

# UAED: Unsupervised Abnormal Emotion Detection Network Based on Wearable Mobile Device

Jiaqi Zhu , Fang Deng , Senior Member, IEEE, Jiachen Zhao , Daoming Liu, and Jie Chen , Fellow, IEEE

**Abstract**—With the development of the internet-of-medical-things, health monitoring through physiological signals has become a critical task. Given this opportunity, research on personal healthcare systems for abnormal emotion detection using physiological signals brings significant benefits to the field of digital healthcare and human-computer interaction. However, it is a challenging task because of the diverse patterns of time series and the lack of labels. In this work, we present a novel model for Unsupervised Abnormal Emotion Detection (UAED) combining Gaussian mixture variational autoencoder (VAE) and convolutional neural networks (CNNs), whose core idea is to reconstruct the input by learning its latent representation thus capturing the normal patterns and applying whitening distance as the anomaly score to detect outliers. In addition, UAED uses stacking operation to transform one-dimensional time series into high-dimension to help the model capture the periodic features and reflect diverse normal patterns. We conduct extensive experiments on four public datasets to demonstrate that our UAED obtains the best performance in various metrics. Furthermore, we deploy UAED in a real environment using a low-cost wearable sensor developed by us to collect electrocardiogram signals and run UAED on mobile terminals with an accuracy of 85%, validating the feasibility of our healthcare system for detecting abnormal emotions.

**Index Terms**—Anomaly detection, physiological signals, variational autoencoder, wearable device.

## I. INTRODUCTION

DIGITAL healthcare receives widespread attention benefiting from the development of artificial intelligence technology and wearable devices. While emotions have a significant

Manuscript received 28 February 2023; accepted 24 April 2023. Date of publication 1 May 2023; date of current version 25 October 2023. This work was supported in part by the National Science Fund for Distinguished Young Scholars of China under Grant 62025301 and in part by the National Natural Science Foundation of China under Grant 61933002. Recommended for acceptance by Dr. Kun Yang. (*Corresponding author: Fang Deng*.)

This work involved human subjects or animals in its research. Medical and Laboratory Animal Ethics Committee of Beijing University of Technology, Application No. 58,2020, and performed in line with the “Multimodal cross-domain biofeedback functional stimulation and healthstate modulation technology and application research”.

Jiaqi Zhu and Jiachen Zhao are with the Department of Automation, Beijing Institute of Technology, Beijing 100081, China (e-mail: sodaq@sina.com; zhao\_jiachen@163.com).

Fang Deng is with the Department of Automation, Beijing Institute of Technology, Beijing 100081, China, and also with the Beijing Institute of Technology Chongqing Innovation Center, Chongqing 401120, China (e-mail: dengfang@bit.edu.cn).

Daoming Liu is with the Beijing Institute of Technology Chongqing Innovation Center, Chongqing 401120, China (e-mail: 1213982008@qq.com).

Jie Chen is with the Shanghai Research Institute for Intelligent Autonomous Systems, Tongji University, Shanghai 200092, China, and also with the Key Laboratory of Intelligent Control and Decision of Complex Systems, Beijing Institute of Technology, Beijing 100081, China (e-mail: chenjie@bit.edu.cn).

Digital Object Identifier 10.1109/TNSE.2023.3271354

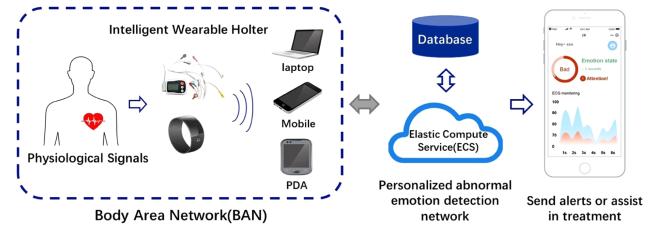


Fig. 1. Intelligent health status monitoring system based on wearable devices.

role to play not only in people’s daily life and work, but also greatly influence human health. Experience of negative emotions is inevitable, even so, excessive, prolonged, or contextually inappropriate negative emotions can affect the body’s hormone production [1], reduce resistance and immunity [2], and even increase the risk of cardiovascular disease [3] and some cancers [4]. Considering the serious damage and suffering caused by negative emotions, it is necessary to detect and give significant attention to them.

In recent years, emotion detection has been widely used in areas such as intelligent healthcare, human-computer interaction, robotics, and rehabilitation assistance. Emotion detection methods can be broadly categorized into two types according to the source of the monitored data, based on external and internal human emotion information. Emotion detection from external information is mainly based on facial expressions and vocal tones, while emotion detection from internal information is based on physiological signals such as electrocardiograms (ECG), electroencephalograms (EEG), electromyography (EMG), galvanic skin response (GSR), heart rate and respiration. Physiological signals are directly regulated by the human autonomic nervous system and endocrine system, which are hardly influenced by subjective factors, thus having the advantage of being objective and realistic compared with other methods [5]. Recent advances in deep learning have demonstrated that ECG is a reliable source for emotion detection [6], [7]. Thus, we use a non-invasive wearable device to capture ECG signals for abnormal emotion detection. The wearable device can monitor an individual’s emotional state over time and will send alerts for health state monitoring or assist in treatment when the individual is chronically in a dysphoric emotion, as shown in Fig. 1.

Currently, various data-driven approaches have been proposed for emotion detection. In spite of all the benefits, compared to the conventional time series anomaly detection (AD) problem, the physiological signal-based abnormal emotion detection problem has its own unique characteristics:

- Waveforms consisting of observations, rather than individual observations, reflect emotional features.
- Normal and abnormal patterns are diverse. Multiple conditions exist for both normal and abnormal waveforms.
- There is significant inter-individual heterogeneity, making the difference between normal waveforms potentially greater than the difference between normal and abnormal waveforms.
- Currently, most physiological signal-based emotion detection protocols are supervised methods, giving some limitations to their application.

To holistically overcome the issues mentioned above, we design a Gaussian mixture convolutional variational autoencoder with data dimension stacking for Unsupervised Abnormal Emotion Detection (UAED) and implement it on wearable devices. To accommodate the first issue mentioned above, we propose a stacking operation to increase the dimension of the one-dimensional ECG signal. The preprocessed data is then fed into a CNN-based VAE network. Along this line, the model's perceptual field will be increased by the stacking operation. Meanwhile, the local temporal invariant of the data is converted into the local spatial invariant, allowing the CNN to efficiently learn the mechanism of the data and facilitate the CNN to capture the multiscale periodic features of the ECG signal. To tackle the second issue, we employ a Gaussian mixture distribution prior in the latent space rather than a single-peaked Gaussian distribution to describe multi-class normal samples appropriately. As for the third issue, we incorporate the whitening distance in the anomaly score, combined with the prior information learned from the Gaussian mixture model (GMM) to enable the model to measure the outliers of the sample under the condition of independent identical distribution. With these operations, the model can be personalized for abnormal emotion detection. Regarding the last issue, we use the VAE framework to perform anomaly detection in an unsupervised mode. The VAE model can learn the deep feature representation and distribution by mapping high-dimensional samples to a low-dimensional latent space and learn the deep feature representation and distribution.

The rest of the paper is organized as follows. Section II summarizes the studies related to emotion detection. Section III presents the network architecture and training steps of UAED. Section IV introduces the experiment procedure and analyzes the experimental result. In final Section V, we conclude the study and describe the future work.

## II. RELATED WORK

### A. Physiological Signals-Based Emotion Detection

Many previous works have highlighted the relationship of physiological signals in multiple contexts such as emotional behavior, social communication, and expression [8], [9], [10]. It has been established that human emotional states are correlated with physiological signals such as EEG and ECG as emotions affect the autonomic nervous system that controls the heartbeat rhythm, and the feasibility of using physiological signals for emotion detection has been analyzed [11]. With the entry of smart wearable devices that can detect human

physiological indicators in real-time, emotion detection based on physiological signals has increasingly become a critical study as well as a spotlight.

One type of emotion detection algorithm based on physiological signals is the classical machine learning algorithm, which extracts artificial features from physiological signals for detection. The works [12] and [13] use the physiological data of individuals in the neutral state for normalization, targeting to overcome inter-individual variability. Nevertheless, this solution relies on the use of individual physiological baseline data, which is hard to apply in practical life. To solve this issue, the stress response factor (SRF) [14] is presented to standardize the data of individuals under free-living conditions and significantly improve the model performance resulting in an average reduction of mean squared error by up to 32%. With the development of deep learning, neural network algorithm based on Big Data gradually becomes the mainstream emotion detection algorithm and makes the detection effect further improved. In [15] and [16], the authors used smart wearable sensors to collect a variety of physiological signals and used CNN with multichannel 1D convolution for mental state detection. The drowsiness detection system in [17] is based on EEG signal for data augmentation and CNN architecture for drowsiness detection, reporting an accuracy of 90.42%. Anusha et al. [18] proposed an automatic preoperative patient stress detection method based on Electrodermal activity (EDA) data using data partitioning-based localized supervised learning algorithm to eliminate inter-individual differences. While the work in [19] used multiple autoencoders to handle unsupervised detection of abnormalities for multimodal data. The model achieved an accuracy of 91.5% for classifying anomalies.

### B. Unsupervised Anomaly Detection

The goal of anomaly detection is to detect rare events (treated as anomalies) that stray from the majority event distribution under imbalanced data. The scarcity and heterogeneous-unexpected nature of anomalies pose a vital challenge to the labeling work [20]. Hence, unsupervised AD methods with the assumption of only normal samples available in the training set become a mainstream research direction owing to the insensitivity to labels. *One-class classification methods* are an intuitive solution that attempts to find the discriminative hyperplane enclosing the normal samples, such as one-class SVM [21] and deep end-to-end one-class classifier [22]. *Density clustering-based methods* intend to find the distribution structure of normal samples to detect points away from clusters as anomalies. The local outlier factor, k-means, and GMM [23], [24], [25] are classical density clustering-based methods. However, most of these methods typically suffer from suboptimal performance when dealing with high-dimensional data.

The philosophy behind *reconstruction-based methods* is that models learning on normal samples cannot reconstruct abnormal samples that did not occur during training. Deep autoencoder (DAE) is a popular reconstruction-based method that extracts tractable representations from high-dimensional data avoiding the curse of dimensionality, including DNN [26], LSTM [27], and VAE [28]. The *GAN-based methods* uses adversarial

training to learn the manifold of normal and achieved excellent performance, such as AnoGAN [29], GANomaly [30] and GAN ensembles [31].

Noteworthy that the work in this article differs from the previous mainly in two-folds. On the one hand, although many studies have been conducted on physiological signals-based emotion detection, most of them treat it as a supervised classification problem limited by data labeling and requiring the fusion of multimodal physiological signals to obtain the desired results [15], [16]. Since abnormal emotions significantly affect human health and endanger social security, it is necessary to pay more attention to the less frequent abnormal emotions compared to normal ones. Thus, we focus on the abnormal emotion detection problem in this work rather than the emotion classification problem. On the other hand, as summarized earlier, the problem of abnormal emotion detection based on physiological signals has its unique issues compared with the conventional anomaly detection task, which will render existing anomaly detection algorithms unable to handle it well. In addition, the existing models such as [7] are complex and difficult to run on mobile devices.

