Stress Detection Using Wearable Devices based on Transfer Learning

Jinting Wu*†, Yujia Zhang*, Xiaoguang Zhao*
*The State Key Laboratory of Management and Control for Complex Systems
Institute of Automation, Chinese Academy of Sciences
†School of Artificial Intelligence
University of Chinese Academy of Sciences
Beijing, China
Email: {wujinting2016, zhangyujia2014, xiaoguang.zhao}@ia.ac.cn

***Abstract*—Excessive stress will have a negative impact on people's physical and mental health, especially for some special occupations. Because stressful stimuli can trigger a variety of physiological responses, analyzing physiological signals collected by wearable devices has become an important way to evaluate the stress state in recent years. However, the number of available subjects of a target group may be small, and collecting a large amount of data when the target group changes is costly and time-consuming. To solve this problem, we propose a stress detection framework for a small target group which uses adversarial transfer learning method to learn shared knowledge about stress between different groups. In order to verify the performance of the framework, we establish a dataset consisting of 264 ordinary college students and 32 police school students, aiming to evaluate the acute stress state of police school students under video stimuli for psychological training in the future. Comprehensive experiments show that our algorithm has achieved a significant improvement in the target group compared with the baseline methods.**

***Index Terms*—Stress Detection, Transfer Learning, Physiological Signal Processing, Wearable Devices**

I. Introduction

Stress monitoring is one of the key aspects in the field of health care, and stress affects people's cognitive behaviors and decision-making. The stress states have a significant impact on people's working efficiency, and long-term excessive stress and negative emotion may even cause harm to a person's physical and mental health [1]. Especially, people with some special occupations (such as police, doctors, and drivers) are vulnerable to acute stressful stimuli in the working environment. Monitoring their responses to stressful stimuli is helpful to protect their health and guarantee their safety.

When people are exposed to stressful events, physiological responses are triggered under the control of the Autonomic Nervous System (ANS). Physiological signals which are commonly used for stress detection include Electrodermal Activity (EDA), Photoplethysmography (PPG), respiration, Electroencephalogram (EEG) and Skin Temperature (ST) [1]. Moreover, cortisol in salivary and blood is related to stress levels [2]. The stress state will also be reflected on behaviors such as eye movements, facial expressions [3], and speech [4]. The above measures have been widely applied to the research of stress detection in the laboratory environment. Physiological signals are more convenient to collect than cortisol, and have better objectivity than behaviors. However, some of the physiological signals such as EEG rely on non-portable equipment, making it difficult to be applied to real-world application scenarios. With the development of wearable sensors, the use of wrist-worn devices provides a portable, unobtrusive, and non-invasive way of physiological signal collection, and has achieved excellent performance. In this paper, we use wearable devices to collect EDA, PPG and ST signals for analysis.

This paper aims to assess the acute stress of a specific target group under multiple stressful stimuli, so as to lay the foundation for subsequent targeted psychological adjustment and training. One of the challenges is that the amount of available subjects of the target group may be small, and it is difficult to train an accurate stress detection model using only these data. In addition, although people have similar stress physiological responses, the intensity of their stress will be affected by various factors such as different ages, genders, occupations, and environments of data collection [1]. When the model trained on the general group is used for a specific group, the data distribution differences between these two groups may lead to large errors. In order to solve these problems, the transfer learning method is adopted in this paper to realize the data distribution alignment of the ordinary group and the target group, thereby achieving the purpose of data enhancement.

Transfer learning can transfer knowledge learned from previous tasks to new problems [5]. The domain that provides prior knowledge is called source domain, and the domain of the new problem to be solved is called target domain. It assists the learning of target domain knowledge by mining the similarities between the two domains in data, tasks, and models. The transfer learning algorithm has good performance for various transfer problems such as different research objects [6] and different research tasks [7]. Especially in the field of physiological signal analysis [8], [9], it can be used to compensate for the errors caused by the factors of individual differences, data distribution differences, and stimulus material differences. By using transfer learning methods, high accuracies can be achieved when the target domain has only a small amount of labeled data.

