PiRL: Participant-Invariant Representation Learning for Healthcare Using Maximum Mean Discrepancy and Triplet Loss

Zhaoyang Cao  
Department of Computational and Applied Mathematics  
Rice University  
Houston, US.  
zc48@rice.edu

Han Yu  
Department of Electrical and Computer Engineering  
Rice University  
Houston, US.  
hy29@rice.edu

Huiyuan Yang  
Department of Electrical and Computer Engineering  
Rice University  
Houston, US.  
hy49@rice.edu

Akane Sano  
Department of Electrical and Computer Engineering  
Rice University  
Houston, US.  
akane.sano@rice.edu

Abstract—Due to individual heterogeneity, person-specific models are usually achieving better performance than generic (one-size-fits-all) models in data-driven health applications. However, generic models are usually preferable in real-world applications, due to the difficulties of developing person-specific models, such as new-user-adaptation issues and system complexities. To improve the performance of generic models, we propose a Participant-invariant Representation Learning (PiRL) framework, which utilizes maximum mean discrepancy (MMD) loss and domain-adversarial training to encourage the model to learn participant-invariant representations. Further, to avoid trivial solutions in the learned representations, a triplet loss based constraint is used for the model to learn the label-distinguishable embeddings. The proposed framework is evaluated on two public datasets (CLAS and Apnea-ECG), and significant performance improvements are achieved compared to the baseline models.

Index Terms—representation learning, maximum mean discrepancy loss, triplet loss, latent space similarity

I. INTRODUCTION

Deep learning has gained popularity in modeling time-series data for solving health-related problems. For example, Oh et al. [1] proposed an automated system that combines a convolutional neural network (CNN) and a long short-term memory network (LSTM) for the diagnosis of arrhythmia. Erdenebayar et al. [2] designed a deep neural network, recurrent neural networks, and a gated-recurrent unit to distinguish apnea and hypopnea events using an electrocardiogram (ECG) signal. In addition to physical health, deep models have also been used for mental health. Yu and Sano [3] applied semi-supervised learning on leveraging unlabeled data to estimate wearable-based momentary stress. Radhika et al. [4], [5] proposed the frameworks that investigate the effectiveness of transfer learning and deep multimodal fusion on CNN stress detection models.

Although deep learning models have achieved promising results in health applications, researchers have observed that person-specific models usually outperform generic models [6]–[11], due to heterogeneous time series data. For example, Bsoul et al. [12] showed that the accuracy of the subject-dependent sleep apnea classification model is 6% higher than that of the subject-independent model. Nath et al. [13] showed a performance gap of 22.5% in accuracy between the subject-dependent and the generic LSTM models in emotion recognition. Although person-specific models have been widely proven to outperform the generic models in health applications [14]–[16], we cannot therefore ignore the challenges of developing person-specific models. For example, person-specific models cannot be easily extended to unseen participants [9], and it is usually impractical to collect enormous datasets from individuals to build person-specific models.

Researchers have explored improving the performance of generic models by introducing person-specific information. For example, Radhika et al. [4], [5] used person-specific information in the testing set during the feature extraction and selection. Wu et al. [17] achieved model personalization from a pre-trained generic model by active learning approach and improved detection precision on each new patient. Similarly, Li et al. [18] fine-tuned a generic convolutional neural network to the tuned dedicated CNN for extracting the characteristic information of a specific patient and obtained the effectiveness on arrhythmia prediction. Bethge et al. [19] utilized MMD loss to impose domain-invariant representations for emotion classification tasks, where each participant had his/her own private encoder with a classifier shared among all. Although improved performances were observed in the generic models,
problems still exist in those studies, as individual encoders result in both high computational costs and difficulties in adapting to new subjects.

In this work, we aim to improve the performances of generic models without introducing extra computational complexity into the model, thus avoiding the aforementioned drawbacks of person-specific methods. We propose a representation learning framework to learn participant-invariant features through two learning phases, phase I: unsupervised learning and phase II: supervised learning. During the unsupervised learning phase, maximum mean discrepancy (MMD) loss is integrated with representation learning, which alleviates the heterogeneous issues and the influence of person-specific information, to minimize the distribution shifts among features from different subjects. Instead of MMD loss, we also test domain classification loss from domain-adversarial training of neural networks (DANN) architecture, which aims to blur the participant-distinguishable information among the learned representations, to make label predictor more robust to the target participant [19]. During the supervised learning phase, triplet loss, which aims to learn the label-distinguishable embedding, is also used in the framework as a constraint to avoid trivial solutions in learned representations. We evaluate the proposed framework using two public datasets, including downstream tasks: sleep apnea detection and stress detection. Our results suggest that the proposed method can help improve the model performance significantly compared to the baseline representation learning model.