To fill the research gap, this article presents an unsupervised abnormal emotion detection scheme based on physiological signals and applies it to wearable devices. Deep learning algorithms are applied to create an intelligent health maintenance system. To the best of our knowledge, we are the first to conduct physiological signal-based unsupervised abnormal emotion detection tasks and show impressive performance. Notably, this work is an extension of our work in [28]. The main difference is that the long short-term memory (LSTM) based AD model presented in the previous work uses one-dimensional perceptual fields to model time series, whereas the CNNs in this work not only expand the perceptual field of the model but also capture the multiscale periodic features of time series. Besides, we design a low-cost wearable device to capture physiological signals and successfully deploy UAED on mobile devices.

To summarize, our contributions are highlighted as follows:

- 1) We propose a novel framework for abnormal emotion detection. As we have known, UAED is the first unsupervised abnormal emotion detection method based on physiological signals, which makes 2D convolution possible to extract the multiscale temporal dependence of physiological signals by stacking operation. This idea is not only applicable to the AD task in this article, but also to other tasks on time series.
- 2) We design a VAE framework with Gaussian mixture priors combined with whitening anomaly score to learn robust latent representations of complex physiological signals and enable the model to perform anomaly detection without relying on the individual's physiological baseline.
- 3) We achieve state-of-the-art performance on seven standard benchmarks for abnormal emotion detection in four public datasets. We also offer extensive experiment analysis including visualization experiments, ablation experiments and parameter sensitivity experiments to demonstrate the validity of the proposed model.
- 4) We design low-cost wearable devices for ECG acquisition and deployed UAED on mobile devices with 85%

detection accuracy, which validated the effectiveness of the UAED system scheme and the feasibility of practical deployment.

### III. METHODS

In this section, we first provide the statement of the abnormal emotion detection problem. After that, we introduce the overview of our model and detail the UAED including the stacking operation and the Gaussian mixture convolutional variational autoencoders. Finally, we present the variational inference and training steps of UAED.

#### A. Problem Statement

The problem of detecting abnormal emotions based on physiological signals, which is the concern of this article, comes under the issue of time series anomaly detection. Time series contains continuous observations that are usually collected with equally spaced timestamps [32]. In this article, time series is defined as  $\mathbf{x} = \{x_1, x_2, \dots, x_t, \dots, x_n\}$ , where  $n$  is the length of  $\mathbf{x}$ , and  $x_t$  is the observation of the time series  $\mathbf{x}$  at time  $t$  ( $t \leq n$ ). Given a time series  $\mathbf{x}$  with length  $n$ , the subsequence of  $\mathbf{x}$  can be presented as  $\mathbf{x}_{t-m:t} = \{x_{t-m}, x_{t-m+1}, \dots, x_{t-1}\}$ , denoting the consecutive position sampling with length  $m$  from  $t - m$  to  $t$  in  $n$ .

For the abnormal emotion detection problem, our goal is to determine whether any subsequence on the observed time series is abnormal or not. The model models the time series and learns the feature representation of its normal pattern. For simplicity of writing, we denote the subsequence on the time series  $\mathbf{x}$  by  $\tilde{\mathbf{x}}$ . The model returns the anomaly score  $\Psi(\tilde{\mathbf{x}})$ , such that  $\Psi(\tilde{\mathbf{x}}_i) > \Psi(\tilde{\mathbf{x}}_j)$  when  $\tilde{\mathbf{x}}_i$  is anomalous and  $\tilde{\mathbf{x}}_j$  is normal.

#### B. Model Overview

The overall structure of the UAED is shown in Fig. 2. First, the one-dimensional physiological signal is up-dimensioned by stacking operation. Then, the preprocessed data is fed to the model training module to learn the representation and distribution of normal patterns. In the proposed framework, this offline training can be performed periodically. Eventually, the trained model will be stored for online AD. The preprocessed observations are fed into the trained model to obtain its anomaly score and then determine whether it is anomalous by its magnitude. Next, we will present the details of the proposed method.

#### C. Stacking Operation

In this work, our goal is to build a model to capture the mapping relationship between physiological signals and emotional states. Routinely, human emotional states are temporally continuous and do not change abruptly. Therefore, it is more significant to investigate a segment of subsequence on the observation time series than the individual observations. In the data preprocessing process, we introduce a stacking operation, the procedure of which is shown in the red box on the left in Fig. 2. First, the original observations are partitioned into multiple subsequences through a window of length  $T$ , denoted as

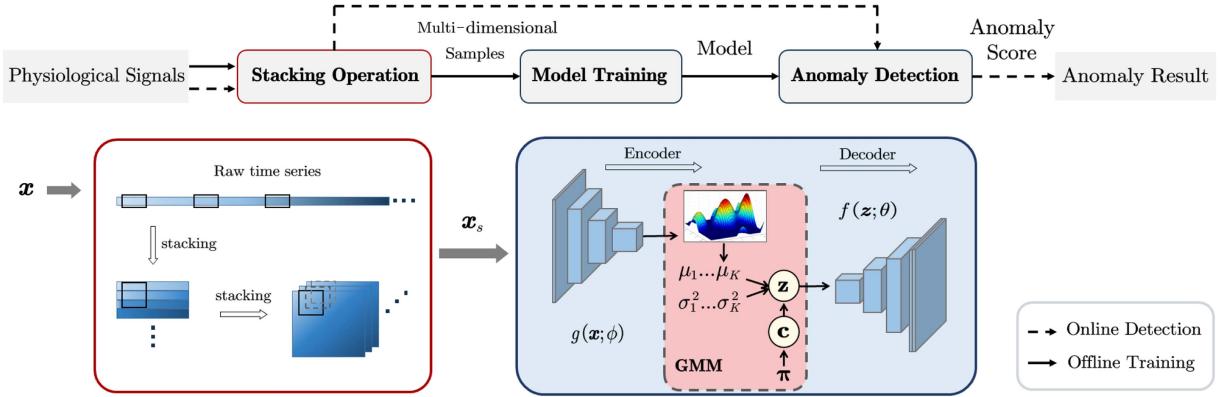


Fig. 2. The workflow of UAED. The red box on the left shows the process of stacking operation, which transform one-dimensional time series into high-dimension to help the model capture the periodic features and reflect diverse normal patterns. Diagram of the proposed Gaussian mixture convolutional variational autoencoders is illustrated in blue box on the right, whose core idea is to reconstruct the input by learning its latent representation thus capturing the normal patterns.

$\mathbf{x}_{t-T:t} = \{x_{t-T}, x_{t-T+1}, \dots, x_{t-1}\}$ , where  $t = T, 2T, \dots, sT$ , and  $sT \leq n + 1$ . The subsequences are stacked along the column direction to form a matrix represented as (1), where  $\mathbf{x}_{s,T}^0$  indicates the first sample output by the stacking operation with the dimension of  $s \times T$ .

$$\begin{aligned} \mathbf{x}_{s,T}^0 &= [\mathbf{x}_{0:T}, \mathbf{x}_{T:2T}, \dots, \mathbf{x}_{(s-1)T:sT}]^T \\ &= \begin{bmatrix} x_0, x_1, \dots, x_{T-1} \\ x_T, x_{T+1}, \dots, x_{2T-1} \\ \vdots \\ x_{(s-1)T}, x_{(s-1)T+1}, \dots, x_{sT-1} \end{bmatrix}_{s \times T} \end{aligned} \quad (1)$$

On this basis, multiple  $s \times T$  matrices can be formed by sliding the window on  $\mathbf{x}$  with stride  $p$ , thus realizing the dimension increase of the one-dimensional time series. The stride  $p$  denotes that the first observation instant of each sample is delayed  $p$  from the previous sample. In practice, to reduce the impact of the skipping problem near the cut point, we repeat the observations of length  $q$  at the beginning of each time of cutting the subsequence as shown in (2). The overlap  $q$  allows stacking operation without destroying the continuity of the original time series and provides the model with more training data.

$$\mathbf{x}_{s,T}^0 = [\mathbf{x}_{0:T}, \mathbf{x}_{T-q:2T-q}, \dots, \mathbf{x}_{(s-1)T-(s-1)q:sT-(s-1)q}]^T \quad (2)$$

$$\mathbf{x}_{s,T}^m = \begin{bmatrix} \mathbf{x}_{mp:T+mp} \\ \mathbf{x}_{T-q+mp:2T-q+mp} \\ \vdots \\ \mathbf{x}_{(s-1)T-(s-1)q+mp:sT-(s-1)q+mp} \end{bmatrix}_{s \times T} \quad (3)$$

Eventually, the one-dimensional signal  $\mathbf{x}$  is converted into  $[\mathbf{x}_{s,T}^0, \mathbf{x}_{s,T}^1, \dots, \mathbf{x}_{s,T}^m, \dots, \mathbf{x}_{s,T}^N]$  of  $N \times s \times T$  by the stacking operation, where  $\mathbf{x}_{s,T}^m$  can be represented as (3),  $N$  is the times of window sliding and  $sT - (s-1)q + Np \leq n$ .

#### D. Gaussian Mixture Convolutional Variational Autoencoders

We denote the observation space and latent space by  $\mathcal{X}$  and  $\mathcal{Z}$ , respectively. For simplicity of writing, the preprocessed observations via the stacking operation and latent variables are

simplified by  $\mathbf{x}$  and  $\mathbf{z}$  hereafter, respectively, where  $\mathbf{x} \in \mathcal{X}$  and  $\mathbf{z} \in \mathcal{Z}$ . VAE consists of an encoder  $g(\mathbf{x}; \phi) : \mathcal{X} \mapsto \mathcal{Z}$  and a decoder  $f(\mathbf{z}; \theta) : \mathcal{Z} \mapsto \mathcal{X}$ , which implements the representation of a high-dimensional input sample  $\mathbf{x}$  into a low-dimensional potential representation  $\mathbf{z}$ , and then reconstructs the original input by using  $\mathbf{z}$ , where  $\phi$  and  $\theta$  are the weight parameters of the encoder and decoder, respectively.

Specifically, inspired by the idea that CNNs can effectively extract image shape features [33], we introduce 2D convolutional layers in both the encoder and decoder of VAE to achieve multi-scale periodic feature extraction. The stacking process ingeniously converts the temporal correlation of signals into spatial correlation, and the generated matrices can be considered as images, which enables CNNs that specialize in image processing to easily extract periodic features of time-series signals. In addition, the proposed stacking operation combined with 2D convolution allows the model to have a larger perceptual field, which helps the network to model the long-range dependence of the time series.

