 3, and 4. As shown in Fig. 15, as the batch proportion size increases, the misdiagnosis rate continuously decreases, approaching zero. When the batch proportion size is from 0.5 to 2, the accuracy continuously improves. This is because increasing the batch proportion size causes an excessive focus on the source domain data while neglecting the features of the target domain data. As the batch proportion size continues to increase, the accuracy begins to show a downward trend. This is because when increasing the batch proportion size, there is an excessive focus on the source domain data, while neglecting the feature of the target domain data. Therefore, we fix the target domain batch size at 20 during the fine-tuning phase while setting the source domain batch size to be twice that of the target domain batch size.

3) Comparison of Different Channel Combinations: In order to explore the impact of different channel combinations on recognition performance, we further conduct a feature ablation study. For the five channels: PPG, IR, Red, GSR, SKT, we use one-channel, two-channel, three-channel, and four-channel respectively to perform cross-subject depression recognition. Due to the certain correlation between IR and Red, only the IR channel in the blood oxygen signal is used for multi-channel combination in the feature ablation study. Fig. 13 and Fig. 16 summarize the results. For one-channel, it can be seen that different channels have varying abilities in recognizing depression. We find that the PPG channel is the most important, with an accuracy of 75.53%, while the SKT channel performs poorly, with an accuracy of only 51.61%. For two-channel, the performance of channel combinations is mostly better than that of sub-channels. Among them, the combination of PPG and SKT channels performs the best, with an accuracy of 82.14%. However, the combination of PPG and GSR channels has reduced the performance of the PPG single channel. For three-channel, the performance of channel combinations is superior to that of the sub-single channel. The combination of PPG, IR, and GSR channels performs the best, with an accuracy of 81.57%, but it does not perform as well as the combination of PPG and SKT channels. This indicates that when using two or three channels, the best choice is the two channels composed of PPG and SKT. For the four-channel combination composed of PPG, IR, GSR, and SKT or the five-channel combination with the addition of Red, the recognition performance of depression using multi-channel combinations is generally better than one-channel, two-channel, and three-channel combinations. This indicates that multi-channel fusion can obtain more comprehensive and multi-angle information, which helps in accurately recognizing depression.

4) Comparison with Related Work: We present a systematic comparison of related work in depression recognition in Table II. Several studies have explored depression detection using traditional machine learning methods [16], [17], [18], [48], [22], [27], [52], while others have adopted deep learning approaches [26], [21], [49], [50], [51]. Generally, single-subject experiments tend to yield higher accuracy than cross-subject evaluations. For instance, traditional machine learning models have achieved up to 100% accuracy in single-subject settings [16]. However, performance drops significantly in cross-subject scenarios—for example, the work [18] reports only 74.4% accuracy using SVM and recursive feature elimination on heart rate variability data. This discrepancy stems from the limitations of traditional models in effectively learning complex physiological features. While expert-crafted features can be useful, they often fail to capture the full spectrum of depression-related patterns. In contrast, the automatic feature extraction capabilities of deep learning have led to growing interest in these methods. For example, the work [50] proposes a multi-scale spatio-temporal contrastive representation framework, and the work [51] introduces a Transformer-based multimodal fusion model combined with a

Authorized licensed use limited to: SHENZHEN UNIVERSITY. Downloaded on July 29,2025 at 15:21:18 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2025.3591096

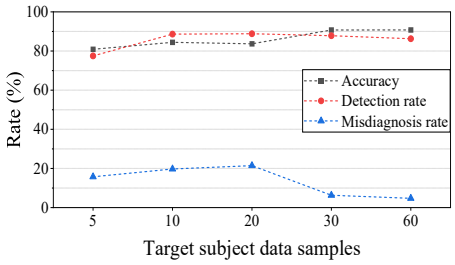


Fig. 11. The performance of different target subject data samples

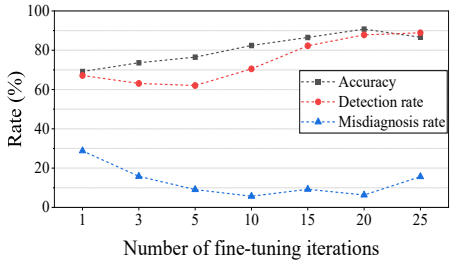


Fig. 12. The performance with the number of fine-tuning iterations

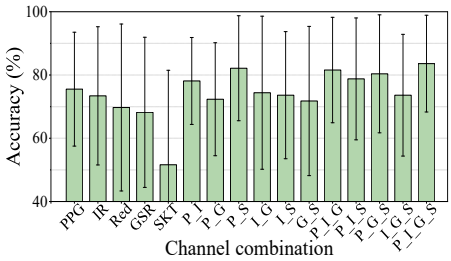


Fig. 13. The results of different channel combinations

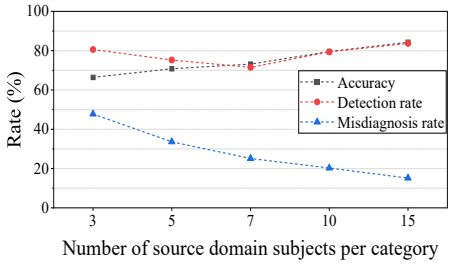


Fig. 14. The performance of different numbers of source domain subjects per category

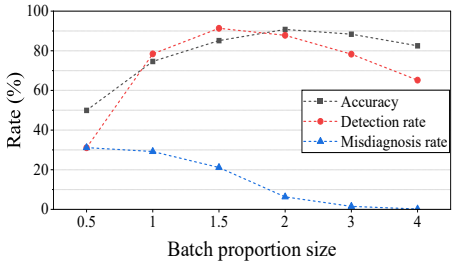


Fig. 15. The performance of different batch proportion between source and target domain

