CoMe: Contrastive Learning-based Mental Disorder Recognition with Wearable Multimodal Physiological Data

Yulan Li*, Shuo Jin*, Kaishun Wu†, Changhe Fan‡, Yongpan Zou*
*College of Computer Science and Software Engineering, Shenzhen University
†Information Hub, Hong Kong University of Science and Technology (Guangzhou)
‡Department of Psychiatry, Affiliated Guangdong Second Provincial General Hospital of Jinan University
yulanlee2024@gmail.com, jinshuo2022@email.szu.edu.cn,
wuks@hkust-gz.edu.cn, changhefan@yahoo.com, yongpan@szu.edu.cn

2025 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 979-8-3315-1557-7/25/$31.00 ©2025 IEEE | DOI: 10.1109/BIBM66473.2025.11357131

***Abstract*—Mental disorders are increasingly prevalent, requiring timely and accurate recognition. Existing methods often depend on specialized devices and single-modality public datasets, limiting their ability to identify diverse mental disorders. Multiclass mental disorder recognition (MDR) faces challenges such as limited labeled data, overlapping symptoms among disorders, and significant individual variability. This paper introduces *CoMe*, a system that leverages low-cost, ubiquitous wearable devices to collect multimodal physiological signals, enabling real-time and daily monitoring to support the early detection of mental disorders. *CoMe* employs a multimodal residual fusion contrastive learning framework to extract both modality-specific and cross-modality knowledge, addressing challenges of sparse labeled data and symptom overlap. Through self-supervised and supervised fine-tuning, *CoMe* personalizes models for new users, facilitating accurate, real-time mental status monitoring in daily life. Validation with 74 participants demonstrates *CoMe*'s superior performance under low-data conditions, outperforming both self-supervised and supervised baselines, and surpassing a state-of-the-art method. *CoMe* achieves average recognition accuracies of 84.5%, 85.36%, 83.25%, and 83.54% for depression, anxiety, bipolar disorders, and control group, respectively.**

***Index Terms*—mental disorder recognition, wearable device, physiological signal, contrastive learning.**

I. Introduction

Mental disorders are highly prevalent worldwide, affecting about 970 million people in 2019 [1], including 31% with anxiety, 28.9% with depression, and 4.1% with bipolar disorder. These disorders severely impair daily functioning and may lead to suicidal thoughts if untreated, emphasizing the urgency of early intervention. Traditional screening methods primarily rely on retrospective self-reports, such as the 9-item Patient Health Questionnaire (PHQ-9) for depression [2], the 7-item Generalized Anxiety Disorder Questionnaire (GAD-7) for anxiety [3], and the Mood Disorder Questionnaire (MDQ) for bipolar disorder [4]. Nevertheless, these questionnaires are subjective, lack continuous monitoring, and are limited in screening multiple disorders. Another approach analyzes complex physiological signals from specialized medical equipment, including Electrocardiogram (ECG) [5], Electroencephalogram (EEG) [6], Magnetic Resonance Imaging (MRI) [7]. Recently, low-cost wearable devices have emerged as promising tools for mental health monitoring. Prior research primarily leverages behavioral data, including activity, location, and sleep, to assess psychological disorders [8], [9]. Nepal *et al.* [8] monitore students' behaviors and mental states over four years via mobile sensing, while Hafiz *et al.* [9] analyze sleep-derived features for bipolar disorder detection. According to the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [10], mental disorders often trigger physiological responses manifested in superficial physiological signals such as Heart Rate Variability (HRV), Oxygen Saturation (SpO_2), Galvanic Skin Response (GSR), and Skin Temperature (SKT) [11]. Recent studies further explore these physiological signals for real-time assessment [12]–[14]. Mishra *et al.* [13] validate GSR and ECG signals from the Empatica E4 for stress detection, and DiSalvo *et al.* [12] examine emotional engagement using GSR.

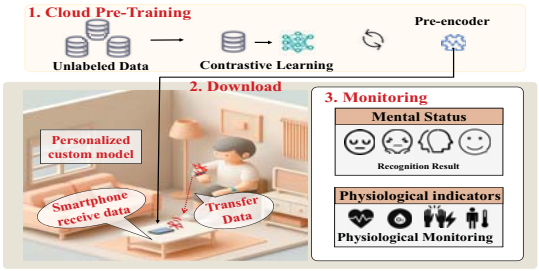

Fig. 1. A typical application scenario of *CoMe*

However, utilizing physiological signals from wearable devices for multi-class mental disorder recognition remains challenging. The lack of high-quality labeled data across categories restricts dataset scalability and reliability. To alleviate data insufficiency, self-supervised methods have been explored [15], [16]. Among them, contrastive learning effectively pretrains unimodal [15], [17] and multimodal physiological data [18]–[20], enhancing downstream performance [21]. Recent studies apply it to mental health analysis. Zhang et al. [22] present a multi-view graph contrastive model for depression recognition using EEG, while Deldari et al. [18] design a cross-modal framework integrating accelerometer, ECG, electromyography,