In this paper, an adversarial transfer learning method is used

978-1-6654-0126-5/21/$31.00 ©2021 IEEE

for training, which contains three main modules: a feature extractor, a domain discriminator and a stress detector. The domain discriminator and the stress detector take the output of the feature extractor as input, and then output domain labels and stress levels respectively. The domain discriminator is used to distinguish whether a sample belongs to the source domain or the target domain, and the purpose of the feature extractor is to generate features that the domain discriminator cannot distinguish. Through adversarial training, the data distribution of the two domains in the feature space can be more consistent. At the same time, the stress detector is also jointly trained with the above two modules to ensure that these features are contributed to the stress detection task, so that the learned model can be applied to stress assessment in both the source domain and the target domain.

The main contributions of this paper are as follows:

- To solve the problem of small amount of data for a specific target group, we propose an adversarial transfer learning framework for the task of physiological-signal-based stress detection, which uses general group data for data enhancement.
- In order to verify the performance of the framework, we designed an experimental protocol to induce stress and established a physiological signal dataset containing 264 ordinary college students and 32 police school students.
- Comprehensive experiments demonstrate that we have achieved robust feature extraction and stress classification on the target group, and the performance of our framework significantly outperforms the baseline method.

II. Related Work

A. Physiological-signal-based Stress Detection

Great progress has been made in stress detection by constructing the relationship between the time-domain and frequency-domain features of a single signal and the stress levels. Among them, EEG [10], [11], ECG [12], and EDA [13], [14] are most commonly used. Besides, more algorithms focus on multiple physiological signals and use multi-modal fusion methods for comprehensive analysis. For example, Ciabattoni et al. analyzed the data of skin response and body temperature captured by smart watches, and used K-Nearest Neighbor (KNN) method to detect real-time mental stress in cognitive tasks [15]. Sriramprakash et al. collected ECG and EDA signals and used Support Vector Machine (SVM) and KNN to detect users' workload [16]. Betti et al. established a wearable sensor system based on multi-modal information (EEG, ECG, and EDA), and used saliva cortisol as the criterion for assessing stress levels [2].

The above algorithms depend on manual features and traditional machine learning models. With the development of deep learning technology in recent years, deep networks have gradually been applied to the fields of physiological-signal-based stress detection and emotion recognition. For example, Jafari et al. converted multi-modal sequence signals into images and used a scalable and low-power embedded Deep Convolutional Neural Network (DCNN) to learn shared features [17]. The algorithm achieves an accuracy of 94% in the task of stress detection. Aristizabal et al. used deep networks to extract the features of physiological signals and behaviors collected by a wearable device [18]. Hassan et al. extracted deep features of EDA, PPG, and Zygomaticus Electromyography (zEMG) by the Deep Belief Network (DBN) and merged them with the statistical features [19]. Siddharth et al. obtained the spectrograms of all signal channels and used a deep network to extract their image features [20]. Although algorithms based on deep learning have achieved more superior performance in stress assessment, they have higher requirements on the scale of training data. The datasets which contain only dozens of subjects in the early research are difficult to meet the requirements of the training of the deep networks.

B. Transfer Learning

Transfer learning algorithms can be divided into 4 categories: instance transfer, feature representation transfer, parameter transfer, and relational knowledge transfer [5]. Most of the current studies focus on feature-based transfer learning, which reduces the gap between the source domain and the target domain through feature transformation [21], [22]. In recent years, deep networks have been used for transfer learning. Some adaptive methods based on the Maximum Mean Discrepancy (MMD) metric [23], [24] have been proposed. Transfer learning algorithms based on Generative Adversarial Network (GAN) [25]–[27] have also achieved great success. In the GAN-based algorithms, the generator aims to extract domain features and make the discriminator unable to distinguish the difference between the two domains, thereby achieving robust feature extraction.

In the field of physiological-signal-based stress recognition, there are only a few studies related to transfer learning, and most of them are used for cross-subject model training. Han et al. considered the trade-offs between task-related and user-related information to learn robust features that eliminate user difference, by introducing an additional adversarial network and a nuisance network [8]. Chen et al. proposed a cross-subject feature evaluation method to select a subset of features suitable for transfer in the driving state detection task [9]. Yin et al. proposed a transfer dynamical autoencoder for EEG feature extraction and mental stress recognition, which transferred the knowledge of EEG signal analysis knowledge in the emotion recognition task to the mental stress recognition task [7]. These works have learned robust physiological signal features and ensured a high recognition accuracy, even when the labeled data of the target domain are insufficient.