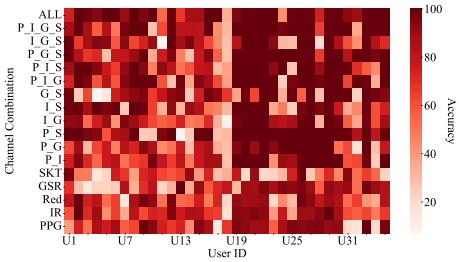


Fig. 16. The accuracy distribution results of different channel combinations for various subjects

TABLE II
THE COMPARISON WITH RELATED WORK IN DEPRESSION RECOGNITION

Relevant work	Participant (Dep/Non-dep)	Signal	Accuracy
[16]	24 / 29	EEG	single-subject 1 channel 76.08% 13 channels 83.96% 128 channels 100%
[26]	21 / 24	EEG	single-subject 99.1%
[17]	34 / 58	EEG	single-subject 89%
[18]	37 / 41	ECG	cross-subject 74.4%
[48]	7 / 16	ECG	single-subject 87%
[21]	10 / 25	GSR, SKT, IBI, IMU	single-subject 87.3%
[22]	33 / 32	PPG	single-subject 98.46%
[49]	26 / 24	EEG	single-subject EDRA dataset 98.33% MODMA dataset 97.56%
[27]	22 / 27	ECG	cross-subject 70%
[50]	24 / 29	EEG, fNIRS	single-subject 86.7%
[51]	46 / 51	EEG	cross-subject 86.84%
[52]	58 / 58	PPG, GSR, acceleration	single-subject at most 90%
UDA-DR (ours)	18 / 17	PPG, SpO2, GSR, SKT	cross-subject 90.75%

multitask learning scheme using standardized questionnaires. These methods have improved the granularity of depression representation through sophisticated architectures and multi-modal signal analysis. Nevertheless, most of these works focus primarily on signal fusion and fine-grained representation, overlooking the challenges posed by inter-subject variability. In contrast, our proposed UDA-DR addresses cross-subject generalization by employing a gradient reversal layer within an adversarial domain adaptation framework, effectively mitigating performance degradation due to individual heterogeneity.

In terms of physiological signal types, EEG-based proprietary devices remain widely used in depression recognition due to their high precision. For example, the work [22] focuses on single-modality signals, limiting its ability to comprehensively capture an individual's peripheral physiological state. In the work [52], manually extracted features are used to construct dynamic representations, followed by a random forest model for prediction. Similarly, the work [48] employs a basic deep neural network, which struggles to effectively fuse multimodal physiological signals. However, their high cost, limited accessibility, and operational complexity significantly hinder real-world application. In contrast, our approach offers superior portability, broader applicability, and enhanced performance, making it more suitable for practical deployment.

F. Depressive Episode Monitoring Results

1) Performance of Single-subject and Cross-subject: In order to verify the feasibility of depressive episode monitoring, we only conduct single-subject and cross-subject depressive episode monitoring evaluations through the LCMTN. As shown in Table III, we observe that all metrics for single-subject results are vastly higher than cross-subject results, with an accuracy difference of 43.5%. The accuracy range for single-subject results is around 99.28%, indicating a consistently high performance. For cross-subject results, the range is from 53.93% to 55.78%. These results indicate the feasibility of depressive episode monitoring. However, we need to solve the issue of subject variability to improve cross-subject performance for depressive episode monitoring.

To evaluate the cross-subject performance of FSL-DEM, we use the LCMTN and the FSL-DEM method designed in Sec. III-C. Fig. 17 shows the accuracy of LCMTN and FSL-DEM for each subject. It can be clearly seen that the performance of FSL-DEM is better than that of LCMTN across

Authorized licensed use limited to: SHENZHEN UNIVERSITY. Downloaded on July 29,2025 at 15:21:18 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2025.3591096