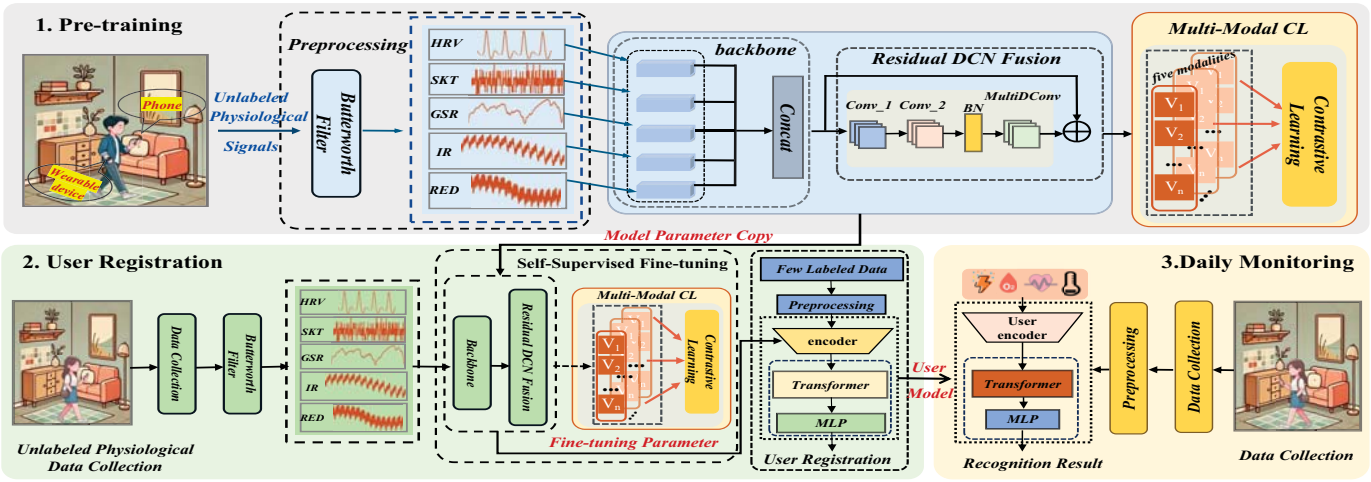

Fig. 2. *CoMe* overview: 1) Pre-t raining phase: Using a large amount of unlabeled physiological data for unsupervised contrastive learning to obtain the pre-encoder. 2) User registration phase: Fine-tuning the pre-encoder with a small amount of unlabeled physiological data from the target domain through contrastive learning. 3) Daily Monitoring phase: Utilizing a customized model for real-time monitoring of mental status.

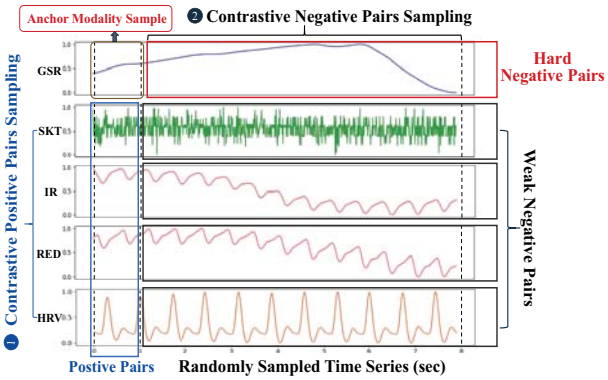

Fig. 3. Sampling principle of multi-modal contrastive learning. An 8-second segment is used for illustration, with GSR as the anchor modality

and GSR for emotion prediction. Yet, these efforts overlook joint recognition of major mental disorders. Additionally, inter-individual variability and overlapping heterogeneous symptoms [23] further hinder model generalization, as in bipolar disorder involving depressive and manic episodes [10].