VAE can be trained with the log-evidence lower bound (ELBO) by stochastic gradient variational estimation (SGVB) [34], which is given by:

$$\begin{aligned} \mathcal{L}_{ELBO} (\mathbf{x}, \theta, \phi) &= \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} [\log p_\theta(\mathbf{x}|\mathbf{z})] \\ &\quad - D_{KL}(q_\phi(\mathbf{z}|\mathbf{x}) \| p_\theta(\mathbf{z})) \end{aligned} \quad (4)$$

where  $D_{KL}$  represents the Kullback-Leibler divergence measure.

Conventional VAEs use an isotropic Gaussian model as the prior distribution  $p_\theta(\mathbf{z})$ . However, when the distribution is not a single cluster, this unimodal prior does not describe the multi-peaked representation well. To this end, we make the Gaussian mixture distribution as the prior of the VAE latent space in this article to better describe this problem.

#### E. Analysis of Variational Lower Bound of UAED

Suppose that the Gaussian distribution has  $K$  clusters where  $K$  is a predefined hyperparameter. Considering that each sample only belongs to one cluster, thus we introduce a discrete random

variable  $c$  denoting the cluster the sample belong to and  $c \in \{1, \dots, K\}$ . Assuming that  $c$  follows a categorical distribution with parameter  $\pi$ , that is,  $c \sim \text{Cat}(\pi)$ , where  $\pi \in \mathbb{R}_+^K$  and  $\sum_{k=1}^K \pi_k = 1$ .

In the VAE, the encoder learns the posterior distribution  $q_\phi(z|x)$  and the decoder learns the likelihood distribution  $p_\theta(x|z)$ , where  $\theta, \phi$  are the variational parameter of VAE decoder and encoder. For simplicity, we simplify  $p_\theta(\cdot), q_\phi(\cdot)$  to  $p(\cdot), q(\cdot)$  hereafter. To this end, the log-likelihood of UAED can be expressed as:

$$\begin{aligned} \log p(\mathbf{x}) &= \sum_c \int_z q(z, c|\mathbf{x}) \log p(\mathbf{x}) dz \\ &= \mathbb{E}_{q(z, c|\mathbf{x})} \left[ \log \frac{p(\mathbf{x}, z, c)}{q(z, c|\mathbf{x})} \right] \\ &\quad + D_{KL}(q(z, c|\mathbf{x}) \| p(z, c|\mathbf{x})) \end{aligned} \quad (5)$$

where  $D_{KL}(\cdot)$  makes the variational posterior probability  $q(z, c|\mathbf{x})$  approximate the true posterior probability  $p(z, c|\mathbf{x})$ . As  $D_{KL}(q(z, c|\mathbf{x}) \| p(z, c|\mathbf{x})) \geq 0$ , it follows that

$$\log p(\mathbf{x}) \geq \mathbb{E}_{q(z, c|\mathbf{x})} \left[ \log \frac{p(\mathbf{x}, z, c)}{q(z, c|\mathbf{x})} \right] = \mathcal{L}_{ELBO}(\mathbf{x}) \quad (6)$$

To this end, we obtain the evidence lower bound (ELBO) of UAED. Based on Bayes theorem, we can decompose the generative process as:

$$p(\mathbf{x}, z, c) = p(\mathbf{x}|z) p(z|c) p(c) \quad (7)$$

Furthermore, assume that  $q(z, c|\mathbf{x})$  is a mean-field distribution. Thus, we can obtain that

$$q(z, c|\mathbf{x}) = q(z|\mathbf{x}) q(c|\mathbf{x}) \quad (8)$$

where  $q(c|\mathbf{x})$  gives the cluster assignment probabilities and  $\sum_{c=1}^K q(c|\mathbf{x}) = 1$ .

Depending on (7) and (8),  $\mathcal{L}_{ELBO}$  in (6) can be rewritten in the following form:

$$\begin{aligned} \mathcal{L}_{ELBO}(\mathbf{x}) &= \sum_c \int_z q(z, c|\mathbf{x}) [\log p(\mathbf{x}, z, c) \\ &\quad - \log q(z, c|\mathbf{x})] dz \\ &= \sum_c q(c|\mathbf{x}) \int_z q(z|\mathbf{x}) \left[ \log \frac{p(\mathbf{x}|z) p(z)}{q(z|\mathbf{x})} \right] dz \\ &\quad + \int_z q(z|\mathbf{x}) \left[ \sum_c \int_z q(c|\mathbf{x}) \left[ \log \frac{p(c|z)}{q(c|\mathbf{x})} \right] dz \right] dz \\ &= \int_z q(z|\mathbf{x}) \left[ \log \frac{p(\mathbf{x}|z) p(z)}{q(z|\mathbf{x})} \right] dz \\ &\quad - \int_z q(z|\mathbf{x}) D_{KL}(q(c|\mathbf{x}) \| p(c|z)) dz \end{aligned} \quad (9)$$

where the former is not related to  $c$  and the latter is non-negative. Thus,  $\mathcal{L}_{ELBO}$  is maximum when  $D_{KL}(q(c|\mathbf{x}) \| p(c|z)) \equiv 0$ .

Then, we can calculate the  $q(c|\mathbf{x})$  term in (9) by sampling  $z^{(l)} \sim q(z|\mathbf{x})$  as shown in (10).

$$q(c|\mathbf{x}) = p(c|\mathbf{z}^{(l)}) = \frac{p(c)p(\mathbf{z}^{(l)}|c)}{\sum_{\hat{c}} p(\hat{c}) p(\mathbf{z}^{(l)}|\hat{c})} \quad (10)$$

Thus, the clustering results can be obtained by the Bayesian formula without participating in gradient descent. More intuitively, the (6) can be rewritten in a form similar to the regular VAE's objective function in (4) as follows:

$$\begin{aligned} \mathcal{L}_{ELBO}(\mathbf{x}) &= \mathbb{E}_{q(z, c|\mathbf{x})} [\log p(\mathbf{x}|z)] \\ &\quad - D_{KL}(q(z, c|\mathbf{x}) \| p(z, c)) \end{aligned} \quad (11)$$

#### F. UAED for Time Series Anomaly Detection

In this subsection, we will demonstrate in detail how to use the UAED model to detect anomalies. During detection, each sample  $\mathbf{x}$  is fed into the encoder to obtain the corresponding mean vectors and standard deviation vectors  $\mu_{z|x,c}, \sigma_{z|x,c}, c \in \{1, \dots, K\}$ . There are  $K$  sets of mean and standard deviation vectors in the latent space of the Gaussian mixture VAE corresponding to  $K$  clusters in the latent space, respectively. We set  $\pi_k = K^{-1}$  to allow  $c \sim \text{Cat}(\pi)$  to be a uniform categorical distribution. A component is selected among the  $K$  components in the GMM according to (10), and then one can determine the latent Gaussian distribution of the sample. Moreover, the Gaussian distribution is sampled  $P$  times to obtain  $P$  latent vectors, and we denote the  $p$ th latent vector by  $\mathbf{z}_p \sim \mathcal{N}(\mu_{z|x,c^*}, \sigma_{z|x,c^*}^2)$ ,  $p = 1, \dots, P$ . The  $P$  latent vectors are reconstruct by the decoder, thus we can obtain the  $P$  corresponding reconstructed mean vectors  $\mu_{\hat{x}|z,c^*}^{(p)}$  and reconstructed standard deviation vectors  $\sigma_{\hat{x}|z,c^*}^{(p)}$ . Then, fitting the input  $\mathbf{x}$  to the Gaussian distribution of the reconstructed mean vector  $\mu_{\hat{x}|z,c^*}^{(p)}$  and the reconstructed standard deviation vector  $\sigma_{\hat{x}|z,c^*}^{(p)}$  learned by UAED, we can obtain the corresponding reconstruction probability  $\mathcal{N}(\mathbf{x}|\mu_{\hat{x}|z,c^*}^{(p)}, \sigma_{\hat{x}|z,c^*}^{(p)})$ . Finally, the reconstruction probability of input  $\mathbf{x}$  is obtained by averaging the reconstruction probabilities corresponding to all  $P$  latent vectors as follows.

$$\mathcal{S}_{RP}(\mathbf{x}) = \frac{1}{P} \sum_{p=1}^P \mathcal{N}(\mathbf{x}|\mu_{\hat{x}|z,c^*}^{(p)}, \sigma_{\hat{x}|z,c^*}^{(p)}) \quad (12)$$

Moreover, we measure the extent of abnormality of a sample using the dispersion of the sample from the corresponding cluster, which we refer to as the whitening distance  $\mathcal{D}_w$  as shown in (13). The whitening distance takes into account the diversity of normal patterns and overcomes intra-class variation allowing UAED to accurately detect abnormalities.

$$\mathcal{D}_w(\mathbf{x}) = \frac{1}{P} \sum_{p=1}^P \frac{\left( \mathbf{x} - \mu_{\hat{x}|z,c^*}^{(p)} \right)^2}{\left( \sigma_{\hat{x}|z,c^*}^{(p)} \right)^2} \quad (13)$$

Integrating the reconstruction probability and whitening distance, the anomaly score of UAED can be denoted as:

$$\mathcal{S} = \mathcal{S}_{RP}(\mathbf{x}) - \lambda \cdot \mathcal{D}_w(\mathbf{x}) \quad (14)$$

**Algorithm 1:** UAED for Unsupervised Anomaly Detection.

---

**Input:** Unsupervised training set  $\mathbf{x}_{train} = \{\mathbf{x}\}_{t=1}^{N_{train}}$ , Test set  $\mathbf{x}_{test} = \{\mathbf{x}\}_{t=1}^{N_{test}}$ , Number of clusters  $K$ ;

**Output:** Detection results  $S(\mathbf{x})$ ;

- 1: Initialize variational parameter of UAED encoder and decoder  $\theta, \phi$ ;
- 2: Train the UAED by  $\mathbf{x}_{train}$  under  $\mathcal{L}_{ELBO}$  in (11);
- 3: **for**  $t = 1$  to  $N_{test}$  **do**
- 4: Feed  $\mathbf{x}$  into the encoder to yield  
 $\mu_{\mathbf{x}|c}, \sigma_{\mathbf{x}|c}^2 = g_\phi(\mathbf{x}; c)$   
 $c = 1, \dots, K$ ;
- 5: Sample a  $c$  on the prior categorical distribution  $c^* \sim Cat(\pi)$ ;
- 6: Sample  $P$  times on the Gaussian distribution yields  $P$  latent vectors corresponding to the samples  $\mathbf{z}_p \sim \mathcal{N}(\mu_{\mathbf{z}|\mathbf{x}, c^*}, \sigma_{\mathbf{z}|\mathbf{x}, c^*}^2)$   $p = 1, \dots, P$ ;
- 7: **for**  $j = 1$  to  $P$  **do**
- 8: Feed the latent vector into the decoder to derive the reconstructed mean vector, reconstructed standard deviation vector  $\mu_{\hat{\mathbf{x}}|\mathbf{z}, c^*}^{(j)}, \sigma_{\hat{\mathbf{x}}|\mathbf{z}, c^*}^{(j)} = f_\theta(\mathbf{z}_j; c^*)$ ;
- 9: **end for**
- 10: Obtain the anomaly score based on (14);
- 11: **if**  $S < \alpha$  **then**
- 12:      $S(\mathbf{x}) = 1$ ,  $\mathbf{x}$  is anomaly
- 13: **else**
- 14:      $S(\mathbf{x}) = 0$ ,  $\mathbf{x}$  is normal
- 15: **end if**
- 16: **end for**

---

where  $\lambda$  is a trade-off hyperparameter controlling the weight of the two parts in the anomaly score, and  $\lambda > 0$ . We recommend a range of 0.1 to 0.5 for  $\lambda$ . More details could be found in the Experiment of parameter sensitivity test. The anomalous is outlying and difficult to reconstruct, which will result in having a small reconstruction probability and a large whitening distance. Then it is obvious that its corresponding anomaly score will be low. To this end, UAED considers samples with abnormal scores less than the threshold as abnormal samples. The method of setting the threshold value is not the focus of this article. In this work, we assume that the anomaly ratios of the data are known, and set threshold  $\alpha$  based on the anomaly ratios and the ranking of the anomaly scores. Therefore, the procedure of UAED is outlined as shown in Algorithm 1.