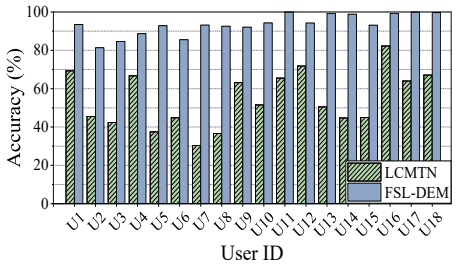


Fig. 17. The results of cross-subject for LCMTN and FSL-DEM

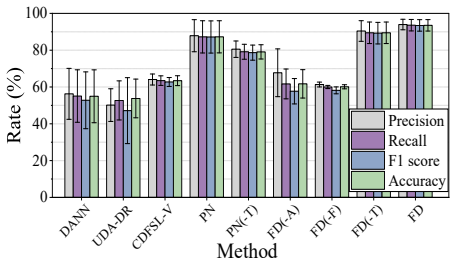


Fig. 18. The performance of cross-subject for different methods

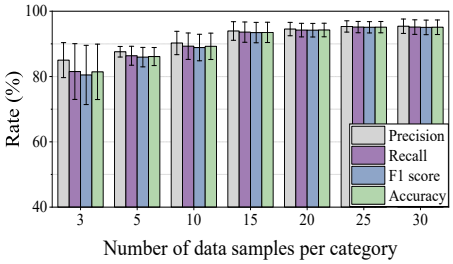


Fig. 19. The performance of different numbers of data samples per category

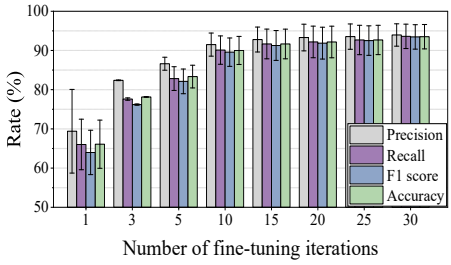


Fig. 20. The performance of different numbers of fine-tuning iterations

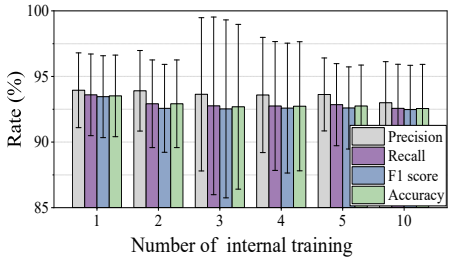


Fig. 21. The performance of different numbers of internal training

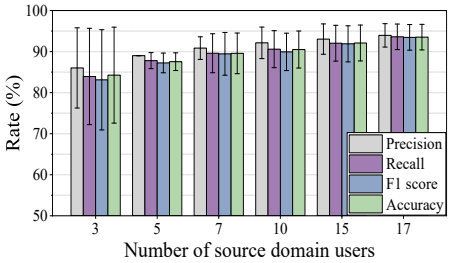


Fig. 22. The performance of different numbers of source domain subjects

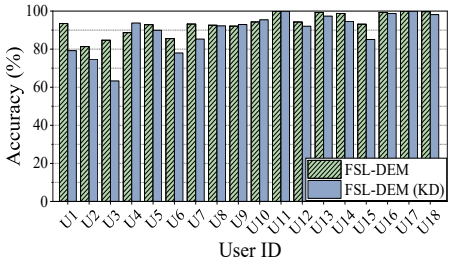


Fig. 23. The results of FSL-DEM and FSL-DEM (KD)

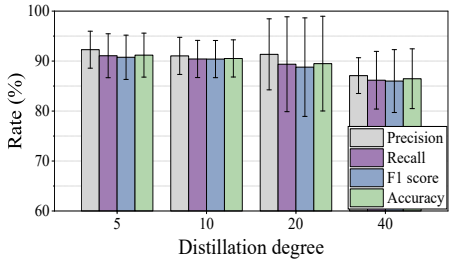


Fig. 24. The performance of different distillation degrees

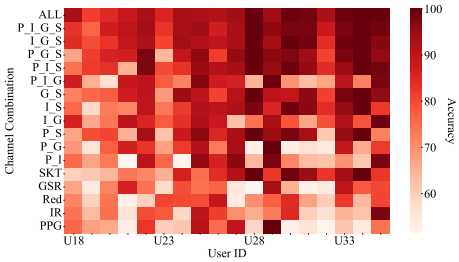


Fig. 25. The accuracy distribution results of different channel combinations for various subjects

all subjects. The average accuracy of FSL-DEM is 93.52%, representing a 39.14% increase compared to LCMTN. Among them, subject U7 shows the largest improvement, with accuracy increasing from 30.39% to 93.17%.

Additionally, Fig. 18 shows the cross-subject depressive episode monitoring results using the unsupervised domain adaptation methods DANN [42], UDA-DR, the few-shot learning methods CDFSL-V [46], PN [47], FSL-DEM, as well as ablation versions of PN and FSL-DEM. The results show that the performance of unsupervised domain adaptation has largely improved compared to using only LCMTN for cross-subject performance. The FSL-DEM exhibits the best performance compared to the other two few-shot learning methods. Furthermore, we conduct an additional ablation study on FSL-DEM. From the comparison between PN and PN(-T), as well as FD and FD(-T), utilizing a subject-based task generation strategy can enhance the cross-subject performance of FSL-DEM. Comparing FD and FD(-F), it is evident that the fine-tuning module is indispensable. The pre-training approach based on meta-learning is superior to the conventional training method. Comparing FD and FD(-A), solely utilizing readily available non-episode samples does not improve performance.

TABLE III
THE PERFORMANCE OF LCMTN FOR SINGLE-SUBJECT AND CROSS-SUBJECT.

	Single-subject	Cross-subject
Precision	99.28%	56.39%
Recall	99.28%	56.16%
F1 score	99.27%	53.93%
Accuracy	99.28%	55.78%

2) Analysis of Performance Influencing Factors: To better select appropriate hyperparameters, we also evaluate four influencing factors affecting the cross-subject performance of FSL-DEM: the number of internal training, the number of data samples per category, the number of fine-tuning iterations, and the number of source domain subjects.

The number of data samples per category. In the fine-tuning phase, we provide 3, 5, 10, 15, 20, 25, and 30 labeled samples for each category. For fine-tuning, the number of fine-tuning iterations is fixed. As shown in Fig. 19, the performance metrics of precision, recall, F1 score, and accuracy increase as the number of samples provided by the target subject increases.

Authorized licensed use limited to: SHENZHEN UNIVERSITY. Downloaded on July 29,2025 at 15:21:18 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2025.3591096

When providing 3 samples for each category, the accuracy can reach 81.43%. After providing 15 samples for each category, the performance metrics tend to stabilize. During the experimental phase of collecting data on depressive episode, it was found that each episode of patients lasted at least 15 seconds. Therefore, considering both the cost of target subject data and the performance of depressive episode monitoring, setting 15 labeled samples for each category as the minimum required data from the target subject is advisable.

The number of fine-tuning iterations. In the fine-tuning phase, we fix the number of samples provided by the target subject for each category at 15 and set fine-tuning iterations to 1, 3, 5, 10, 15, 20, 25, and 30, respectively. Observing from Fig. 20, it can be noted that the performance metrics of precision, recall, F1 score, and accuracy exhibit a rapid increase followed by a gradual stabilization as the number of fine-tuning iterations increases. After 15 fine-tuning iterations, the performance improvement is small. Specifically, the accuracy rates are 91.66%, 92.17%, 92.68%, and 93.52% for fine-tuning iterations of 15, 20, 25, and 30, respectively. Therefore, we fixed the number of fine-tunings to 20 times.