To overcome these challenges, we propose *CoMe*, a low-cost wearable-based system for early detection of depression, anxiety, and bipolar disorders through multimodal physiological sensing. *CoMe* adopts a two-stage self-supervised contrastive learning framework to leverage both unlabeled and limited labeled data. In Stage 1, a multimodal pre-encoder is trained on large-scale unlabeled data, with user variability simulated via a leave-one-user-out domain split. A multimodal residual fusion encoder with deformable convolutions is developed to effectively extract common knowledge across modalities while preserving modality-specific characteristics. We further propose a coarse-grained and fine-grained contrastive learning (CL) scheme to enhance discriminative representation learning. In Stage 2, the pre-encoder is personalized via unsupervised fine-tuning on target data, followed by supervised adaptation using limited source-labeled samples, enabling generalization to unseen users while preserving individual traits. As illustrated in Fig. 1, *CoMe* follows a cloud–edge architecture. In summary, the main contributions of this paper are as follows:

- We propose *CoMe*, the first multi-class mental disorder monitoring system via low-cost wearables, capturing multimodal signals for daily identification of depression, anxiety, bipolar disorder. Notably, we construct a multimodal physiological dataset encompassing these mental conditions and healthy controls.
- We design a two-stage self-supervised contrastive framework with a multimodal residual fusion pre-training model, extracting cross-modal common and modality-specific knowledge from unlabeled data to adapt to new users efficiently.
- Experiments demonstrate *CoMe* outperforms baselines. With 25% labeled and 2-minute unlabeled target data, it achieves 84.5%, 85.36%, 83.25% accuracy for depression, anxiety, and bipolar disorder respectively, 83.54% for health, 93.76% mental disorder detection rate, and 10.41% healthy misdiagnosis rate, validating its efficacy.

II. System Design

A. System Overview

CoMe comprises pre-training and personalized customization phases, as illustrated in Fig. 2. Raw signals are preprocessed with Butterworth filtering and normalization to ensure signal quality. The pre-encoder is trained using unlabeled source domain data. Pre-processed signals are encoded using modality-specific temporal extractors and fused via a residual Deformable Convolution Network (DCN) module. The pre-encoder is optimized by a contrastive learning method to improve representation discriminability, and is later adapted to each user with a short unlabeled segment for personalization. The customized model subsequently supports continuous and real-time monitoring of mental states in daily scenarios.

B. Data Pre-processing

Raw data inevitably contain noise, requiring preprocessing and normalization to improve MDR performance. *CoMe* collects HRV using a photoplethysmogram (PPG) sensor, SpO_2 with infrared (IR) and red (RED) light sensor, GSR with

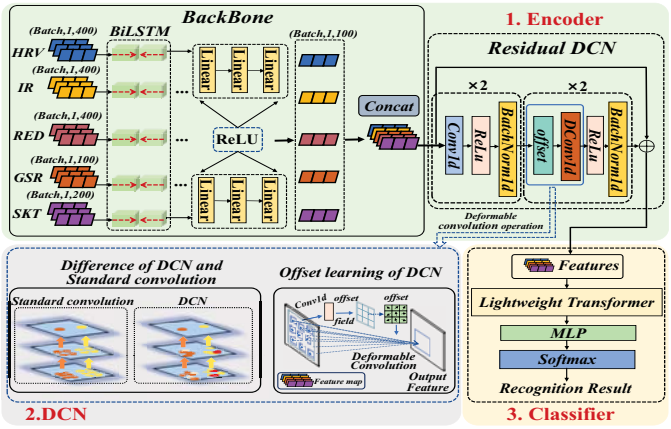

Fig. 4. Illustration of residual DCN fusion.

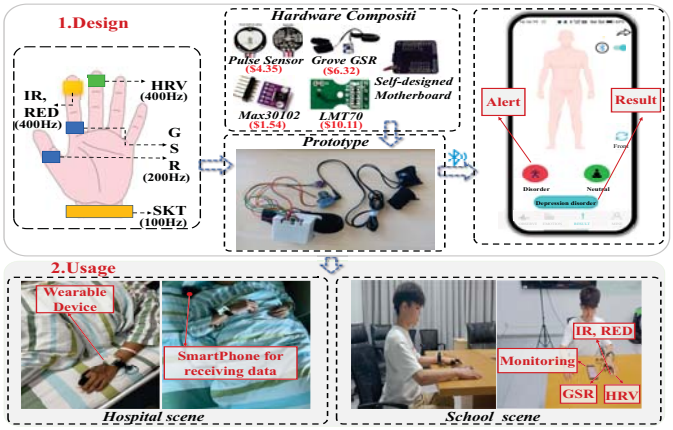

Fig. 5. Design and usage for *CoMe*: 1) The hardware components and software design for *CoMe*. 2) Instructions for participants on how to wear the device and the experimental scenes.

a skin conductance sensor, and SKT with a temperature sensor. Short-Time Fourier Transform analysis reveals that informative frequencies concentrate within limited bands, with noise distributed outside. Thus, Butterworth filters are employed for denoising. To mitigate subject variability, min–max normalization scales each windowed signal to [0, 1].