#### IV. EXPERIMENTS AND RESULTS

In this section, we will evaluate the performance of UAED through two scenarios of experiments. The first is conducted on four public datasets, comparing the results with state-of-the-art baseline models, aiming to validate the effectiveness of the proposed method. The second one is based on our self-developed wearable ECG acquisition device to evaluate the detection performance of UAED in real-world scenarios. This study and the experimental procedures are approved by the Medical and Laboratory Animal Ethics Committee of Beijing University of Technology (approval no.58).

#### A. Dataset

In studies related to emotion model theory, there are two main approaches to represent the emotion states [39]: the discrete categorical model and the continuous dimensional model. The discrete categorical model represents emotion states as several categories. While the continuous dimensional model represents all emotion states in a low-dimensional space.

In this work, we used four publicly available datasets to deeply evaluate the performance of the proposed framework in abnormal emotion detection, covering the two emotion representation methods mentioned above. Detailed information about the datasets is summarized in Table I, and a brief description of each dataset is given below.

1) *DREAMER* [35]: DREAMER is a dataset designed for emotion recognition using a low-cost device and consists of EEG and ECG readings recorded by exposure to sensors from 23 participants. ECG signals were captured using a SHIMMER sensor [40] at a sampling rate of 256 Hz. Movie clips were used to stimulate different emotions in participants, and at the end of each clip participants scored their self-assessment on a 5-point arousal, valence, and dominance scale. In this experiment, we treat states with a valence of 1 as abnormal and the rest of the emotion states as normal.

2) *DRIVEDB* [36]: The Stress Recognition in Automobile Drivers database (DRIVEDB) is a multimodal dataset that records drivers' stress levels in different environments taken from the online public database in PhysioNet.<sup>1</sup> The dataset is designed to capture naturally occurring emotions rather than emotions elicited in a laboratory setting. ECG data collected from a custom wearable system for 16 different participants used in this study were sampled at 496 Hz. In this experiment, we treat states with high stress as abnormal and the remaining states as normal.

3) *MAHNOB-HCI* [37]: Mahnob-HCI-tagging database (MAHNOB-HCI) was collected using the Biosemi active II system with active electrodes to capture physiological signals such as ECG, EEG, respiratory amplitude, and skin temperature from 27 participants at a sampling rate of 256 Hz. Participants were asked to record their emotion status after each trial, including emotional label, arousal, valence, dominance, and predictability both on a scale from 1 to 9. In this experiment, we refer to the emotional label to treat the anxiety state marked as 12 as abnormal and the rest of the emotional states as normal.

4) *WESAD* [38]: Wearable Stress and Affect Detection (WESAD) is a multimodal dataset for stress and emotion detection. A RespiBAN Professional<sup>2</sup> sensor worn on the chest was used to acquire ECG signals from 15 participants at a sampling rate of 700 Hz with the aim of studying four different emotional states, namely neutral, stress, amusement, and meditation. In this experiment, we treated the stress state as abnormal and the rest of the affective states as normal.

<sup>1</sup>PhysioNet: The Research Resource for Complex Physiologic Signals. URL <https://www.physionet.org/>

<sup>2</sup><https://www.pluxbiosignals.com/en/respiban-professional>

TABLE I  
DESCRIPTION OF THE PUBLIC DATASETS

Dataset	Participants	Attributes	#Instances <sup>1</sup>	#Channels	Sampling Rate	Window Size	#Processed Samples	Outliers ( $\mu\%$ ) <sup>2</sup>
DREAMER [35]	23	A, V, D <sup>3</sup>	15269888	2	256Hz	96	3313	valence=1(13.88%)
DRIVEDB [36]	16	Stress level	967520	1	496Hz	24	3359	high stress(3.78%)
MAHNOB-HCI [37]	27	Emotional label	11447745	3	256Hz	48	9937	anxiety(1.29%)
WESAD [38]	15	Emotional label	31470603	1	700Hz	96	13659	stress(22.16%)

<sup>1</sup> Number of instances in the original dataset without stacking operations.

<sup>2</sup> The abnormal classes and anomaly ratios.

<sup>3</sup> Arousal, Valence, Dominance.

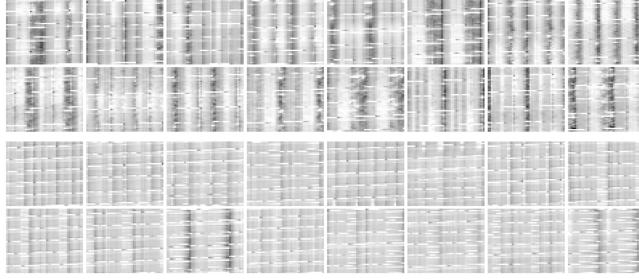


Fig. 3. Visualization of samples from DREAMER dataset processed by stacking operation. The top two rows are normal samples and the bottom two rows are abnormal samples.

### B. Pre-Processing

Although multimodal signal fusion achieves desirable results for many tasks, it often requires more sensors, more complex networks, and more computational resources. Considering the requirement of economy and comfort for daily life applications, for the task of abnormal emotion detection based on wearable devices, we prefer to design networks that can obtain high accuracy detection results based on simple inputs at the same time. Therefore, in this work we use a unimodal ECG signal as the input to the network.

Following [7], [41], we only do simple data cleaning including removal of incomplete data and normalization to better understand the learning ability of UAED for physiological signal feature representation. We map the observations in the same range of [0,1] for each dataset by min-max normalization. The data is then fed for stacking operation, and the processing results for each dataset are shown in Table I.

After the stacking operation, we transformed the original 1D time series into 3D samples, and the visualization is presented in Fig. 3. It can be noticed that: 1. The stacking operation embeds the sequence into the image while retaining the original temporal information. 2. The periodic pattern of the time series can be intuitively reflected in the images. 3. The difference between normal and abnormal samples after stacking is more obvious, which indicates that the stacking operation can better reflect the abnormal pattern of the time series and will help the detection of abnormal states.

### C. Network Structure

CNNs have been widely used for image recognition and detection as an effective image processing and feature extraction

TABLE II  
THE NETWORK STRUCTURE ADOPTED FOR EACH DATASET

Dataset	Backbone Structure
DREAMER [35]	<i>Conv2_1(8, 11, 2)-Conv2_2(16, 5, 2)-Conv2_3(32, 5, 2)-Conv2_4(64, 5, 2)</i>
	<i>Dconv2_1(32, 5, 2)-Dconv2_2(16, 5, 2)-Dconv2_3(8, 5, 2)-Dconv2_4(2, 11, 2)</i>
DRIVEDB [36]	<i>Conv2_1(8, 5, 2)-Conv2_2(16, 3, 2)-Conv2_3(32, 3, 1)</i>
	<i>Dconv2_1(16, 3, 1)-Dconv2_2(8, 3, 2)-Dconv2_3(1, 5, 2)</i>
MAHNOB-HCI [37]	<i>Conv2_1(32, 5, 2)-Conv2_2(32, 3, 2)-Conv2_3(64, 3, 2)-Conv2_4(64, 3, 1)</i>
	<i>Dconv2_1(64, 3, 1)-Dconv2_2(32, 3, 2)-Dconv2_3(32, 3, 2)-Dconv2_4(3, 5, 2)</i>
WESAD [38]	<i>Conv2_1(8, 11, 2)-Conv2_2(16, 5, 2)-Conv2_3(32, 5, 2)-Conv2_4(64, 5, 2)</i>
	<i>Dconv2_1(32, 5, 2)-Dconv2_2(16, 5, 2)-Dconv2_3(8, 5, 2)-Dconv2_4(1, 11, 2)</i>

method. However, conventional deep CNN models are not suitable for wearable devices [42] and cloud computing scenarios [43] due to their large number of learnable parameters and excessive requirements of memory and computational resources for arithmetic operations. For abnormal emotion detection tasks based on wearable devices, we want to design a lightweight CNN that can guarantee accuracy simultaneously.

In this work, the architecture of VAE includes an encoder consisting mainly of several convolutional layers as well as a fully connected layer and a decoder whose structure is paired with the encoder. The schematic diagram of the network structure is shown in Fig. 4, and the specific network structure used for each dataset is shown in Table II. Where, *Conv2\_L(c, k, s)* and *Dconv2\_L(c, k, s)* denote the 2D convolution layer and 2D deconvolution layer, L denotes the number of layers, c,k,s denotes the number of channels, kernel size, stride size, respectively. Except for the last layer of the encoder and decoder, the remaining convolutional and deconvolutional layers are followed by a BatchNormalization (BN) layer and a RELU activation layer to reduce the effect of overfitting and help the network converge.