The number of internal training. In the meta-learning phase, the number of internal training not only affects the training duration but also influences the model performance. Specifically, we conduct performance evaluations for depressive episode monitoring by setting the internal training iterations to 1, 2, 3, 4, 5, and 10. As shown in Fig. 21, precision, recall, F1 score, and accuracy all decrease as the number of internal training iterations increases, but the decrease is very small. This is because the increase in the number of internal training iterations leads to overfitting of the model on specific tasks. Among them, when the number of internal training iterations is 1, all performance metrics are optimal. Therefore, choosing an internal training iteration of 1 not only reduces meta-learning training time but also results in better performance in depressive episode monitoring.

The number of source domain subjects. The diversity of subjects is closely related to the generalization performance of the model. In the meta-learning phase, apart from each individual subject, we randomly select 3, 5, 7, 10, 15, 17 subjects from the remaining subjects as source domain subjects. As shown in Fig. 22, as the number of source domain subjects increases, the performance of depressive episode monitoring gradually improves. With only three source domain subjects, the accuracy of depressive episode monitoring reaches 84.26%, which is a 29.8% improvement compared to the cross-subject accuracy of the LCMTN. From having 3 source domain subjects to 15, the accuracy increases from 84.26% to 92.1%. Further increasing to include all source domain subjects results in only a 1.42% improvement in accuracy. Even if the number of source domain subjects is further increased, the room for performance improvement is extremely limited while the data cost continues to increase.

3) Impact of Knowledge Distillation: A knowledge distillation module is designed to reduce computational complexity and memory usage, while ensuring the performance of depressive episode monitoring. The distillation degree τ refers to a parameter that quantifies the intensity of knowledge distillation during the knowledge distillation process. In this regard, we use FSL-DEM and FSL-DEM (KD) with a distillation degree of 10 for the encoder. As shown in Fig. 23, the accuracy after distillation for most subjects is slightly lower than the accuracy before distillation. The average accuracy before and after distillation is 93.52% and 90.53, with a difference of only 2.99%. Additionally, subjects U4, U9, and U10 have improved accuracy through knowledge distillation, indicating that knowledge distillation can optimize performance for certain subjects. Furthermore, we evaluate the impact of knowledge distillation at different degrees on the performance of depressive episode monitoring. We respectively set the degrees of knowledge distillation for the encoder to 5, 10, 20, and 40, and obtain the results shown in Fig. 24. It can be observed that as the degree of knowledge distillation increases, various performance metrics show a gradual decrease. It is worth noting that although the degree of knowledge distillation increases from 5 to 40, the performance has declined slightly, but not significantly. Therefore, when selecting the degree of knowledge distillation, it is necessary to strike a balance between model performance and computational resources.

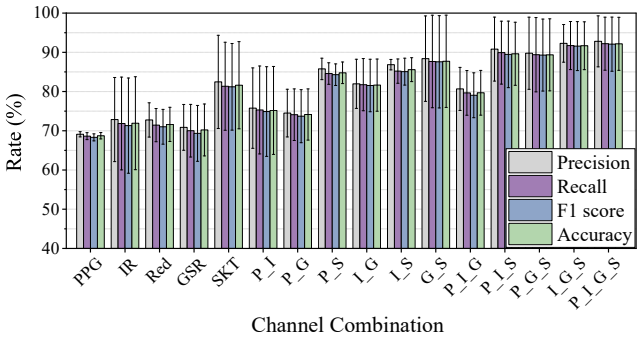


Fig. 26. The performance of different channel combinations

4) Performance with Different Channel Combinations: In order to explore the impact of different channel combinations on depressive episode monitoring performance, we further conduct a feature ablation study. The evaluation settings are consistent with Sec. V-E3. Fig. 25 and Fig. 26 summarize the results. For one-channel, it can be observed that the SKT channel performs the best, with an accuracy of 81.61%. This finding suggests that depressed subjects exhibit significant changes in body temperature during depressive episodes. For two-channel, the performance of channel combinations is better than of sub-channels. When the SKT channel is combined with other channels, it significantly enhances the performance of the other channels. Especially, the combination of SKT and GSR exhibits the best performance, with an accuracy of 87.72%. For three-channel and four-channel, channel combinations' performance is better than sub-channel combinations. These results indicate that integrating information from different channels can complement and enhance the performance of the model in cross-subject depressive episode monitoring.

G. Evaluation of System Running Performance

1) Battery Consumption: At the outset of our assessment, we confirm that both the smartphone and wearable hardware

Authorized licensed use limited to: SHENZHEN UNIVERSITY. Downloaded on July 29,2025 at 15:21:18 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2025.3591096

are powered up to 100%. Participants are instructed to equip the hardware and keep the corresponding app active on their smartphones, with all other apps shut down. This two-hour assessment offers key insights into battery performance. We utilize the Android Debug Bridge, a valuable tool for retrieving the device's live battery status. Fig. 27 shows that after two hours, the battery of the smartphone and the wearable device drop by 19% and 38%, respectively. According to this trend, the entire system can run continuously for over five hours. To further enhance its usability and adapt it to daily usage habits, we also introduce a low power mode (LPM) for the wearable hardware. In this mode, the data transfer rate is reduced and messages are sent every 5 minutes. The results show that the wearable device consumes 20% less power when running in the LPM for two hours, with the battery level dropping by 18%. At the same time, the smartphone experiences a 14% battery drop in the LPM over two hours, representing a 5% power saving compared to normal mode. These findings indicate that the hardware system can run in the LPM for over twelve hours, which demonstrates remarkable endurance capability.