C. Pre-training

The pre-training stage includes Multi-modal Contrastive Learning and Multi-modal Fusion, which are detailed in the following subsections.

1) Multi-modal Contrastive Learning: While existing contrastive learning methods often rely on coarse pairings across augmented or temporal views [24], [25], recent multimodal extensions [18], [19] remain limited in capturing fine-grained cross-modal dependencies. Building on these insights, *CoMe* integrates both coarse-grained and fine-grained pairings to enhance multimodal feature representations. Fig. 3 illustrates the sampling principle. In the time-aligned setting, one modality is chosen as the anchor, and data from other modalities at the same time point form positive pairs. Samples from the same modality at different times serve as hard negatives, while samples from other modalities at different times form weak negatives. The objective is to maximize the similarity of positive pairs while minimizing that of negative pairs. By learning such modality-aware representations from unlabeled data, *CoMe* improves generalization and enhances discrimination between related disorder types.

Given a batch of size N with 1-second window samples, after passing through the pre-encoder $f(\cdot)$, the features of this batch are obtained as y. For a chosen anchor modality m and a modality n from M ($m \neq n$), among the N 1-second windows, we select the i-th and j-th time steps ($i \neq j$). For the positive pairs loss, the computation is as follows:

$$\mathcal{L}_{pos} = \sum_{n \in M} \exp(sim(y_m^i, y_n^i)/\tau) \tag{1}$$

where sim is a similarity function measuring the distance between positive pairs, τ is a temperature parameter, y_m^i and y_n^i represent the positive pairs of modalities m and n ($m \neq n$) at time i.

For the negative pairs loss, the computation is as follows:

$$\mathcal{L}_{hard\_neg} = \sum_{j \neq i, j=1}^{2N} \exp(sim(y_m^i, y_m^j)/\tau) \tag{2}$$

where y_m^i and y_m^j represent the hard negative pairs of modality m at the i-th and j-th time steps.

$$\mathcal{L}_{weak\_neg} = \sum_{n \in M} \sum_{j \neq i, j=1}^{2N} \exp(sim(y_m^i, y_n^j)/\tau) \tag{3}$$

where y_m^i and y_n^j represent the weak negative samples of modality m at the i-th time step and modality n at the j-th time step.

The hard and weak negative pairs provide fine-grained representations across different modalities, and by summing them, a coarse-grained representation can be obtained along the temporal dimension. After summing, the final contrastive loss is as follows:

$$\mathcal{L} = -\sum_{i=1}^{N} \log \frac{\mathcal{L}_{pos}}{\mathcal{L}_{pos} + \mathcal{L}_{hard\_neg} + \mathcal{L}_{weak\_neg}} \tag{4}$$

where $\mathcal{L}$ the final objective function applied to the MDR task.

2) Multi-modal Fusion: Deformable convolution adaptively adjusts sampling positions and has been extended to temporal data analysis [26]. *CoMe* employs a residual DCN fusion scheme that integrates coarse-grained and fine-grained sampling to extract both local fused and modality-specific global representations. Fig. 4 illustrates the residual DCN fusion process. *CoMe* uses a 1-second window as the sample size for segmenting various physiological signal.

Each physiological modality $M = \{HRV, IR, RED, GSR, SKT\}$ is processed by a Bidirectional Long Short-Term Memory (BiLSTM) followed by a linear layer to extract modality-specific temporal features. Representations from all modalities are concatenated into a unified vector $h(t)$, processed by two 1D convolution layers and a DCN. Offsets for the five channels are computed via 1D convolution and applied to the DCN to fuse information across spatial–temporal positions:

TABLE I
THE AVERAGE RECOGNITION ACCURACY PERFORMANCE OF *CoMe* AND THE BASELINE METHODS

Baselines		Depression Disorder	Anxiety Disorder	Bipolar Disorder	Control group
Supervised Learning	HRV (100%)	**68.43%**	71.25%	**66.49%**	**69.86%**
	IR (100%)	61.64%	**71.8%**	64.07%	61.07%
	RED (100%)	58.84%	67.58%	60.93%	59.26%
	GSR (100%)	50.94%	49.34%	49.57%	51.5%
	SKT (100%)	52.56%	49.73%	54.89%	50.76%
	Multi-Modal (100%)	**73.67%**	**78.82%**	**72.37%**	**75.04%**
Self-Supervised Learning	CPC (100% 4min)	80.53%	82.37%	80.2%	82.46%
	CMC (100% 3min)	79.56%	80.53%	82.34%	81.43%
	COCOA (75% 3min)	80.67%	84.47%	81.11%	81.79%
	Ours (25% 2min)	**84.5%**	**85.36%**	**83.25%**	**83.54%**