### D. Evaluation Metrics

We use the Area Under Receiver Operating Characteristic Curve (AUC-ROC), the Area Under Precision-Recall Curve (AUC-PR), Mean Intersection over Union (MIoU) and F1 score to compare AD performance intuitively. AUC-ROC is the area under ROC curve, which plots with the false-negative (FN) rate

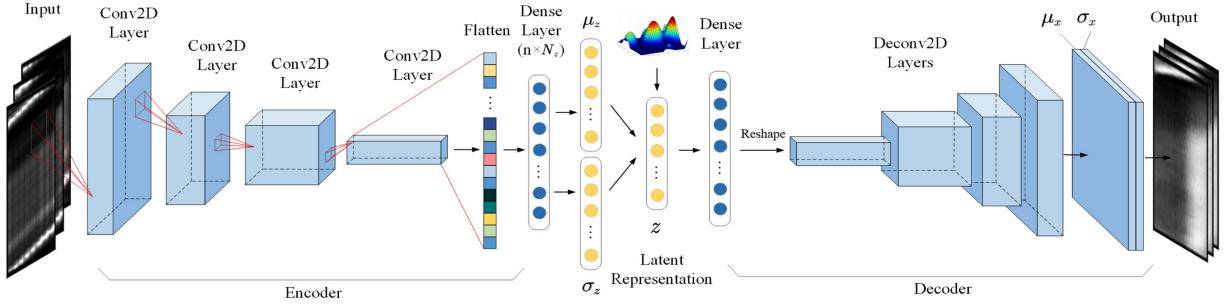


Fig. 4. The network structure of UAED.

as the x-axis and true-positive (TP) rate as the y-axis at various threshold settings. AUC-PR is the area under PR curve, which plots the precision against recall at various thresholds. Both AUC-ROC and AUC-PR are immune to thresholds and thus provide an objective measure of model performance. MIoU is the average of the ratio of intersection and union of detection and reality for each class. In (15),  $p_{ij}$  and  $k + 1$  denotes the number of samples of category  $i$  detected as category  $j$  and the total number of categories, respectively. In this task, there are only two classes, abnormal and normal, thus MIoU can be calculated as  $\frac{1}{2}(\frac{TP}{TP+FP+FN} + \frac{TN}{TN+FN+FP})$ , where FP and TN denote false-positive rate and true-negative rate, respectively. F1 score as a harmonic mean of precision and recall takes both into account, which is expressed as (16). The *precision* and *recall* can be calculated as  $precision = \frac{TP}{TP+FP}$  and  $recall = \frac{TP}{TP+FN}$ , respectively. The values of all four evaluation metrics range in the interval of [0,1]. The higher value of the evaluation metrics indicates better detection results.

$$MIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}} \quad (15)$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \quad (16)$$

### E. Baseline Methods

We employ the following seven baselines to compare with our model.

Isolation forest (ISF) [44], which employs a random segmentation strategy that understand anomalous samples as points that are spread sparsely and away from a dense cluster.

One class SVM (OCSVM) [21], which maps data to a high-dimensional feature space using a kernel function in which an optimum hyperplane is identified to enable maximum separation of the target data from the coordinate origin. We employ the widely adopted radial basis function (RBF) kernel in the experiment.

LSTM-based autoencoder (AE-LSTM) [27], which uses a classical LSTM-based encoder-decoder framework for unsupervised learning with adaptive thresholding of performance metrics based on support vector regression (SVR).

Deep autoencoding gaussian mixture model (DAGMM) [45], which integrates the autoencoder with a GMM with end-to-end

joint optimization of parameters. The model is implemented using both fully connected layers and LSTM layers, respectively.

Deep support vector data description (DSVDD) [46], which locates a hypersphere in the high-dimensional feature space that encloses as many target sample points as possible while minimizing its volume.

LSTM-based Variational autoencoder (VAE-LSTM) [47], which uses LSTM-based encoder and regularizes the data distribution in the encoding space enables the model to obtain a more efficient predictive likelihood function and a tighter log-likelihood lower bound.

Detect outlier machine instances (DOMI) [48], which uses a 1D-CNN based Gaussian mixture VAE framework and detects anomalies based on reconstruction probabilities.

For apple-to-apple comparison, all the baselines are unsupervised AD methods. Among them, ISF and OCSVM are shallow models, while all other methods are deep neural networks. For fairness, all deep models share the same infrastructure as possible with UAED. Also, we use LSTM layer instead of the fully connected layer in the paper for the experiments for DAGMM, and VAE methods. We implement the baselines using open-source packages or the code of the original paper and give preference on the parameter settings recommended by the original paper. Adam [49] is used as an optimizer in all models.

### F. Experimental Results and Analysis

Table III shows the performance of UAED and seven baseline methods on four datasets DREAMER, DRIVE DB, MAHNOB-HCI, and WESAD, where the best results are marked in bold. The reported AUC-ROC, AUC-PR, MIoU and F1 score are the average of a nested cross-validation leave-one-subject-out (LOSO) procedure. Concretely, we leave the data from one subject at a time as the test set and the rest of the data as the training set, thus avoiding overfitting. Further, we perform a 10-fold cross-validation on the training set as an inner loop to determine the optimal parameters of the model and we will perform further sensitivity analysis on the parameters in Section IV-I. Taking DREAMER as an example, it contains data from 23 participants, so we leave one participant's data as test set. For the other 22 participants' data, we divide them into 10 parts and rotate 9 of them as training data and 1 of them as validation data to determine the hyperparameters. For different datasets, the ratio of the training set was in the range of 86%–96%, with

TABLE III  
PERFORMANCE OF UAED AND SEVEN BASELINES ON FOUR DATASETS DREAMER, DRIVEDB, MAHNOB-HCI, AND WESAD

Methods	Metric	DREAMER	DRIVEDB	MAHNOB-HCI	WESAD	Average	Improvement
ISF [44]	AUC-ROC	0.528	0.511	0.652	0.538	0.557	+0.196
	AUC-PR	0.576	0.542	0.736	0.669	0.631	+0.155
	MIoU	0.361	0.311	0.532	0.388	0.403	+0.184
	F1	0.315	0.198	0.682	0.602	0.449	+0.219
OCSVM [21]	AUC-ROC	0.506	0.502	0.674	0.702	0.596	+0.157
	AUC-PR	0.526	0.598	0.812	0.764	0.675	+0.111
	MIoU	0.265	0.255	0.350	0.605	0.369	+0.218
	F1	0.128	0.012	0.556	0.658	0.339	+0.329
AE-LSTM [27]	AUC-ROC	0.656	0.603	0.723	0.580	0.641	+0.112
	AUC-PR	0.686	0.659	0.832	0.675	0.713	+0.073
	MIoU	0.462	0.431	0.586	0.525	0.501	+0.086
	F1	0.598	0.517	0.620	0.416	0.538	+0.130
DAGMM [45]	AUC-ROC	0.568	0.529	0.629	0.642	0.592	+0.161
	AUC-PR	0.604	0.668	0.240	0.618	0.533	+0.253
	MIoU	0.362	0.298	0.497	0.457	0.404	+0.183
	F1	0.412	0.210	0.665	0.523	0.453	+0.215
DSVDD [46]	AUC-ROC	0.532	0.500	0.698	0.520	0.563	+0.190
	AUC-PR	0.596	<b>0.718</b>	0.788	0.597	0.675	+0.111
	MIoU	0.318	0.304	0.427	0.405	0.364	+0.223
	F1	0.482	0.472	0.632	0.518	0.526	+0.142
VAE-LSTM [47]	AUC-ROC	0.652	0.580	0.702	0.604	0.635	+0.118
	AUC-PR	0.584	0.651	<b>0.850</b>	0.695	0.695	+0.091
	MIoU	0.490	0.428	<b>0.680</b>	0.406	0.501	+0.086
	F1	0.436	0.564	0.606	0.594	0.550	+0.118
DOMI [48]	AUC-ROC	0.668	0.591	<b>0.860</b>	0.688	0.702	+0.051
	AUC-PR	0.725	0.658	0.814	0.741	0.735	+0.051
	MIoU	<b>0.509</b>	0.442	0.636	0.618	0.551	+0.036
	F1	0.564	0.552	0.692	0.659	0.617	+0.051
UAED (Pro.)	AUC-ROC	<b>0.692</b>	<b>0.687</b>	0.852	<b>0.782</b>	0.753	-
	AUC-PR	<b>0.769</b>	0.708	0.849	<b>0.816</b>	0.786	-
	MIoU	0.496	<b>0.492</b>	0.665	<b>0.694</b>	0.587	-
	F1	0.546	<b>0.612</b>	<b>0.726</b>	<b>0.789</b>	0.668	-

an average of 92%. The CPU of the experimental device is Intel Core i7-8700@3.20 GHz and the GPU is NVIDIA Quadro P400.

From Table III, we can draw the following conclusions. 1. The average performance of UAED on the four publicly datasets outperformed all comparison methods. In particular, the UAED scheme achieves 7.26%, 6.94%, 6.53%, and 8.27% improvement compared to the suboptimal methods in AUC-ROC, AUC-PR, MIoU, and F1 score, respectively. The main reason is that UAED enables 2D convolution to extract periodic features and shape features of physiological signals by stacking operations. Meanwhile, the normal pattern of the physiological signal is successfully captured by the VAE framework with a Gaussian mixture priors. Note that the described stacking operation is a general operation that can be applied to other time series tasks. 2. The reconstruction-based methods achieves robust results on this task compared to the one-class classification methods (e.g., OCSVM and DSVDD) as well as the ISF. The one-class-based approach has been proved that its performance cannot be guaranteed in dataful scenarios, and thus the performance is poor in this task. In contrast, UAED achieves good performance by learning reasonable latent representations. 3. VAE-based methods (e.g., VAE-LSTM, DOMI, and UAED) achieved better results on this task compared to AE-based methods (e.g., AE-LSTM, DAGMM). This is due to the fact that VAE regularizes the sample distribution in the latent space and uses variational inference to enable the model to obtain a more efficient predictive likelihood function and a tighter log-likelihood lower bound. Moreover, both DOMI and UAED use a Gaussian mixture prior rather than an isotropic Gaussian prior, allowing the latent representation learned by the

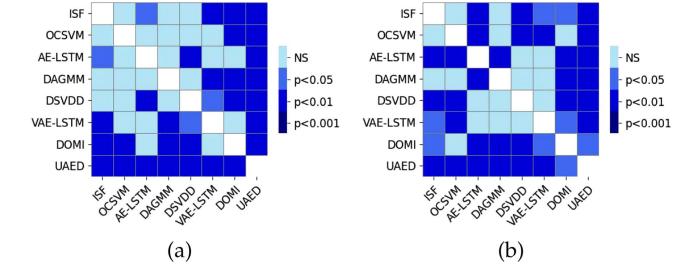


Fig. 5. The heatmap of Nemenyi scores according to AUC-ROC and F1 score.  
(a) AUC-ROC. (b) F1 score.

model to represent complex normal prototype features, which leads to further performance improvements.

Furthermore, we perform a statistical assessment to confirm the advancement and robustness of the UAED. We utilize the Nemenyi test [50], a nonparametric significance test that examines whether there is a significant discrepancy in average performance between pairs of algorithms on multiple datasets. We compare pairs of methods based on 99%, 95%, and 90% confidence intervals. Fig. 5 shows the heatmap of the Nemenyi scores based on AUC-ROC and F1 score, whose last row (or column) indicates that our UAED performs better than all other methods at a 95% confidence level.