2) FLOPs and Latency: To further validate the effectiveness of knowledge distillation when deploying models on edge devices, we test the floating point operations (FLOPs) and inference latency of the student model and the teacher model in the inference stage with two different lengths of physiological signals as input. As shown in Table IV, the teacher model exhibits significantly higher computational demands with FLOPs reaching 2.04×10^8 for an input length of 32 seconds and 4.09×10^8 for 64 seconds. Accordingly, its latency is also much higher, with 2.586 seconds for an input length of 32 seconds and 4.061 seconds for an input length of 64 seconds. In contrast, the student model demonstrates superior efficiency, requiring only 1.53×10^7 FLOPs and 3.06×10^7 for two inputs, respectively. The latency is also reduced to 0.806 seconds and 0.961 seconds, respectively. These results highlight the student model's suitability for edge deployment, emphasizing the benefits of knowledge distillation for resource-constrained devices.

3) CPU and Memory Load: In this experiment, we demonstrate the dynamics of the CPU and memory footprint of a student model deployed on an edge device for acquisition. All other applications and services are closed to ensure accurate measurements of CPU and memory load. Android Profiler in USB debug mode is used to monitor real-time CPU and memory usage for a duration of 110 seconds. As shown in Fig. 28, the results are mainly divided into three phases. In the initial stage of the silent state (highlighted in gray), CPU usage remains minimal, and memory usage stabilizes at around 240 MB. During the data acquisition and visualization phase (highlighted in yellow), memory usage rises to approximately 550 MB, while CPU load increases slightly to 10%. We acquire 60 seconds of physiological signaling data during this phase. In the data processing and model prediction phase (highlighted in blue), which lasts about 3 seconds, resource consumption surges, with memory usage reaching about 720 MB and CPU utilization peaking at 92%, highlighting the computational intensity of the predictive modeling process.

TABLE IV
EVALUATION OF FLOPS AND LATENCY FOR TEACHER AND STUDENT MODELS

Model	Input Length (s)	FLOPs	Latency (s)
Teacher	32	2.04×10^8	2.586
(1.66 M parameters)	64	4.09×10^8	4.061
Student	32	1.53×10^7	0.806
(0.14 M parameters)	64	3.06×10^7	0.961

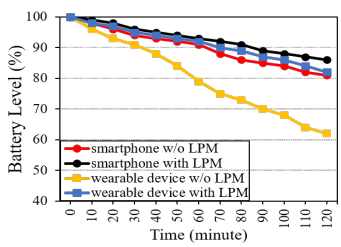


Fig. 27. Battery consumption

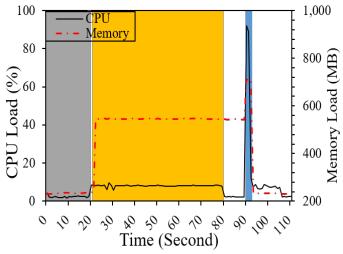


Fig. 28. CPU and memory load

VI. DISCUSSION AND FUTURE WORK

In this section, we discuss some important details of our device and the future work.

A. Wearable Devices

As shown in Table V, we compare our self-designed hardware with existing commercial devices in terms of size, weight, cost, and data interface. We can see that these commercial devices are much more expensive than ours or do not provide access to sensor data. This is why we need to design such a device by ourselves instead of using commercial ones. Specifically, the cost of our device (Ver 1.0) encompasses not only the four sensors detailed in Table I but also expenditures of the PCB circuit board (0.70 \$) and 3D-printed casing (6.382 \$). The entire hardware is encapsulated within a custom-designed 3D printed enclosure with dimensions of 7 cm×4 cm×2.5 cm. We can also see that our device (Ver 1.0) is larger and heavier than other commercial devices, as our design incorporates several physiological sensors through an embedded system approach. However, by enhancing the integration density of hardware circuits and optimizing the lightweight design of the enclosure, we have enabled the device (Ver 2.0) to achieve more compact size and reduced weight, as shown in Table V.

It is worth noting that, despite our device's affordability, it still achieves accurate depression detection recognition for two key reasons. On the one hand, as detailed in Table I, our device supports much higher sampling rates for four kinds of sensors than commercial devices such as Empatica E4 with a maximum sampling rate of 64 Hz. The high sampling rates allow us to capture higher resolution and more detailed temporal signals. On the other hand, we have proposed novel methods to address the issue of human heterogeneity and enhance cross-subject recognition performance.

[2]https://product.tdk.com/
[3]https://store.google.com/tw/product
[4]https://www.empatica.com/en-int/research/e4/

Authorized licensed use limited to: SHENZHEN UNIVERSITY. Downloaded on July 29,2025 at 15:21:18 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2025.3591096

TABLE V
THE COMPARISON OF OUR SELF-DESIGNED HARDWARE WITH COMMERCIAL DEVICES

Device	Size (cm)	Mass (g)	Price ($)	Data Interface
Silmee W20 [2]	$2.05 \times 6.5 \times 1.25$	27.5	538.56	not available
Fitbit Versa 4 [3]	$4.05 \times 4.05 \times 1.12$	37.64	120	not available
Empatica E4 [4]	$4.4 \times 4 \times 19$	25	1080	pay 1080 $ for 3-year data access
Ours (Ver 1.0)	$7 \times 4 \times 2.5$	120	23.56	available
Ours (Ver 2.0)	$3.8 \times 3.8 \times 1.1$	28.6	18.57	available

B. Real-world Adaptation

Although disease recognition and monitoring increasingly depend on data from real-world, dynamic environments in the current era of ubiquitous computing, our study remains limited in adapting to this trend.

(1) Environmental inconsistencies and noise interference. The differences of data-collection environments create significant challenges. Real-world deployment faces unpredictable factors, including environmental noise, device variability, and patients' daily routines. At present, we require subjects to remain in a resting state during data collection, while noises generated by human motion artifacts inevitably affect the data quality of the physiological signals. In the future, we will explore reasonable artifact noise processing algorithms (e.g., wavelet algorithm) to mitigate motion artifacts and enhance robustness under naturalistic conditions.