III. Methodology

The brief flow of our stress detection framework is shown in Fig. 1. The multi-modal signals are first pre-processed to remove abnormal fragments and noise. Next, manual features and deep features of the signals are extracted and fused. Then, during training, three modules are jointly trained in our proposed transfer learning method: feature extractors (it

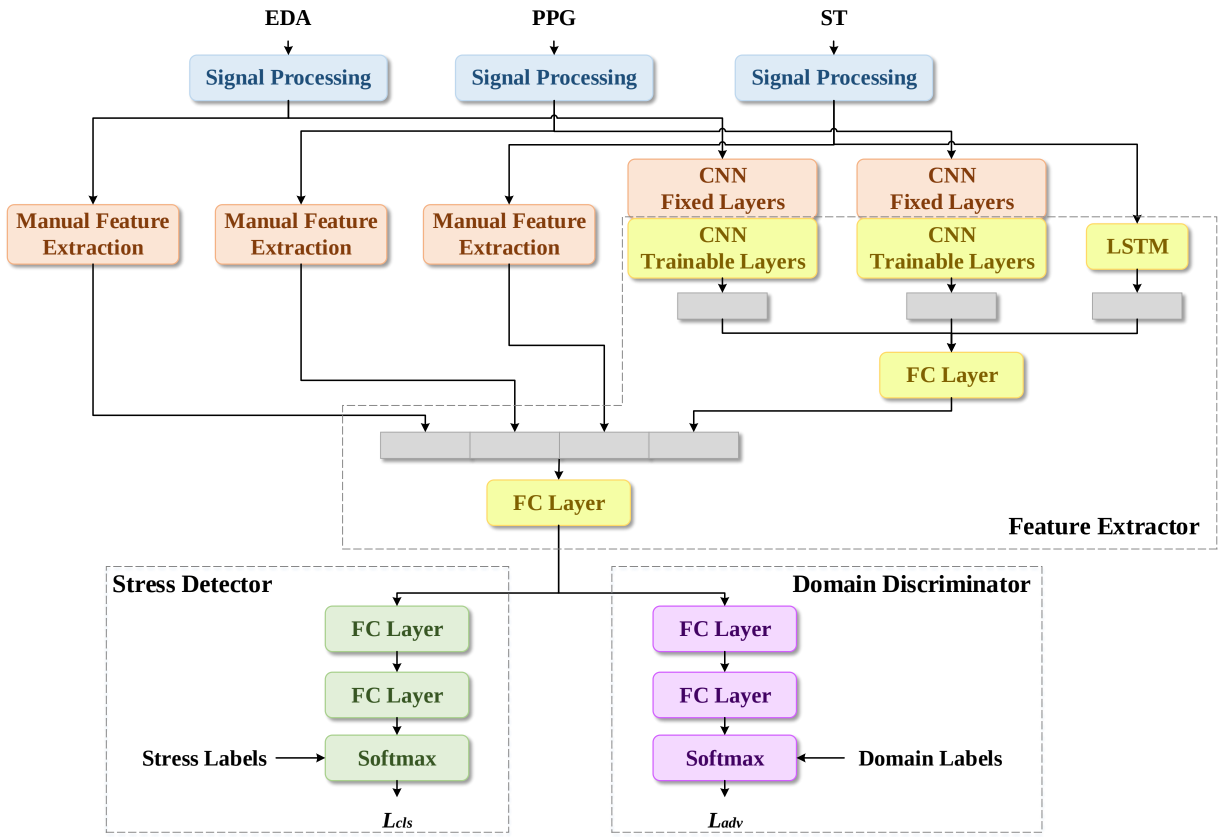

Fig. 1. The brief flow of our framework. The multi-modal signals are first pre-processed (blue) to remove abnormal fragments and noise. Then, manual features and deep features of multi-modal signals are extracted and fused. The manual feature extraction module and part of the deep network with fixed parameters (red) will not be updated during training. The proposed adversarial transfer learning algorithm contains the trainable feature extractor (yellow), a stress detector (green), and a domain discriminator (purple). After joint training, a model which is not sensitive to domain differences can be obtained.