To empirically study the computational cost and space complexity of the proposed method, we analysis the floating point operations (FLOPs) and the number of learnable parameters of UAED and several state-of-the-art methods based on deep neural networks. For all methods we test on the WESAD dataset and

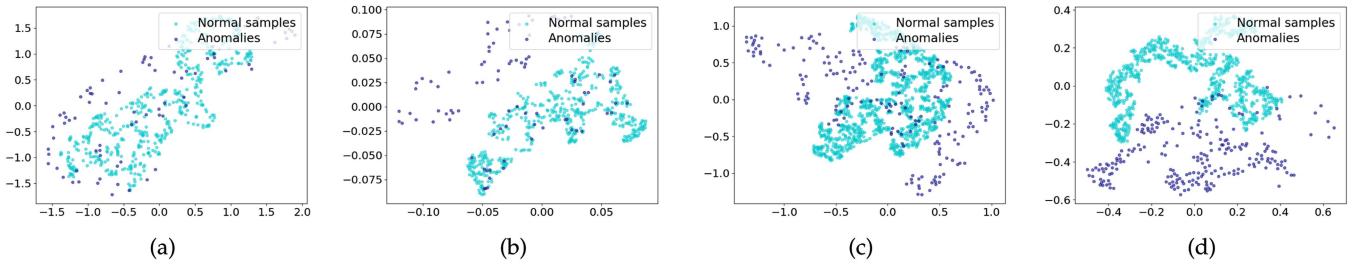


Fig. 6. Visualization of the distribution of the latent representation. (a) UAED with a single Gaussian prior for DREAMER. (b) UAED for DREAMER. (c) UAED with a single Gaussian prior for WESAD. (d) UAED for WESAD.

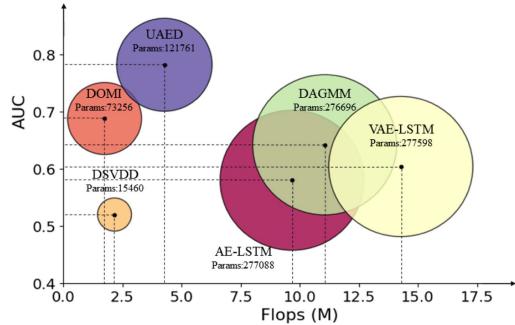


Fig. 7. AUC-ROC, FLOPs, and parameter counts (circle area size) of UAED and several state-of-the-art methods on WESAD dataset.

visualize the results in Fig. 7. It can be found that DSVDD has the lowest number of parameters and the second lowest FLOPs due to its fully connected layer, but suffers from a lower AUC-ROC performance. Whereas UAED achieves the best AUC-ROC performance while occupying a relatively small amount of parameters with high computational efficiency.

#### G. Visual Analysis

We further demonstrate the validity of UAED by visualizing the latent representation extracted by UAED through T-SNE analysis. Fig. 6 shows the representation with an isotropic Gaussian prior and a Gaussian mixture prior (UAED) in the 2D latent space by sampling from  $q_\phi(\mathbf{z}, c|\mathbf{x})$  for the DREAMER and WESAD datasets, respectively. As shown in Fig. 6, UAED has a more distinct demarcation between the normality and abnormality compared to the model with isotropic Gaussian prior, thereby facilitating the detection of abnormalities. In addition, we find that UAED learns a more concentrated distribution of features than the model with isotropic Gaussian prior, suggesting that the Gaussian mixture prior facilitates the convergence of features. Further, we calculate the average distance between features of each individual in the WESAD dataset in the 2D space for both single Gaussian prior and Gaussian mixture prior as 1.285 and 0.027, respectively. It is evident that the UAED can significantly reduce the inter-individual variability.

#### H. Ablation Experiments

In this section, we consider the following variants of UAED as a comparison to demonstrate the necessity of the individual

TABLE IV  
ABLATION EXPERIMENTS RESULTS

Model	Stacking	GM*	Whitening Score	AUC
Variant I	✗	✗	✗	0.621
Variant II	✗	✓	✗	0.636
Variant III	✗	✓	✓	0.670
Variant IV	✓	✗	✗	0.708
Variant V	✓	✓	✗	0.723
UAED	✓	✓	✓	<b>0.753</b>

\* VAE with Gaussian mixture prior

AUC represents the average AUC-ROC of UAED on DREAMER, DRIVEDB, MAHNOB-HCI, and WESAD.

components in UAED. We design four variants of UAED to verify their average performance. In Variants I and II, 1D convolutional layers are used instead of the original 2D convolutional layers. All models are trained in an unsupervised manner and share the same network structure and hyperparameters as the original UAED as much as possible. We take the average effect of LOSO cross-validation on four datasets as shown in Table IV, and the specific performance on each dataset as shown in Fig. 10.

We can find that the experimental results of UAED are significantly better than all its variants on all datasets. Specifically, we can get the following findings. 1. Comparing Variant I with VAE-LSTM in Table III, the performance of 1D convolution is inferior to LSTM for multi-channel datasets in this scenario, while there is no significant difference between the two performances for single-channel data. The LSTM can capture the causal relationship between the order of the temporal data, while the 1D convolution captures the features without order but is better at finding some unusual values of points to detect anomalies. 2. GMM with whitening anomaly score can further improve the model performance, with a 6.36% AUC gain. 3. Comparing Variant IV with Variant I shows that the stacking operation with 2D convolution dramatically improves the model performance, delivering a 14.00% AUC gain, and this effect is more pronounced on multi-channel datasets. To this end, the superiority of each component of UAED in dealing with time series AD is visually verified through the ablation analysis.

#### I. Parameter Sensitivity Test

We further discuss the effect of parameters on UAED performance by selecting different parameter values, including window size  $T$ , overlap  $q$ , stride  $p$ , dimensions of  $\mathbf{z}$ , clusters  $K$  in the Gaussian distribution, and trade-off parameter  $\lambda$ . A

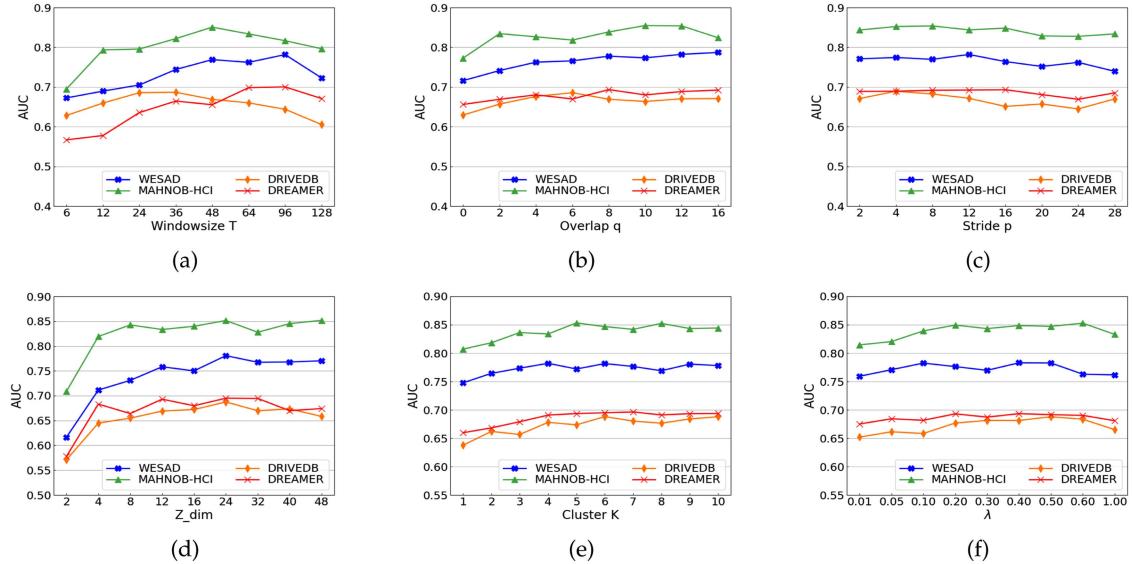


Fig. 8. Sensitivity test for key parameters. (a) Performance for  $T$ . (b) Performance for  $q$ . (c) Performance for  $p$ . (d) Performance for  $z$ . (e) Performance for  $K$ . (f) Performance for  $\lambda$ .

reasonable threshold interval is set for each parameter, while other parameters and network structure remain the same. The performance of UAED with various parameter sets is presented in Fig. 8, where we use four colors to indicate different datasets.

The first column of Fig. 8 displays the performance of parameters  $T$ ,  $q$ , and  $p$ , which are all related to the stacking operation. For the parameter  $T$ , the model performance generally shows a trend of increasing and then decreasing with the increase of windowsize in each dataset. This is attributed to too large a  $T$  that can lead to insufficiently processed datasets, and since the DRIVEDB dataset is relatively small, too large  $T$  will cause UAED's performance on it decrease significantly. On the other hand, too small  $T$  will cause the stacking operation to fail to capture the periodic features of the time series and thus CNN cannot perform well, which will also lead to performance degradation. Since the data structure of the dataset MAHNOB-HCI is relatively complex, the performance of UAED on it degrades significantly when  $T$  is too small. For the parameter  $q$  and  $p$ , the model is more robust when the former  $q > 4$  and the latter is within the interval 2 to 12. When  $q$  is too small, the model performance decreases on all datasets, which may be influenced by the skipping problem near the cutting point.

Subplots (d) and (e) in Fig. 8 display the test results of  $z$  and  $K$ . Adequate dimensions of  $z$  can decrease the impact of dimensionality reduction, whereas too small latent space dimensions may lead to underfitting of the model and thus reduce the ability of  $z$  to reconstruct high-dimensional data. Furthermore,  $K$  characterizes the complexity of the distribution in the latent space. It can be found that the curves are in a smooth trend in most intervals. Specifically, we can see that the model performance becomes stable when  $z > 4$  and  $k > 4$ . Subplots (f) shows the results of the  $\lambda$ , which is a trade-off parameter between reconstruction probability and whitening distance anomaly score. The experimental results show that UAED is highly robust to  $\lambda$ , especially in the interval 0.1 to 0.5,

TABLE V  
AUC-ROC PERFORMANCE OF 5 TIMES NESTED LOSO EXPERIMENTS ON THE WESAD DATASET

Methods	#1	#2	#3	#4	#5	avg	SD
DOMI	0.613	0.728	0.756	0.699	0.636	0.686	0.054
UAED	0.776	0.782	0.806	0.772	0.775	0.782	0.012

avg and SD represent the average and standard deviation of the AUC-ROC performance, respectively.

so we recommend preferentially taking values in this interval when there is no prior knowledge.

Overall, UAED is robust to all six hyperparameters within a large interval, thus the researcher does not need to worry about their values affecting the UAED's performance.