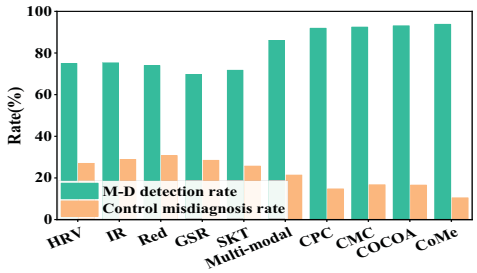

Fig. 6. The performance comparison with baselines under optimal condition.

$$y(t) = \left(\sum_{c \in M} \sum_{k \in \mathcal{R}} w_c(k) \cdot h_c(t + k + \Delta k_c) \right) + h(t) \quad (5)$$

where $y(t)$ is the local fusion and global features obtained at time t through the residual DCN, $w_c(k)$is the weight of the convolution kernel in modality c and k is the position of the convolution kernel relative to the center time point, $\mathcal{R}$ is the sampling range of the convolution kernel in the time dimension, Δk_c is the learned offset for modality c.

D. Personalized Customization

When a new user registers on *CoMe*, the wearable device collects a few minutes of unlabeled multimodal physiological data. This data is locally preprocessed and used to fine-tune the pre-trained encoder $f(\cdot)$ via self-supervised contrastive learning, yielding a personalized encoder $f'(\cdot)$. To further adapt the system, a small set of labeled data is randomly selected and encoded by $f'(\cdot)$, with the resulting features input to a lightweight classifier composed of a Transformer and MLP. This process customizes the encoder $g_e(\cdot)$ and classifier $g_c(\cdot)$ to individual users, enabling daily mental health monitoring. In our implementation, *CoMe* simulates this personalization using the target domain dataset, split into fine-tuning and test sets. The fine-tuning set mimics data collected at registration, and the test set simulates post-personalization inference.

III. EXPERIMENT AND EVALUATION

A. Prototype Implementation

1) Implementation Details: Commercial devices like the Empatica E4 [27] support multimodal physiological sensing but remain expensive and less accessible. *CoMe* designs a low-cost, compact wearable replicating these functions and integrated with software for analysis and visualization. As shown in Fig. 5, the hardware consists of a custom board with Bluetooth and multiple sensors enclosed in a 3D-printed shell. Powered by battery, it enables continuous sensing and wireless data transfer. The total cost is $25.13, significantly lower than commercial counterparts [27].

B. Data Collection

We collect labeled and unlabeled physiological signals using our custom-designed wearable over ten days, with each subject completing a 30-minute daily session. A total of 74 subjects (40 males, 34 females; aged 14–60) participate, including 56 with mental disorders (18 with depression, 22 with anxiety, and 16 with bipolar disorder) confirmed by psychiatrists through standardized assessments. The remaining 18 healthy controls are recruited from a school and screened negative for mental disorders using SDS [28], SCL-90 [29], and SAS [30]. Individuals with severe cardiovascular, hypertensive, or respiratory conditions are excluded to minimize confounding effects. As shown in Fig. 5, data are collected in natural indoor settings such as hospital wards and school rooms. Subjects are instructed to remain still to avoid sensor displacement. Informed consent is obtained from all participants, and the study protocol is approved by the institutional ethics committee.

C. Evaluation Protocol

1) Evaluation Metrics: We use three clinically relevant metrics: accuracy, detection rate, and misdiagnosis rate. Accuracy measures overall classification performance as the ratio of correctly classified samples to total samples. Detection rate indicates the ability to identify mental disorder (M-D) cases, computed as the fraction of samples recognized as disorders per participant. Misdiagnosis rate captures false positives in the control group, measured as the fraction of healthy samples misclassified as disorders, where lower values indicate higher specificity.

2) Baselines: To assess the performance of *CoMe*, we compare it with supervised and self-supervised baselines. For supervised learning, baselines are trained from scratch with cross-entropy loss using the same architecture as *CoMe*. We implement five single-modal baselines (HRV, IR, RED, GSR, SKT), each trained on one physiological channel without the fusion module, and one multimodal baseline combining all modalities. For self-supervised baselines, we consider three contrastive methods. CPC [20] constructs temporal context-based pairs. CMC [19] contrasts representations across modal-

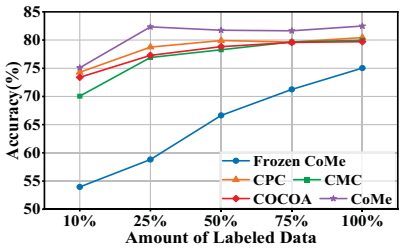

Fig. 7. Accuracy comparison with baselines across Labeled Data Amounts.