is composed of the trainable part of the above deep feature extraction network and feature fusion network), domain discriminators, and stress detectors. Among them, the domain discriminator is used to distinguish whether a sample belongs to the source domain or the target domain. The pressure detector is used to give the classification result of the stress level. The feature extractor aims to generate the features required for stress detection and make the domain discriminator unable to distinguish between the two domains. After adversarial training of the three modules, a stress detection model that is insensitive to domain differences can be obtained.

A. Signal Processing

The goals of this module are denoising and removing abnormal data points of multi-modal signals. First, the hand movements are estimated based on the acceleration sensor signal, and data fragments of low quality caused by the intense movements are removed. Then, the PPG signal is filtered with a fourth-order Butterworth band-pass filter with a cut-off frequency of 0.5-3 Hz. The EDA signal is filtered by a fourth-order Butterworth low-pass filter with a cut-off frequency of 1 Hz. The ST signal is filtered by a moving average filter. Finally, the outlier detection based on wavelet transform is performed on the signals. The outliers are determined and eliminated by detecting the modulus maximum point of the coefficient of the wavelet transform.

B. Feature Extraction

Manual Features. The EDA signal is decomposed into tonic component and phasic component which are related to the Skin Conductance Level (SCL) and the Skin Conductance Response (SCR) respectively. Based on these components, the features such as mean, standard deviation, number of startles, and rise time are calculated [2]. For the PPG signal, the peak detection algorithm in HeartPy toolbox [28] is first performed to extract the RR Intervals for the calculation of Heart Rate Variability (HRV). The extracted PPG features include time-domain features such as Beats Per Minute (BPM), Inter-Beat Interval (IBI), and frequency-domain features such as Low Frequency (LF) power and High Frequency (HF) power. For the ST signal, we calculate its average and standard deviation as features. All manual features and their descriptions are shown in Table I.

TABLE I
MANUAL FEATURES AND DESCRIPTIONS

Number	Signal	Features	Description
1	EDA	TC_mean	Mean of the tonic component
2		TC_SD	Standard deviation of the tonic component
3		TC_mean1	Mean of the first-order differential of the tonic component
4		TC_SD1	Standard deviation of the first-order differential of the tonic component
5		PC_mean	Mean of the phasic component
6		PC_SD	Standard deviation of the phasic component
7		Startle	Number of detected startles
8		Startle_mean	Mean of the amplitude of the startles
9		Startle_SD	Standard deviation of the amplitude of the startles
10		Rise_time_mean	Mean of the rise time of the startles
11		Rise_time_SD	Standard deviation of the rise time of the startles
12		Fall_time_mean	Mean of the fall time of the startles
13		Fall_time_SD	Standard deviation of the fall time of the startles
14	PPG	BPM	Beats per minute
15		IBI	Inter-beat interval
16		SDNN	Standard deviation of all normal RR intervals
17		SDSD	Standard deviation of the squared differences between adjacent normal RR intervals
18		RMSSD	Square root of the mean of the squared differences between adjacent normal RR intervals
19		PNN20	Percentage of differences between adjacent normal RR intervals exceeding 20 ms
20		PNN50	Percentage of differences between adjacent normal RR intervals exceeding 50 ms
21		LF	Signal power in low frequency (0.04-0.15Hz)
22		HF	Signal power in high frequency (0.15-0.4Hz)
23		LF/HF	Ratio between LF and HF powers
24	ST	ST_mean	Mean of the skin temperature
25		ST_SD	Standard deviation of the skin temperature

Deep Features. A short-time Fourier transform with a window length of 1s is used to obtain the frequency components of the PPG and EDA signals in different time periods. The time-series data are converted to a frequency domain-based image representation [20]. Then, the spectrograms are input to the pre-trained Convolutional Neural Networks (CNN) for deep feature extraction. Among them, the parameters of the last several layers of this network are involved in network training, and the parameters of other layers are fixed. Because the frequency analysis of the ST signal contains less information, we only use a Long Short-term Memory (LSTM) network [29] for temporal feature extraction. These deep features are cascaded with the above manual features after dimension reduction, and the fused features are input into the stress detector and the domain discriminator.