### J. Stability Analysis

Since anomalies often lead to serious consequences, the anomaly detection task is highly concerned with the stability of the model. In this section, we analyze the stability of the model from three aspects: 1. the convergence of the algorithm, 2. model performance variation in one training session, 3. the stability of model performance in independent training sessions. Fig. 9 extensively shows the convergence and the change of the AUC-ROC performance of the model on different datasets during one training session. It is noticeable that UAED has a more stable AUC-ROC performance along with a higher AUC-ROC performance in the late stage of training. And it has good convergence on all datasets. Whereas DOMI, although it has faster convergence on both MAHNOB-HCI and DRIVEDB datasets in the early stage of training, it suffers a larger performance fluctuation problem compared with our method. This is problematic and risky for anomaly detection tasks. Moreover, following the same experimental setup as in Section 4.6, 5 times nested LOSO experiments are performed on the WESAD dataset as shown in Table V. The stability of the UAED performance

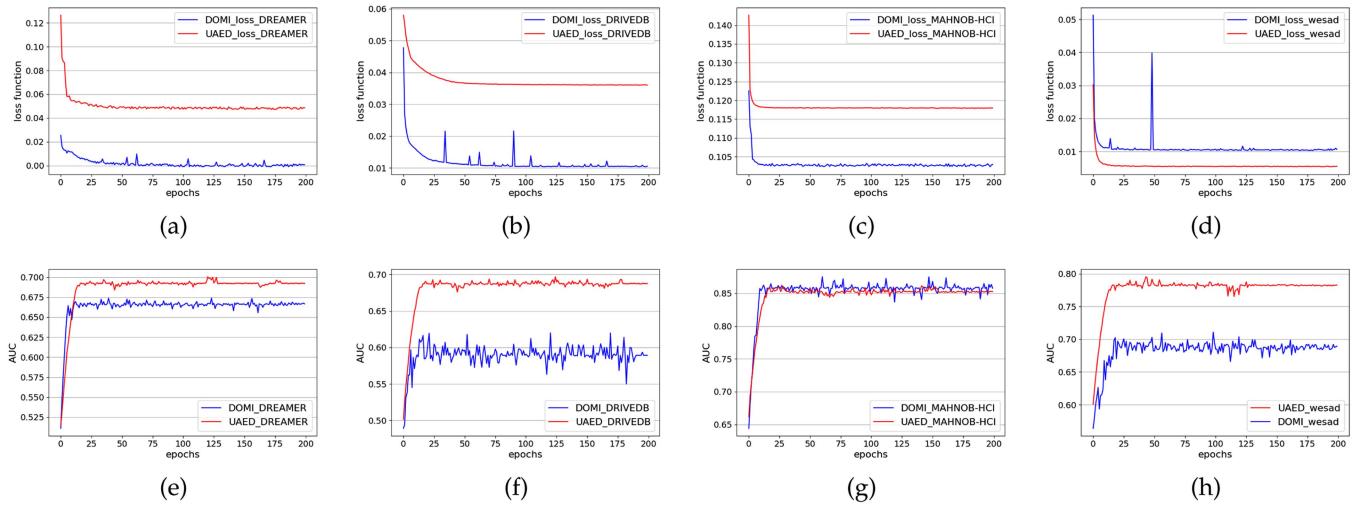


Fig. 9. Stability analysis. (a)–(d) Loss curves on DREAMER, DRIVEDB, MAHNOB-HCI, and WESAD datasets. (e)–(h) AUC-ROC changing curves during one training session on DREAMER, DRIVEDB, MAHNOB-HCI, and WESAD datasets.

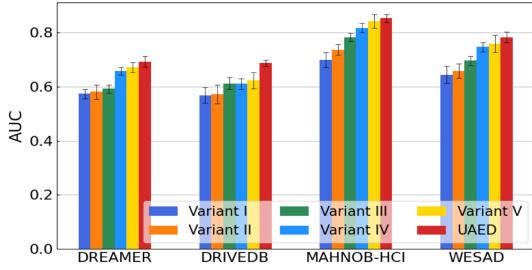


Fig. 10. Ablation experiments on four datasets DREAMER, DRIVEDB, MAHNOB-HCI, and WESAD.

over multiple independent training sessions can be demonstrated from the lower standard deviation (SD). To summarize, the above experimental results demonstrate that UAED has good stability.

#### K. Preliminary Application Experiment

To validate the feasibility of UAED application, we recorded ECG signals of 12 adult participants when abnormal emotions arose in an office setting, as well as participants' self-assessment of their emotional status after each stimulus, including subjective experiences of task load, emotion, and perceived stress. We elicited abnormal emotion responses by controlling participants' task completion times and applying distractions. Written informed consent was provided by all the participants prior to the enrollment of this study.

In the experiment, the ECG signal is captured using a low-cost flexible wearable wireless sensor developed by us, which uses an ECG detection module based on the AD8232 biochip and Ag/AgCl medical electrodes to collect ECG data, and transmits the values in ASCII form to the terminal to record ECG using Bluetooth. The LA electrode is placed on the back of the left hand, and RA and LA are placed on the back of the right hand. The sampling frequency of the ECG acquisition is 200 Hz. Pre-processing of the acquired signal is performed using the

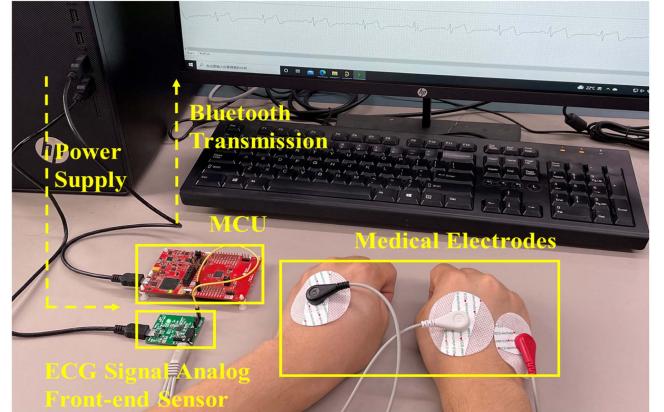


Fig. 11. Experimental scenario and ECG acquisition device. We visualize the acquired signal in real-time on a computer screen.

MCU's ADC module, designing a 1–100 Hz band-pass filter to retain the main signal, while setting a 50 Hz trap filter to remove the industrial frequency interference, which produces a filtered signal with clear peaks. The experimental scenario and the ECG acquisition device are illustrated in Fig. 11.

The experiment consists of three stages: First, there is a relaxation phase of approximately 10 minutes. Then, participants would receive work task prompts for a time-free and undisturbed work state, which is called *relaxed working state* of approximately one hour, at the end of which the subjects were allowed to take a short break. Finally, participants would work for approximately one hour under the mandatory work completion time and in a disturbing work state, which is called *stressed working state*. At the end of each work state, participants were invited to fulfill a questionnaire about their current subjective experience. A total of three recordings were taken during each stage of work, each lasting five minutes and at intervals of 10 minutes.

The model is trained using our self-collected data, and the data is segmented in the same way as in the previous section.

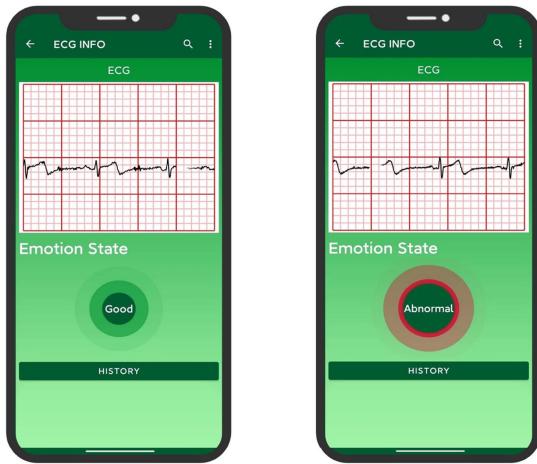


Fig. 12. Demo application on mobile phone.

The actual inference time of UAED is 108 ( $\pm 30$ ) ms. Ultimately, we obtained an AUC of 0.742 and an F1 of 0.709 for the model to detect abnormal emotions in the daily work environment, with a detection accuracy of 85%, which validated the effectiveness of the UAED system scheme and the feasibility of practical deployment. Compared with some advanced wearable device based emotion detection models in recent years, acceptable performance have been achieved, such as a method based on a wearable smart watch [51], which takes the user's real-time activity scene into consideration with an accuracy of 74.3%, an attention-based LSTM system in [52] that uses a combination of sensors from a smartphone and a wristband with the detection accuracy of 89.2%. Inherently, the model in [52] is a supervised model using multimodal information and thus achieving higher accuracy, while our work is unsupervised and requires only unimodal signals to obtain comparable results. However, since the experiments use medical electrodes as sensors to collect data as shown in Fig. 11, they are disturbed by hand activity, such as using keyboard. In addition, the ambiguity of the emotional boundaries can cause some inaccuracies in the annotation. These will have a negative impact on the experimental outcome, so there is still space for further enhancement in future work.

Moreover, we run UAED on a real mobile device. Our goal was to build a personal health system for detecting abnormal emotions. The trained model was saved as a .pth file, which we converted to a .pt file. The application runs on the Android 11 operating system. In Fig. 12, we provide several screenshots of the actual test in the application.

## V. CONCLUSION

In this article, we propose UAED, a personal health system framework for abnormal emotion detection. The stacking operation increases the dimension of physiological signals and inputs them into a CNN-based VAE model. Specifically, a Gaussian mixture prior is used rather than an isotropic Gaussian prior to better characterize the prototypical characteristics of the normal pattern. In addition, better detection results are achieved by using the whitening distance anomaly score to measure the extent of

abnormality of a sample during the detection process. Experimental results on public datasets show that UAED outperforms the suboptimal methods by 7.26%, 6.94%, and 8.27% based on AUC-ROC, AUC-PR, and F1 score, respectively. Meanwhile, we deployed the proposed abnormal emotion detection system on a mobile device in real-world scenarios and achieved 85% accuracy with an inference time of 108 ms, which verified the feasibility of the system. In the future, we will focus on the vulnerability to interference from skin state and human activity during hand detection, and apply UAED to wearable and mobile medical devices to establish a closer connection with end-users.