(2) Model generalizability and adaptability. Model adaptability is also an important issue. Models optimized for specific tasks and data distributions in controlled environments may struggle with the diverse and dynamic conditions and requirements in real-world settings. For example, a model trained to extract depression-related features from particular physiological signals may not promptly recognize and adjust to the wider range of symptoms and behaviors exhibited by patients in the wild, potentially resulting in suboptimal effectiveness. We will continue to improve the *DepGuard* by enhancing its flexibility and adaptability to a broader population and more dynamic conditions, potentially drawing on recent advances in multi-modal federated learning system [53].

(3) Longitudinal validation. Due to the short hospitalization time of depressed patients, our work lacks a cross-session test to validate longitudinal performance across time. This limitation hinders our understanding of model consistency and reliability over extended periods. In future work, we aim to distribute the hardware and software to experimenters, enabling them to collect richer data and achieve real-time monitoring in their daily lives. This will facilitate long-term evaluation and validation.

VII. CONCLUSION

In summary, we present *DepGuard*, a wearable depression assessment system that integrates a wrist-worn device with a mobile app for both depression recognition and real-time monitoring of depressive episodes. The system is designed for potential clinical application to help alleviate the workload of healthcare professionals. To address cross-domain generalization challenges, we employ unsupervised domain adaptation and few-shot learning techniques. Extensive experiments validate the effectiveness of our approach. Results indicate a strong correlation between heart rate and depression recognition, as well as between skin temperature and depressive episode detection, further reinforcing findings in psychiatric research.

REFERENCES

[1] W. H. Organization *et al.*, "Depressive disorder (depression) world health organization," *Geneva, Switzerland*, 2023.
[2] D. American Psychiatric Association, A. P. Association *et al.*, *Diagnostic and statistical manual of mental disorders: DSM-5*. American psychiatric association Washington, DC, 2013, vol. 5, no. 5.
[3] H. Herrman, V. Patel, C. Kieling, M. Berk, C. Buchweitz, P. Cuijpers, T. A. Furukawa, R. C. Kessler, B. A. Kohrt, M. Maj *et al.*, "Time for united action on depression: a lancet–world psychiatric association commission," *The Lancet*, vol. 399, no. 10328, pp. 957–1022, 2022.
[4] X. Xu, B. Yao, Y. Dong, S. Gabriel, H. Yu, J. Hendler, M. Ghassemi, A. K. Dey, and D. Wang, "Mental-llm: Leveraging large language models for mental health prediction via online text data," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 8, no. 1, pp. 1–32, 2024.
[5] X. Xu, P. Chikersal, J. M. Dutcher, Y. S. Sefidgar, W. Seo, M. J. Tumminia, D. K. Villalba, S. Cohen, K. G. Creswell, J. D. Creswell *et al.*, "Leveraging collaborative-filtering for personalized behavior modeling: a case study of depression detection among college students," *IMWUT*, vol. 5, no. 1, pp. 1–27, 2021.
[6] Y. Rykov, T.-Q. Thach, I. Bojic, G. Christopoulos, J. Car *et al.*, "Digital biomarkers for depression screening with wearable devices: cross-sectional study with machine learning modeling," *JMIR mHealth and uHealth*, vol. 9, no. 10, p. e24872, 2021.
[7] M. Tlachac, R. Flores, M. Reisch, K. Houskeeper, and E. A. Rundensteiner, "Deprest-cat: Retrospective smartphone call and text logs collected during the covid-19 pandemic to screen for mental illnesses," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 2, pp. 1–32, 2022.
[8] M. A. Wani, M. A. ELAffendi, K. A. Shakil, A. S. Imran, and A. A. Abd El-Latif, "Depression screening in humans with ai and deep learning techniques," *IEEE Transactions on Computational Social Systems*, 2022.
[9] C. Lin, P. Hu, H. Su, S. Li, J. Mei, J. Zhou, and H. Leung, "Sensemood: depression detection on social media," in *Proceedings of the 2020 International Conference on Multimedia Retrieval*, 2020, pp. 407–411.
[10] S. Nepal, W. Liu, A. Pillai, W. Wang, V. Vojdanovski, J. F. Huckins, C. Rogers, M. L. Meyer, and A. T. Campbell, "Capturing the college experience: A four-year mobile sensing study of mental health, resilience and behavior of college students during the pandemic," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 8, no. 1, pp. 1–37, 2024.
[11] X. Xu, X. Liu, H. Zhang, W. Wang, S. Nepal, Y. Sefidgar, W. Seo, K. S. Kuehn, J. F. Huckins, M. E. Morris *et al.*, "Globem: cross-dataset generalization of longitudinal human behavior modeling," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 4, pp. 1–34, 2023.
[12] K. Opoku Asare, Y. Terhorst, J. Vega, E. Peltonen, E. Lagerspetz, and D. Ferreira, "Predicting depression from smartphone behavioral markers using machine learning methods, hyperparameter optimization, and feature importance analysis: exploratory study," *JMIR mHealth and uHealth*, vol. 9, no. 7, p. e26540, 2021.
[13] Y. Pan, Y. Shang, T. Liu, Z. Shao, G. Guo, H. Ding, and Q. Hu, "Spatial–temporal attention network for depression recognition from facial videos," *Expert Systems with Applications*, vol. 237, p. 121410, 2024.
[14] C. Koch, M. Wilhelm, S. Salzmann, W. Rief, and F. Euteneuer, "A meta-analysis of heart rate variability in major depression," *Psychological Medicine*, vol. 49, no. 12, pp. 1948–1957, 2019.
[15] S. Saganowski, B. Perz, A. G. Polak, and P. Kazienko, "Emotion recognition for everyday life using physiological signals from wearables: A systematic literature review," *IEEE Transactions on Affective Computing*, vol. 14, no. 3, pp. 1876–1897, 2022.