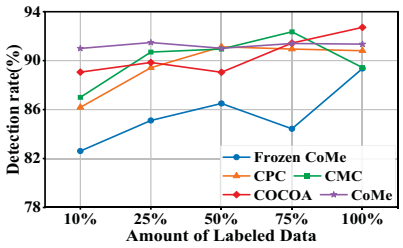

Fig. 8. M-D group detection rate comparison with baselines across Labeled Data Amounts.

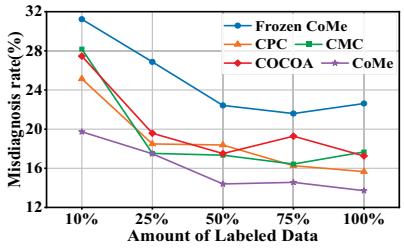

Fig. 9. Control group misdiagnosis rate comparison with baselines across Labeled Data Amounts.

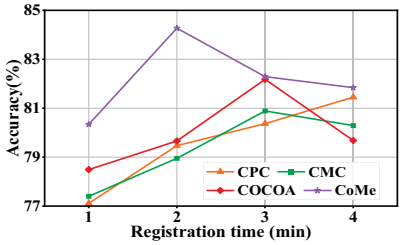

Fig. 10. Accuracy comparison with baselines over Registration Time.

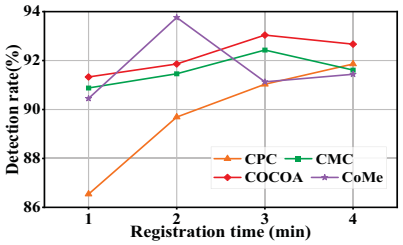

Fig. 11. M-D group detection rate comparison with baselines over Registration Time.

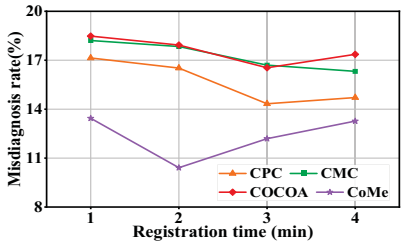

Fig. 12. Control group misdiagnosis rate comparison with baselines over Registration Time.

ities. COCOA [18], a representative multimodal method, conducts cross-modal and intra-modal temporal contrast on aligned samples without distinguishing negative types or granularities. All self-supervised baselines share the same network and training settings as *CoMe* for fair comparison.

3) Implementation Details: To mitigate random data selection effects, each participant undergoes five trials with averaged results. In each trial, 12 minutes of data are randomly sampled. Each 1-second window yields 720 samples, with 120 for registration and 600 for evaluation. The training uses Adam optimizer with a learning rate of 1e-4 and batch size of 256.

D. Results

1) Main Performance Comparison: To evaluate the performance of *CoMe* and baselines under low-data conditions, we conduct extensive experiments. Table I presents the average recognition accuracy of *CoMe* and baselines, with parentheses denoting the percentage of labeled data and the registration time used for personalization, which corresponding to the optimal performance. *CoMe* demonstrates remarkable data efficiency, achieving competitive performance with only 2 minutes of registration and 25% labeled data. By fusing five physiological channels, *CoMe* surpasses supervised models trained on full labels, improving by over 3% compared to the best single-modal model and by 6% over multi-modal baselines. In self-supervised settings, *CoMe* achieves optimal registration at 2 minutes, outperforming the best baseline (COCOA) by 2.15% on average while using far fewer labels. Supervised methods, limited by scarce labels, often overfit and fail to distinguish such cases. In contrast, under conditions of individual heterogeneity and overlapping symptom features, *CoMe* improves inter-class separability by aligning positive pairs and repelling negatives, resulting in more robust and generalizable representations.

Fig. 6 illustrates detection and misdiagnosis rates under optimal settings. Compared with supervised methods, *CoMe* yields an 18.37% higher detection rate for M-D group and reduces the control misdiagnosis rate by 15.34%. In self-supervised comparisons, although the detection rate gain is modest, *CoMe* achieves a 4.3% lower misdiagnosis rate than the best baseline, with a final rate of 10.41%, within clinically accepted ranges. Overall, *CoMe* effectively enhances disorder detection while minimizing the impact on healthy individuals.