C. Classification based on Transfer Learning

Let the data of the source domain can be denoted as $S_s = \{X_s, Y_s, G_s\}$, where X_s represents the intermediate results of feature extraction which is input to the trainable network, $Y_s \in \{1, \ldots, K\}$ represents the corresponding stress labels, K is the number of stress categories, and $G_s = \{0, 0, \cdots, 0\}$ represents the domain labels of the source domain. Similarly, the data of the target domain can be expressed as $S_t = \{X_t, Y_t, G_t\}$, where $G_t = \{1, 1, \cdots, 1\}$ represents the domain labels of the target domain. Our goal is to construct a discriminant model that can predict the stress level y of a given x and is not sensitive to domain differences.

Inspired by [25], we propose a transfer learning method based on adversarial networks, which includes a feature extractor F (the yellow part in Fig. 1), a stress detector C, and a domain discriminator D to determine which domain the sample belongs to. The goal of D is to distinguish the source domain from the target domain depending on the features generated by F, while the goal of F is to generate features that are not related to the domains. By competing with each other, F and D can make the transformed feature distribution of the source domain and target domain tend to be consistent. At the same time, the training of C ensures that F can extract the features that contribute to stress detection so that the model can be applied to the stress assessment of both domains.

Both D and C are composed of multiple Fully Connected (FC) layers and a softmax layer. For an input x, the output of F is denoted as $F(x)$, the output of D is denoted as $D(F(x))$, and the output of C is denoted as $C(F(x))$. Then the classification loss of C and D can be expressed as:

$$
\begin{aligned}
&L_{cls}(X_s, Y_s, X_t, Y_t) = \\
&\quad - \mathbb{E}_{(x,y)\sim(X_s,Y_s)\cup(X_t,Y_t)} \sum_{k=1}^{K} \mathbb{1}_{[k=y]} logC(F(x)),
\end{aligned} \tag{1}
$$

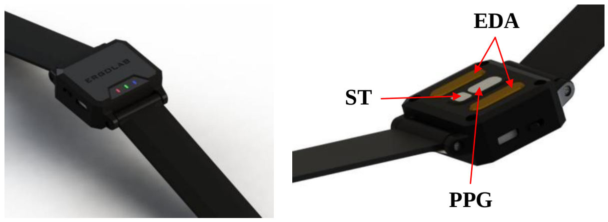

Fig. 2. The wearable device in our experiment.

$$L_{adv}(X_s, X_t) = -\mathbb{E}_{x\sim X_s}(\log D(F(x))) - \mathbb{E}_{x\sim X_t}(\log(1 - D(F(x)))). \quad (2)$$

Based on the above analysis of the goals of the three modules, C is optimized by minimizing L_{cls}, D is optimized by minimizing L_{adv}, and F is optimized by minimizing L_{cls} and maximizing L_{adv}. The optimization formulas of C, D and F can be expressed as:

$$\min_C L_{cls}(X_s, Y_s, X_t, Y_t \,|F), \quad (3)$$

$$\min_D L_{adv}(X_s, X_t \,|F), \quad (4)$$

$$\min_F (L_{cls}(X_s, Y_s, X_t, Y_t \,|C) - L_{adv}(X_s, X_t \,|D)), \quad (5)$$

During training, the parameters of each module are updated alternately. During testing, D is not used, and the output of the whole model is the stress levels given by C.

IV. Materials

A. Instruments

We use a wrist-worn wearable device for data collection, as shown in Fig. 2. This device can acquire signals including EDA, PPG, ST, and accelerometer data. It is connected to a smart phone via Bluetooth and can display waveforms of multiple signals in real-time in a supporting APP.

B. Participants

In this study, we recruited participants who belonged to two groups: police school students (group P) and ordinary college students (group O). Compared with group O, group P has better physical fitness and stronger psychological endurance, so there are differences in data distribution between the two groups.