Authorized licensed use limited to: SHENZHEN UNIVERSITY. Downloaded on July 29,2025 at 15:21:18 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2025.3591096

[16] G. Tasci, H. W. Loh, P. D. Barua, M. Baygin, B. Tasci, S. Dogan, T. Tuncer, E. E. Palmer, R.-S. Tan, and U. R. Acharya, "Automated accurate detection of depression using twin pascal's triangles lattice pattern with eeg signals," *Knowledge-Based Systems*, vol. 260, p. 110190, 2023.

[17] H. Lin, C. Jian, Y. Cao, X. Ma, H. Wang, F. Miao, X. Fan, J. Yang, G. Zhao, and H. Zhou, "Mdd-tsvm: A novel semisupervised-based method for major depressive disorder detection using electroencephalogram signals," *Computers in Biology and Medicine*, vol. 140, p. 105039, 2022.

[18] S. Byun, A. Y. Kim, E. H. Jang, S. Kim, K. W. Choi, H. Y. Yu, and H. J. Jeon, "Detection of major depressive disorder from linear and nonlinear heart rate variability features during mental task protocol," *Computers in Biology and Medicine*, vol. 112, p. 103381, 2019.

[19] D. Wang, J. Weng, Y. Zou, and K. Wu, "Emotracer: A wearable physiological and psychological monitoring system with multi-modal sensors," in *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers*, 2022, pp. 444–449.

[20] Y. Zou, D. Wang, D. Wang, C. Zheng, Q. Song, Y. Zhu, C. Fan, and K. Wu, "Research on wearable emotion recognition based on multi-source domain adversarial transfer learning (in chinese)," *Chinese Journal of Computers*, vol. 47, no. 2, pp. 266–286, 2024.

[21] S. Hassantabar, J. Zhang, H. Yin, and N. K. Jha, "Mhdeep: Mental health disorder detection system based on wearable sensors and artificial neural networks," *ACM Transactions on Embedded Computing Systems*, vol. 21, no. 6, pp. 1–22, 2022.

[22] Y. Jiao, X. Wang, C. Liu, G. Du, L. Zhao, H. Dong, S. Zhao, and Y. Liu, "Feasibility study for detection of mental stress and depression using pulse rate variability metrics via various durations," *Biomedical Signal Processing and Control*, vol. 79, p. 104145, 2023.

[23] Y. Hu, J. Chen, J. Chen, W. Wang, S. Zhao, and X. Hu, "An ensemble classification model for depression based on wearable device sleep data," *IEEE Journal of Biomedical and Health Informatics*, 2023.

[24] W. Kuang, S. Jin, D. Wang, Y. Zheng, Y. Zou, and K. Wu, "Emotracer: A user-independent wearable emotion tracer with multi-source physiological sensors based on few-shot learning," in *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2024, pp. 2680–2687.

[25] A. Ahmed, J. Ramesh, S. Ganguly, R. Aburukba, A. Sagahyroon, and F. Aloul, "Evaluating multimodal wearable sensors for quantifying affective states and depression with neural networks," *IEEE Sensors Journal*, 2023.

[26] G. Sharma, A. Parashar, and A. M. Joshi, "Dephnn: a novel hybrid neural network for electroencephalogram (eeg)-based screening of depression," *Biomedical Signal Processing and Control*, vol. 66, p. 102393, 2021.

[27] F. Zhang, M. Wang, J. Qin, Y. Zhao, X. Sun, and W. Wen, "Depression recognition based on electrocardiogram," in *International Conference on Computer and Communication Systems*. IEEE, 2023, pp. 1–5.

[28] C.-C. Huang, I. Low, C.-H. Kao, C.-Y. Yu, T.-P. Su, J.-C. Hsieh, Y.-S. Chen, and L.-F. Chen, "Meg-based classification and grad-cam visualization for major depressive and bipolar disorders with semi-cnn," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2022, pp. 1823–1826.

[29] K. Shao, Y. Hao, L. Hu, X. Zong, and M. Chen, "Data augmentation and pseudo-sequence of fnirs for depression recognition," in *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2023, pp. 2223–2226.

[30] M. Mousavian, J. Chen, Z. Traylor, and S. Greening, "Depression detection from smri and rs-fmri images using machine learning," *Journal of Intelligent Information Systems*, vol. 57, pp. 395–418, 2021.

[31] S. Pange and V. Pawar, "Depression analysis based on eeg and ecg signals," in *International Conference for Emerging Technology (INCET)*. IEEE, 2023, pp. 1–6.

[32] Y. Tazawa, K.-c. Liang, M. Yoshimura, M. Kitazawa, Y. Kaise, A. Takamiya, A. Kishi, T. Horigome, Y. Mitsukura, M. Mimura *et al.*, "Evaluating depression with multimodal wristband-type wearable device: screening and assessing patient severity utilizing machine-learning," *Heliyon*, vol. 6, no. 2, 2020.

[33] I. Moshe, Y. Terhorst, K. Opoku Asare, L. B. Sander, D. Ferreira, H. Baumeister, D. C. Mohr, and L. Pulkki-Råback, "Predicting symptoms of depression and anxiety using smartphone and wearable data," *Frontiers in Psychiatry*, vol. 12, p. 625247, 2021.

[34] Y. Fang, J. Wu, Q. Wang, S. Qiu, A. Bozoki, and M. Liu, "Source-free collaborative domain adaptation via multi-perspective feature enrichment for functional mri analysis," *Pattern Recognition*, vol. 157, p. 110912, 2025.

[35] T. Chen, Y. Guo, S. Hao, and R. Hong, "Semi-supervised domain adaptation for major depressive disorder detection," *IEEE Transactions on Multimedia*, 2023.