## REFERENCES

- [1] M. T. V. Yamuza et al., "Human emotion characterization by heart rate variability analysis guided by respiration," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 6, pp. 2446–2454, Nov. 2019.
- [2] A. Öhman, *Fear and Anxiety as Emotional Phenomena: Clinical Phenomenology, Evolutionary Perspectives, and Information-Processing Mechanisms*. New York, NY, USA: Guilford Press, 1993.
- [3] L. C. Gallo and K. A. Matthews, "Understanding the association between socioeconomic status and physical health: Do negative emotions play a role?" *Psychol. Bull.*, vol. 129, no. 1, 2003, Art. no. 10.
- [4] C. Xu et al., "Prognostic value of negative emotions on the incidence of breast cancer: A systematic review and meta-analysis of 129,621 patients with breast cancer," *Cancers*, vol. 14, no. 3, 2022, Art. no. 475.
- [5] M.-G. Kim and S. B. Pan, "Deep learning based on 1-D ensemble networks using ECG for real-time user recognition," *IEEE Trans. Ind. Informat.*, vol. 15, no. 10, pp. 5656–5663, Oct. 2019.
- [6] Siddharth, T.-P. Jung and T. J. Sejnowski, "Utilizing deep learning towards multi-modal bio-sensing and vision-based affective computing," *IEEE Trans. Affect. Comput.*, vol. 13, no. 1, pp. 96–107, Jan.–Mar. 2022.
- [7] P. Sarkar and A. Etemad, "Self-supervised ECG representation learning for emotion recognition," *IEEE Trans. Affect. Comput.*, vol. 13, no. 3, pp. 1541–1554, Jul.–Sep. 2022.
- [8] S. D. Kreibig, "Autonomic nervous system activity in emotion: A review," *Biol. Psychol.*, vol. 84, no. 3, pp. 394–421, 2010.
- [9] N. Merrill and C. Cheshire, "Trust your heart: Assessing cooperation and trust with biosignals in computer-mediated interactions," in *Proc. ACM Conf. Comput. Supported Cooperative Work Social Comput.*, 2017, pp. 2–12.
- [10] M. A. Feijt, J. H. Westerink, Y. A. De Kort, and W. A. IJsselsteijn, "Sharing biosignals: An analysis of the experiential and communication properties of interpersonal psychophysiology," *Hum.–Comput. Interaction*, vol. 38, pp. 49–78, 2023.
- [11] F. Agrafioti, D. Hatzinakos, and A. K. Anderson, "ECG pattern analysis for emotion detection," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 102–115, Jan.–Mar. 2012.
- [12] Y. Shi et al., "Personalized stress detection from physiological measurements," in *Proc. Int. Symp. Qual. Life Technol.*, 2010, pp. 28–29.
- [13] G. Tanev, D. B. Saadi, K. Hoppe, and H. B. D. Sorensen, "Classification of acute stress using linear and non-linear heart rate variability analysis derived from sternal ECG," in *Proc. IEEE 36th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2014, pp. 3386–3389.
- [14] B. Lamichhane, U. Großkathöfer, G. Schiavone, and P. Casale, "Towards stress detection in real-life scenarios using wearable sensors: Normalization factor to reduce variability in stress physiology," in *Proc. eHealth 360°: Int. Summit eHealth*, 2017, pp. 259–270.
- [15] A. Saeed, S. Trajanovski, M. Van Keulen, and J. Van Erp, "Deep physiological arousal detection in a driving simulator using wearable sensors," in *Proc. IEEE Int. Conf. Data Mining Workshops*, 2017, pp. 486–493.
- [16] S. Chakraborty, S. Aich, M.I. Joo, M. Sain, and H.C. Kim, "A multichannel convolutional neural network architecture for the detection of the state of mind using physiological signals from wearable devices," *J. Healthcare Eng.*, vol. 2019, 2019, Art. no. 5397814.
- [17] S. Chaabene, B. Bouaziz, A. Boudaya, A. Hökelmann, A. Ammar, and L. Chaari, "Convolutional neural network for drowsiness detection using EEG signals," *Sensors*, vol. 21, no. 5, 2021, Art. no. 1734.
- [18] A. A. S. et al., "Electrodermal activity based pre-surgery stress detection using a wrist wearable," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 1, pp. 92–100, Jan. 2020.

- [19] R. Musthyala, S. Arawkar, and M. Gupta, "Anomaly identification using multimodal physiological signals on the edge," in *Proc. 2nd Workshop Deep Learn. Wellbeing Appl. Leveraging Mobile Devices Edge Comput.*, 2021, pp. 7–12.
- [20] J. Zhu, F. Deng, J. Zhao, and J. Chen, "Adaptive aggregation-distillation autoencoder for unsupervised anomaly detection," *Pattern Recognit.*, vol. 131, 2022, Art. no. 108897.
- [21] Y. L. Lee, D.C. Juan, X.A. Tseng, Y. T. Chen, and S. C. Chang, "DC-prophet: Predicting catastrophic machine failures in datacenters," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discov. Databases*, 2017, pp. 64–76.
- [22] M. Sabokrou, M. Fathy, G. Zhao, and E. Adeli, "Deep end-to-end one-class classifier," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 675–684, Feb. 2021.
- [23] M. M. Breunig, H.P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density-based local outliers," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2000, pp. 93–104.
- [24] A. Zimek, E. Schubert, and H.P. Kriegel, "A survey on unsupervised outlier detection in high-dimensional numerical data," *Stat. Anal. Data Mining: The ASA Data Sci. J.*, vol. 5, no. 5, pp. 363–387, 2012.
- [25] Y. J. Cha and Z. Wang, "Unsupervised novelty detection-based structural damage localization using a density peaks-based fast clustering algorithm," *Struct. Health Monit.*, vol. 17, no. 2, pp. 313–324, 2018.
- [26] Z. Wang and Y. J. Cha, "Unsupervised deep learning approach using a deep auto-encoder with a one-class support vector machine to detect damage," *Struct. Health Monit.*, vol. 20, no. 1, pp. 406–425, 2021.
- [27] H. Chen, H. Liu, X. Chu, Q. Liu, and D. Xue, "Anomaly detection and critical SCADA parameters identification for wind turbines based on LSTM-AE neural network," *Renewable Energy*, vol. 172, pp. 829–840, 2021.
- [28] J. Zhu, F. Deng, J. Zhao, Z. Ye, and J. Chen, "Gaussian mixture variational autoencoder with whitening score for multimodal time series anomaly detection," in *Proc. IEEE 17th Int. Conf. Control Automat.*, 2022, pp. 480–485.
- [29] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," in *Proc. 25th Int. Conf. Inf. Process. Med.*, 2017, pp. 146–157.
- [30] S. Akcay, A. Atapour-Abarghouei, and T. P. Breckon, "Ganomaly: Semi-supervised anomaly detection via adversarial training," in *Proc. 14th Asian Conf. Comput. Vis.*, 2019, pp. 622–637.
- [31] X. Han, X. Chen, and L.-P. Liu, "GAN ensemble for anomaly detection," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 4090–4097.
- [32] C. Luo et al., "Correlating events with time series for incident diagnosis," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2014, pp. 1583–1592.
- [33] R. Pal, A. K. Tripathi, A. C. Pandey, M. A. Khan, V. G. Menon, and H. Mittal, "A N2CNN-based anomaly detection method for cardiovascular data in cyber-physical system," *IEEE Trans. Netw. Sci. Eng.*, vol. 14, no. 8, pp. 1–10, Aug. 2015.
- [34] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," Dec. 10, 2013, *arXiv:1312.6114v11*.
- [35] S. Katsigianis and N. Ramzan, "DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 1, pp. 98–107, Jan. 2018.
- [36] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 156–166, Jun. 2005.
- [37] M. Soleimani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 42–55, Jan.–Mar. 2012.
- [38] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing wesad, a multimodal dataset for wearable stress and affect detection," in *Proc. 20th ACM Int. Conf. Multimodal Interact.*, 2018, pp. 400–408.
- [39] R. A. Calvo and S. Mac Kim, "Emotions in text: Dimensional and categorical models," *Comput. Intell.*, vol. 29, no. 3, pp. 527–543, 2013.
- [40] A. Burns et al., "SHIMMER—A wireless sensor platform for noninvasive biomedical research," *IEEE Sensors J.*, vol. 10, no. 9, pp. 1527–1534, Sep. 2010.
- [41] H. Gao, B. Qiu, R. J. Duran Barroso, W. Hussain, Y. Xu, and X. Wang, "TSMAE: A novel anomaly detection approach for Internet of Things time series data using memory-augmented autoencoder," *IEEE Trans. Netw. Sci. Eng.*, vol. 10, no. 5, pp. 2978–2990, Sep.–Oct. 2023.
- [42] S. B. Shuvo, S. N. Ali, S. I. Swapnil, T. Hasan, and M. I. H. Bhuiyan, "A lightweight CNN model for detecting respiratory diseases from lung auscultation sounds using EMD-CWT-based hybrid scalogram," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 7, pp. 2595–2603, Jul. 2021.
- [43] J. Acharya and A. Basu, "Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning," *IEEE Trans. Biomed. Circuits Syst.*, vol. 14, no. 3, pp. 535–544, Jun. 2020.
- [44] F. T. Liu, K. M. Ting, and Z.H. Zhou, "Isolation-based anomaly detection," *ACM Trans. Knowl. Discov. Data*, vol. 6, no. 1, pp. 1–39, 2012.
- [45] B. Zong et al., "Deep autoencoding Gaussian mixture model for unsupervised anomaly detection," in *Proc. Int. Conf. Learn. Representations*, 2018.
- [46] L. Ruff et al., "Deep one-class classification," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 4393–4402.
- [47] S. Lin, R. Clark, R. Birke, S. Schönborn, N. Trigoni, and S. Roberts, "Anomaly detection for time series using VAE-LSTM hybrid model," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2020, pp. 4322–4326.
- [48] Y. Su et al., "Detecting outlier machine instances through Gaussian mixture variational autoencoder with one dimensional CNN," *IEEE Trans. Comput.*, vol. 71, no. 4, pp. 892–905, Apr. 2022.
- [49] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Jan. 30, 2017, *arXiv:1412.6980v*.
- [50] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, 2006.
- [51] Z. Wang, Z. Yu, B. Zhao, B. Guo, C. Chen, and Z. Yu, "Emotionsense: An adaptive emotion recognition system based on wearable smart devices," *ACM Trans. Comput. Healthcare*, vol. 1, no. 4, pp. 1–17, 2020.
- [52] K. Yang et al., "Behavioral and physiological signals-based deep multimodal approach for mobile emotion recognition," *IEEE Trans. Affect. Comput.*, vol. 14, no. 2, pp. 1082–1097, Apr.–Jun., 2023.



**Jiaqi Zhu** received the B.E. degree in automation from the Nanjing Institute of Technology, Nanjing, China, in 2019. She is currently working toward the Ph.D. degree in control science and engineering from the Beijing Institute of Technology, Beijing, China. Her research interests include image processing, time series analysis, and unsupervised learning.



**Fang Deng** (Senior Member, IEEE) received the B.E. and Ph.D. degrees in control science and engineering from the Beijing Institute of Technology, Beijing, China, in 2004 and 2009, respectively. He is currently a Professor with the School of Automation, Beijing Institute of Technology. His research interests include nonlinear estimation, fault diagnosis, control of renewable energy resources, and wireless sensor networks.



**Jiachen Zhao** received the B.E. degree in control science and engineering in 2016 from the Beijing Institute of Technology, Beijing, China, where he is currently working toward the Ph.D. degree in control science and engineering. His research interests include machine learning, time series analysis, and data mining.



**Daoming Liu** received the B.E. degree in automation from the Beijing University of Chemical Technology, Beijing, China, in 2018. He is an Engineer with the Beijing Institute of Technology Chongqing Innovation Center. His current work is about wearable system and wireless sensor networks.



**Jie Chen** (Fellow, IEEE) received the B.E., M.S., and Ph.D. degrees from the Beijing Institute of Technology, Beijing, China, in 1986, 1996, and 2001, respectively. He is a Member of the China Engineering Academy, a Professor with the State Key Laboratory of Intelligent Control and Decision of Complex Systems, Beijing Institute of Technology, and also a Professor with the Shanghai Research Institute for Intelligent Autonomous Systems, Tongji University, Shanghai, China. His research interests include complex systems, multiagent systems, multiobjective optimization and decision, constrained nonlinear control, and optimization methods.