2) Performance with different amounts of labeled data: To compare the performance of *CoMe* with other contrastive self-supervised baselines under different labeled data proportions, we use 10 minutes of data per participant, excluding registration time. This setup allows exploration of the optimal label quantity for a general model and facilitates encoder adaptation to unlabeled target data. For *CoMe*, we adopt two fine-tuning strategies, one freezes the pre-trained encoder and trains only the classifier, while the other jointly tunes both modules. The same settings are applied to all baselines for fairness.

As shown in Fig. 7, performance improves with more labels but plateaus later. *CoMe* consistently surpasses baselines and shows larger gains with limited data. With 25% labeled data, it exceeds CPC, CMC, and COCOA by 1.87%, 2.21%, and 2.73%. When the encoder is frozen, the average accuracy reaches 75.06%. The decline with more labels indicates that leveraging extra unlabeled data enhances representation robustness and limits overfitting. Fig. 8 and Fig. 9 show that *CoMe* achieves similar detection and misdiagnosis rates using only 25% labeled data, differing by merely 1.3% and 1.1% from the full-label case. It demonstrates higher data efficiency under limited supervision.

3) Performance with different registration time: To compare *CoMe* with other self-supervised contrastive baselines under varying registration times, the total sample length is fixed at 10 minutes, and 1–4 minutes of unlabeled target-domain data are used to fine-tune encoder $f(\cdot)$. As shown in Fig. 10, *CoMe* attains an average four-class accuracy of 84.16% with 2-minute registration, 1.89% higher than that in Section III-D2. All baselines, including *CoMe*, improve

with short registration but degrade with longer durations, likely due to overfitting to user-specific patterns and reducing generalizability. Fig. 11 and Fig. 12 show that *CoMe* achieves a 93.76% detection rate with 2-minute registration, 2.28% higher than that in Sec. III-D2. These results suggest that providing 2 minutes of unlabeled physiological data during the registration phase, along with a 10-minute monitoring window, enables effective real-time mental state assessment for new users.

IV. Conclusion

In this work, we presente *CoMe*, a wearable-based monitoring system for multi-class mental disorder recognition in daily environments. Using multimodal contrastive learning with limited labeled data, *CoMe* effectively captures cross-modal common and modality-specific physiological patterns. Experimental results on real-world data validate its superior generalization and adaptability, enabling rapid model personalization with minimal data. The results demonstrate *CoMe's* potential for continuous mental health screening and early risk detection in real-life scenarios.

Acknowledgment

This work was supported in part by China NSFC (No. 62172286, 62472366); in part by key program of DEGP (No. 2024ZDZX1011, 2024GCZX003, 2023KCXTD042); in part by Guangdong Provincial Key Lab of Integrated Communication, Sensing and Computation for Ubiquitous Internet of Things(No.2023B1212010007); in part by Shenzhen Science and Technology Program (No. JCYJ20250604181429038); in part by Shenzhen Science and Technology Foundation (No. ZDSYS20190902092853047); in part by 111 Center (No. D25008); in part by Guangdong Yiyang Healthcare Charity Foundation (No. JZ2022001-3); in part by Talents' plan Foundation of Guangdong Second Provincial General Hospital (No. 2024C003). Both Changhe Fan and Yongpan Zou are the corresponding authors.

References

[1] W. H. Organization, *World mental health report: Transforming mental health for all*. World Health Organization, 2022.

[2] K. Kroenke, R. L. Spitzer, and J. B. Williams, "The phq-9: validity of a brief depression severity measure," *Journal of general internal medicine*, vol. 16, no. 9, pp. 606–613, 2001.

[3] R. L. Spitzer, K. Kroenke, J. B. Williams, and B. Löwe, "A brief measure for assessing generalized anxiety disorder: the gad-7," *Archives of internal medicine*, vol. 166, no. 10, pp. 1092–1097, 2006.

[4] R. M. Hirschfeld, J. B. Williams, R. L. Spitzer, J. R. Calabrese, L. Flynn, P. E. Keck Jr, L. Lewis, S. L. McElroy, R. M. Post, D. J. Rapport *et al.*, "Development and validation of a screening instrument for bipolar spectrum disorder: the mood disorder questionnaire," *American journal of psychiatry*, vol. 157, no. 11, pp. 1873–1875, 2000.

[5] 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.

[6] 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.

[7] R. Sujatha, K. Tejesh, H. Krithi, and H. R. Shri, "Detection of bipolar disorder using machine learning with mri." in *ISIC*, 2021, pp. 445–452.

[8] 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 IMWUT*, vol. 8, no. 1, pp. 1–37, 2024.

[9] P. Hafiz, K. W. Miskowiak, A. Maxhuni, L. V. Kessing, and J. E. Bardram, "Wearable computing technology for assessment of cognitive functioning of bipolar patients and healthy controls," *Proceedings of the ACM IMWUT*, vol. 4, no. 4, pp. 1–22, 2020.