Group P contains 32 healthy participants (24 males and 8 females) with an average age of 20.9 years. Group O contains 264 healthy participants (151 males and 113 females) with an average age of 22.8 years. People suffering from heart disease, high blood pressure, depression, etc. were excluded. All participants signed an informed consent form before being included in the study.

Fig. 3. The description of our experiment protocol.

C. Experimental Protocol

We use video stimuli to induce stress. In the preparation of video stimuli, we collected the blood cortisol concentrations of 10 additional subjects before and after watching the video. This experimental result proved that the selected videos in this study can effectively trigger stress responses.

The experimental protocol (which is shown in Fig. 3) includes 4 phases: Relax 1, Video 1, Relax 2, Video 2. Participants were required to wear a wrist-worn device, sit in front of the screen for more than 10 minutes, and follow the prompts to sit quietly or watch the video. All participants were required to sit comfortably and keep their body (especially wrists) still during the whole procedure. At the end of each stage, the subjects were required to verbally provide the self-reporting level (5 levels) of their current stress. The experimental protocol is designed as follows:

- Relax 1: Sit still for 2 minutes.
- Video 1: Watch a clip from the movie ”Final Destination 5”, which lasts 2 minutes and 25 seconds.
- Relax 2: Sit still for 2 minutes.
- Video 2: Watch a clip of a documentary related to the riot, which lasts 2 minutes and 56 seconds. This stimulus is closely related to the future working environment of police school students.

D. Data Annotations

During data collection, we obtained the self-reporting stress levels of each phase, which range from 1 to 5 (from low to high). They are used as training labels, and the stress detection task is transformed into a 5-class classification problem. In addition, we re-label the first and second levels as low stress, and levels from the third to the fifth as high stress. A 2-class stress detector is also trained. The experimental results with the self-reporting labels are shown in Section V-B1.

Finally, we also use different phases in our experimental protocol as training labels. The data are divided into two categories: relaxed (Relax 1 and Relax 2) and stressful (Video 1 and Video 2). The experimental results with the phase labels are shown in Section V-B2.

V. Experiments

A. Implementation Details

For the PPG and EDA signals, we choose the pre-trained VGG16 network [30], a classic network, for deep feature extraction due to its effectiveness. The number of trainable layers is set to 2. For the ST signals, a two-layer BLSTM

TABLE II
EXPERIMENTAL RESULTS WITH SELF-REPORTING LABELS

Method	5-class Accuracy	2-class Accuracy
SVM (Group O)	37.50%	66.41%
SVM (Group P)	33.33%	62.29%
JDA+SVM	40.91%	72.65%
Our Framework	43.18%	75.00%

TABLE III
EXPERIMENTAL RESULTS WITH PHASE LABELS

Method	Accuracy
SVM (Group O)	76.56%
SVM (Group P)	80.05%
JDA+SVM	84.37%
Our Framework	86.76%

network is used to extract features, and the numbers of LSTM neurons are set to 64. Both FC layers which fuse the deep features and the manual features have 50 units. In the stress detector, the numbers of units of the two FC layers are set to 50 and 20. The numbers of units in the discriminator are set to 30 and 10. The activation function in the network is ReLU. During training, the batch size is set to 8. The Adam optimizer is utilized to minimize the loss, and the learning rate is set to 0.0001. The hyper-parameters are selected by 5-fold cross-validation.

B. Experimental Results and Analysis

1) Comparisons with Baseline Methods Using Self-reporting Labels: In this section, we compare our proposed framework with the following baseline and state-of-the-art methods: a). an SVM trained with the manual features of group O; b). an SVM trained with the manual features of group P; c). a state-of-the-art transfer learning algorithm called Joint Distribution Adaptation (JDA) [22] and an SVM trained with the manual features of the two group. All models use 5-fold cross-validation to divide the group O, and the test samples are all from the group P. Model b), c) and our algorithm use 5-fold cross-validation to divide group P. Besides, in model c) and our algorithm, the training samples of group P combine all samples of group O to train 5 models. The experimental result of each method is the average accuracy of 5 models, which is shown in Table II.