[36] Z. Zhang, Q. Meng, L. Jin, H. Wang, and H. Hou, "A novel eeg-based graph convolution network for depression detection: incorporating secondary subject partitioning and attention mechanism," *Expert Systems with Applications*, vol. 239, p. 122356, 2024.

[37] Y. Fang, M. Wang, G. G. Potter, and M. Liu, "Unsupervised cross-domain functional mri adaptation for automated major depressive disorder identification," *Medical Image Analysis*, vol. 84, p. 102707, 2023.

[38] Z. Zhang, C. Xu, L. Jin, H. Hou, and Q. Meng, "A depression level classification model based on eeg and gcn network with domain generalization," in *2024 43rd Chinese Control Conference (CCC)*. IEEE, 2024, pp. 8582–8587.

[39] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[40] C. Grady, "Institutional review boards: Purpose and challenges," *Chest*, vol. 148, no. 5, pp. 1148–1155, 2015.

[41] D. P. Kingma, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[42] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. March, and V. Lempitsky, "Domain-adversarial training of neural networks," *Journal of Machine Learning Research*, vol. 17, no. 59, pp. 1–35, 2016.

[43] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7167–7176.

[44] J. Wang, J. Chen, J. Lin, L. Sigal, and C. W. de Silva, "Discriminative feature alignment: Improving transferability of unsupervised domain adaptation by gaussian-guided latent alignment," *Pattern Recognition*, vol. 116, p. 107943, 2021.

[45] R. Hu, L. Chen, S. Miao, and X. Tang, "Swl-adapt: An unsupervised domain adaptation model with sample weight learning for cross-user wearable human activity recognition," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 5, 2023, pp. 6012–6020.

[46] S. Samarasinghe, M. N. Rizve, N. Kardan, and M. Shah, "Cdfsl-v: Cross-domain few-shot learning for videos," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 11 643–11 652.

[47] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[48] M. Kobayashi, G. Sun, T. Shinba, T. Matsui, and T. Kirimoto, "Development of a mental disorder screening system using support vector machine for classification of heart rate variability measured from single-lead electrocardiography," in *IEEE Sensors Applications Symposium*. IEEE, 2019, pp. 1–6.

[49] L. Yang, Y. Wang, X. Zhu, X. Yang, and C. Zheng, "A gated temporal-separable attention network for eeg-based depression recognition," *Computers in Biology and Medicine*, vol. 157, p. 106782, 2023.

[50] K. Shao, R. Wang, Y. Hao, L. Hu, M. Chen, and H. Arno Jacobsen, "Multimodal physiological signals representation learning via multiscale contrasting for depression recognition," in *Proceedings of the 32nd ACM International Conference on Multimedia*, 2024, pp. 5692–5701.

[51] C. Fu, K. Su, Y. Su, F. Qian, Y. Zhang, C. Liu, S. Song, L. Yang, X. Lv, P. Shan *et al.*, "M 3 add: A novel benchmark for physiology signal-based automatic depression detection with multimodal multitask multievent framework," in *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2025, pp. 1–5.

[52] X. Shui, H. Xu, S. Tan, and D. Zhang, "Depression recognition using daily wearable-derived physiological data," *Sensors*, vol. 25, no. 2, p. 567, 2025.

[53] X. Ouyang, X. Shuai, Y. Li, L. Pan, X. Zhang, H. Fu, S. Cheng, X. Wang, S. Cao, J. Xin *et al.*, "Admarker: A multi-modal federated learning system for monitoring digital biomarkers of alzheimer's disease," in *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, 2024, pp. 404–419.

Authorized licensed use limited to: SHENZHEN UNIVERSITY. Downloaded on July 29,2025 at 15:21:18 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Mobile Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMC.2025.3591096

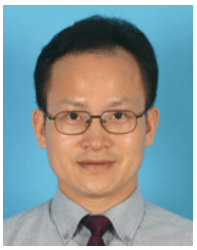
Yufei Zhang received the B.S. and M.S. degree in the School of Computer and Information Engineering, Shanghai Polytechnic University (SSPU), in 2021 and 2024, respectively. He is currently pursuing the PhD degree in the College of Computer Science and Software Engineering, Shenzhen University. His research interests include affective computing, mobile computing, and spatial-temporal data analysis.

Shuo Jin is currently pursuing a master's degree in the College of Computer Science and Software Engineering at Shenzhen University, China. His research focuses on mobile computing and affective computing.

Wenting Kuang 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.

Yuda Zheng is currently pursuing a bachelor's degree in the College of Computer Science and Software Engineering at Shenzhen University, China. His research focuses on mobile application development and affective computing.

Qifeng Song received the MS degree from College of Computer Science and Software Engineering, Shenzhen University, in 2024. His research interests include affective computing and mobile computing.



Changhe Fan is currently a chief physician from the Department of Psychiartry, Guangdong Second Provincial General Hospital. His research interests include biological psychiatry and behavioral medicine.

Yongpan Zou (Member, IEEE) 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.

Victor C. M. Leung (Life Fellow, IEEE) received the PhD degree in electrical engineering from the University of British Columbia, in 1981. He currently is the dean with the Artificial Intelligence Research Institute, Shenzhen MSU-BIT University. His research focuses on wireless networks and mobile systems with more then 60,000 citations. He has received numerous accolades, such as the APEBC Gold Medal, NSERC Postgraduate Scholarships, and IEEE Vancouver Section Centennial Award.

Kaishun Wu (Fellow, IEEE) 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.

Authorized licensed use limited to: SHENZHEN UNIVERSITY. Downloaded on July 29,2025 at 15:21:18 UTC from IEEE Xplore. Restrictions apply.
© 2025 IEEE. All rights reserved, including rights for text and data mining and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.