[10] D. American Psychiatric Association, D. American Psychiatric Association *et al.*, *Diagnostic and statistical manual of mental disorders: DSM-5*. American psychiatric association Washington, DC, 2013, vol. 5, no. 5.

[11] F. Zapata, E. Smith, and V. Kreinovich, "Systems approach explains why low heart rate variability is correlated with depression (and suicidal thoughts)," in *Uncertainty, Constraints, and Decision Making*. Springer, 2023, pp. 49–53.

[12] B. DiSalvo, D. Bandaru, Q. Wang, H. Li, and T. Plötz, "Reading the room: Automated, momentary assessment of student engagement in the classroom: Are we there yet?" *Proceedings of the ACM IMWUT*, vol. 6, no. 3, pp. 1–26, 2022.

[13] V. Mishra, S. Sen, G. Chen, T. Hao, J. Rogers, C.-H. Chen, and D. Kotz, "Evaluating the reproducibility of physiological stress detection models," *Proceedings of the ACM IMWUT*, vol. 4, no. 4, pp. 1–29, 2020.

[14] Y. Zhang, S. Jin, W. Kuang, Y. Zheng, Q. Song, C. Fan, Y. Zou, V. C. M. Leung, and K. Wu, "Depguard: Depression recognition and episode monitoring system with a ubiquitous wrist-worn device," *IEEE Transactions on Mobile Computing*, pp. 1–18, 2025.

[15] E. Eldele, M. Ragab, Z. Chen, M. Wu, C. K. Kwoh, X. Li, and C. Guan, "Time-series representation learning via temporal and contextual contrasting," *arXiv preprint arXiv:2106.14112*, 2021.

[16] B. Guo, W. Zuo, S. Wang, W. Lyu, Z. Hong, Y. Ding, T. He, and D. Zhang, "Wepos: Weak-supervised indoor positioning with unlabeled wifi for on-demand delivery," *Proceedings of the ACM IMWUT*, vol. 6, no. 2, pp. 1–25, 2022.

[17] M. Caron, I. Misra, J. Mairal, P. Goyal, P. Bojanowski, and A. Joulin, "Unsupervised learning of visual features by contrasting cluster assignments," *Advances in neural information processing systems*, vol. 33, pp. 9912–9924, 2020.

[18] S. Deldari, H. Xue, A. Saeed, D. V. Smith, and F. D. Salim, "Cocoa: Cross modality contrastive learning for sensor data," *Proceedings of the ACM IMWUT*, vol. 6, no. 3, pp. 1–28, 2022.

[19] Y. Tian, D. Krishnan, and P. Isola, "Contrastive multiview coding," in *In Proceedings of Springer ECCV*, 2020, pp. 776–794.

[20] A. v. d. Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," *arXiv preprint arXiv:1807.03748*, 2018.

[21] J. Jiang, Y. Shu, J. Wang, and M. Long, "Transferability in deep learning: A survey," *arXiv preprint arXiv:2201.05867*, 2022.

[22] S. Zhang, H. Wang, Z. Zheng, T. Liu, W. Li, Z. Zhang, and Y. Sun, "Multi-view graph contrastive learning via adaptive channel optimization for depression detection in eeg signals." *International Journal of Neural Systems*, vol. 33, no. 11, pp. 2 350 055–2 350 055, 2023.

[23] D. Goldberg, "The heterogeneity of "major depression"," *World Psychiatry*, vol. 10, no. 3, p. 226, 2011.

[24] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *In Proceedings of the PMLR ICML*, 2020, pp. 1597–1607.

[25] S. Tonekaboni, D. Eytan, and A. Goldenberg, "Unsupervised representation learning for time series with temporal neighborhood coding," *arXiv preprint arXiv:2106.00750*, 2021.

[26] S. Xu, L. Zhang, W. Huang, H. Wu, and A. Song, "Deformable convolutional networks for multimodal human activity recognition using wearable sensors," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–14, 2022.

[27] Empatica, "Empatica e4 wristband," 2021, accessed: 2024. [Online]. Available: https://www.empatica.com/research/e4/

[28] W. W. Zung, "A self-rating depression scale," *Archives of general psychiatry*, vol. 12, no. 1, pp. 63–70, 1965.

[29] L. R. Derogatis and R. Unger, "Symptom checklist-90-revised," *The Corsini encyclopedia of psychology*, pp. 1–2, 2010.

[30] W. W. Zung, "A rating instrument for anxiety disorders." *Psychosomatics: Journal of Consultation and Liaison Psychiatry*, 1971.