The recognition accuracy of the SVMs which train only on group O or group P is the lowest. Because group P has few training samples, it is prone to problems such as outlier interference and category unbalance during training, which leads to a decrease in the robustness of the model. The sample amount of group O is sufficient to support the training of the model a), but the physique and stress tolerance of the two groups are different. Using this model to predict the samples of group P will bring obvious bias. After the data adaptation of the JDA algorithm, the feature distribution of the two groups tends to be consistent, and the recognition accuracy has increased significantly.

Because the difference between the multiple stress levels is difficult to distinguish, the accuracy of the 5-class setting is not satisfactory. Especially there are only a few samples at level 4 or 5 (high stress), the misclassification rates of these methods at level 4 or 5 are high. However, it can be observed that the framework proposed in this paper achieves the highest recognition accuracy in both the 5-class and 2-class settings. The fusion features with abundant information and the adversarial transfer method contribute to the improvement of the detection performance.

2) Comparisons with Baseline Methods Using Phase Labels: We compare the above algorithms in the task of distinguishing whether a sample is related to a relaxed state or a stress state which is triggered by video stimuli. The experimental results are shown in Table III. We can observe that the transfer learning algorithms still show obvious advantages, and our framework achieves the highest accuracy rate.

It is worth noting that these results are much higher than the 2-class accuracy in Table II. Although the self-report rating is a commonly used way to establish stress ground truth, the stress state is a subconscious process in some cases, and the self-report rating is highly subjective [1]. In addition, stress is a multifaceted experience, and the intensity and reaction time of the self-experience and physiological response may be varied. Although sometimes people's inner stress feeling is not obvious, their physiological reactions will still be stimulated. Compared with ordinary students, this phenomenon is more obvious among police school students. Their psychological endurance may be stronger, but it also brings the problem that it is difficult to detect and adjust abnormal states in time under pressure stimulation. In future work, we will conduct a comprehensive analysis with reference to various pressure marking methods.

VI. CONCLUSION

In this paper, we propose a stress detection framework based on transfer learning to solve the problem of insufficient samples of the target group. It is obtained through adversarial training of three modules: a feature extractor, a domain discriminator, and a stress detector. To evaluate the effectiveness of our method, we design an experimental protocol for data collection and established a physiological signal dataset containing ordinary college students and police school students. In the experiments, we compare the proposed framework with the baseline and the state-of-the-art algorithms, and analyze the experimental results based on self-report labels and phase labels. The experimental results demonstrate the effectiveness and superiority of our framework. In future work, we will explore the impact of different feature components, framework modules, and network parameters on model performance in detail. In addition, we also aim to develop a real-time stress detection system with stronger anti-interference capabilities and apply it in real-world environments.

Acknowledgment

This work was supported by the National Key Research and Development Project of China (Grants No. 2019YFB1310601), the National Key R&D Program of China (Grants No. 2017YFC0820203), and the National Natural Science Foundation of China (Grant No. 62103410).

References

[1] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis, "Review on psychological stress detection using biosignals," *IEEE Transactions on Affective Computing*, 2019.

[2] S. Betti, R. M. Lova, E. Rovini, G. Acerbi, L. Santarelli, M. Cabiati, S. Del Ry, and F. Cavallo, "Evaluation of an integrated system of wearable physiological sensors for stress monitoring in working environments by using biological markers," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 8, pp. 1748–1758, 2017.

[3] G. Giannakakis, M. Pediaditis, D. Manousos, E. Kazantzaki, F. Chiarugi, P. G. Simos, K. Marias, and M. Tsiknakis, "Stress and anxiety detection using facial cues from videos," *Biomedical Signal Processing and Control*, vol. 31, pp. 89–101, 2017.

[4] I. Lefter, G. J. Burghouts, and L. J. Rothkrantz, "Recognizing stress using semantics and modulation of speech and gestures," *IEEE Transactions on Affective Computing*, vol. 7, no. 2, pp. 162–175, 2015.

[5] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.

[6] D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.

[7] Z. Yin, M. Zhao, W. Zhang, Y. Wang, Y. Wang, and J. Zhang, "Physiological-signal-based mental workload estimation via transfer dynamical autoencoders in a deep learning framework," *Neurocomputing*, vol. 347, pp. 212–229, 2019.

[8] M. Han, O. Özdenizci, Y. Wang, T. Koike-Akino, and D. Erdoğmuş, "Disentangled adversarial autoencoder for subject-invariant physiological feature extraction," *IEEE signal processing letters*, vol. 27, pp. 1565–1569, 2020.

[9] L.-l. Chen, A. Zhang, and X.-g. Lou, "Cross-subject driver status detection from physiological signals based on hybrid feature selection and transfer learning," *Expert Systems with Applications*, vol. 137, pp. 266–280, 2019.

[10] A. R. Subhani, W. Mumtaz, M. N. B. M. Saad, N. Kamel, and A. S. Malik, "Machine learning framework for the detection of mental stress at multiple levels," *IEEE Access*, vol. 5, pp. 13 545–13 556, 2017.

[11] F. Al-Shargie, T. B. Tang, N. Badruddin, and M. Kiguchi, "Towards multilevel mental stress assessment using svm with ecoc: an eeg approach," *Medical & biological engineering & computing*, vol. 56, no. 1, pp. 125–136, 2018.

[12] M. Salai, I. Vassányi, and I. Kósa, "Stress detection using low cost heart rate sensors," *Journal of healthcare engineering*, vol. 2016, 2016.

[13] R. Zangróniz, A. Martínez-Rodrigo, J. M. Pastor, M. T. López, and A. Fernández-Caballero, "Electrodermal activity sensor for classification of calm/distress condition," *Sensors*, vol. 17, no. 10, p. 2324, 2017.

[14] O. Dehzangi, V. Rajendra, and M. Taherisadr, "Wearable driver distraction identification on-the-road via continuous decomposition of galvanic skin responses," *Sensors*, vol. 18, no. 2, p. 503, 2018.

[15] L. Ciabattoni, F. Ferracuti, S. Longhi, L. Pepa, L. Romeo, and F. Verdini, "Real-time mental stress detection based on smartwatch," in *2017 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 2017, pp. 110–111.

[16] S. Sriramprakash, V. D. Prasanna, and O. R. Murthy, "Stress detection in working people," *Procedia computer science*, vol. 115, pp. 359–366, 2017.

[17] A. Jafari, A. Ganesan, C. S. K. Thalisetty, V. Sivasubramanian, T. Oates, and T. Mohsenin, "Sensornet: A scalable and low-power deep convolutional neural network for multimodal data classification," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 66, no. 1, pp. 274–287, 2018.

[18] S. Aristizabal, K. Byun, N. Wood, A. F. Mullan, P. M. Porter, C. Campanella, A. Jamrozik, I. Z. Nenadic, and B. A. Bauer, "The feasibility of wearable and self-report stress detection measures in a semi-controlled lab environment," *IEEE Access*, vol. 9, pp. 102 053–102 068, 2021.

[19] M. M. Hassan, M. G. R. Alam, M. Z. Uddin, S. Huda, A. Almogren, and G. Fortino, "Human emotion recognition using deep belief network architecture," *Information Fusion*, vol. 51, pp. 10–18, 2019.

[20] S. Siddharth, T.-P. Jung, and T. J. Sejnowski, "Utilizing deep learning towards multi-modal bio-sensing and vision-based affective computing," *IEEE Transactions on Affective Computing*, 2019.

[21] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE transactions on neural networks*, vol. 22, no. 2, pp. 199–210, 2010.

[22] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 2200–2207.

[23] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," in *International conference on machine learning*. PMLR, 2015, pp. 97–105.

[24] Y. Zhu, F. Zhuang, and D. Wang, "Aligning domain-specific distribution and classifier for cross-domain classification from multiple sources," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 5989–5996.

[25] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *The journal of machine learning research*, vol. 17, no. 1, pp. 2096–2030, 2016.

[26] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7167–7176.

[27] K. Saito, D. Kim, S. Sclaroff, T. Darrell, and K. Saenko, "Semi-supervised domain adaptation via minimax entropy," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 8050–8058.

[28] P. van Gent, H. Farah, N. van Nes, and B. van Arem, "Heartpy: A novel heart rate algorithm for the analysis of noisy signals," *Transportation research part F: traffic psychology and behaviour*, vol. 66, pp. 368–378, 2019.

[29] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.