II. METHODOLOGY

We propose a representation learning framework that aims to extract participant-invariant representations, named PiRL. The main PiRL framework is visualized in Figure 1 and detailed architecture is shown in Experimental Setting III-B. We employ a 1D CNN-based auto-encoder structure as a deep feature extractor from input wearable data. On top of the auto-encoder, MMD loss and domain classification loss are utilized to constrain the representations from distribution shifts, therefore encouraging the learning embedding to be participant-invariant. During training for downstream tasks, we optimize the model with a triplet loss for label-distinguishable representations. The following subsections will introduce the aforementioned components in detail.

A. Representation learning

We utilize a 1D CNN-based auto-encoder to extract learning representations from raw time-series sequences $X$. The encoder extracts latent representations from input sequences as $e$ with a series of 1D CNN layers; whereas the decoder aims to output the reconstructed signal as $\hat{X}$ from $e$ using up-sampling layers. The objective of the auto-encoder is:

$$\mathcal{L}_{ae} = \|X - \hat{X}\|_2^2$$  \hspace{1cm} (1)

1) MMD loss: To encourage the model to learn the participant-invariant representations, we constrain the model with an MMD loss function, which is widely used in eliminating distribution shifts among different groups of data [20].

Given training samples from two different subjects as $X_i$ and $X_j$ with the total number of subjects $N$, the MMD loss can be considered as follows:

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \mathcal{L}_{mmd}(p, q, \mathcal{H}) = \sup_{f \in \mathcal{H}, ||f||_{\infty} \leq 1} \left( \mathbb{E}_{p(X_i)}[f(X_i)] - \mathbb{E}_{q(X_j)}[f(X_j)] \right)$$

where $p$ and $q$ are the distributions of variable $X_i$ and $X_j$, $\mathcal{H}$ is the Hilbert space, and $f$ is the mapping function. During the training process, the model is optimized to minimize the distribution distances between the pairs of subjects. Thus, the overall objective of unsupervised learning is:

$$\mathcal{L} = \mathcal{L}_{ae} + \lambda \cdot \mathcal{L}_{mmd}$$  \hspace{1cm} (2)

where $\lambda$ is the coefficient of MMD loss and set to 0.2.

2) Domain classification loss: A domain classifier in DANN [19] can help learn a classifier that is more robust in target data by accomplishing the following adversarial tasks: minimizing the loss of label prediction while maximizing the loss of domain classification. Subsequently, with the addition of a gradient reversal layer before the domain classifier, the overall objective function is the sum of two minimizing problems [21]. In our case, we treat the individual participant as an independent domain, and the number of domains is equal to the number of participants. The corresponding domain classification loss is given below:

$$\mathcal{L}_{domain} = - \sum_{i=1}^{N} (p_i \cdot \log(q_i))$$  \hspace{1cm} (3)

where $N$ is the number of labels, $p_i$ is the true probability distribution of one-hot encoding of the source domain, and $q_i$ is the predicted probability distribution of one-hot encoding computation of the target domain. As a result, the overall objective of unsupervised learning is:

$$\mathcal{L} = \mathcal{L}_{ae} + \mu \cdot \mathcal{L}_{domain}$$  \hspace{1cm} (4)

where $\mu$ is the coefficient of domain classification loss and set to -1. At last, the final objective function of this study is

$$\mathcal{L} = \mathcal{L}_{ae} + \lambda \cdot \mathcal{L}_{mmd} + \mu \cdot \mathcal{L}_{domain}$$  \hspace{1cm} (5)

B. Fine-tuning with triplet loss

The supervised learning models for downstream tasks are built on top of the learned representations and in this work, a fully connected layer is used as a classifier. Then the supervised learning structure is fine-tuned according to different downstream tasks. Furthermore, we also apply the triplet loss [22] to optimize the representation of the training labels and avoid the trivial solutions learned in the pre-training procedure. The triplet loss is given below:

$$\mathcal{L}_{triplet} = \max(d(a, p) - d(a, n) + \text{margin}, 0)$$  \hspace{1cm} (6)

where $a$ represents anchor sample data, $p$ represents ‘positive’ sample data with the same class label from the anchor, $n$ represents ‘negative’ sample data with a different class label from the anchor.
Fig. 1. PiRL Network Architecture (a) step I: unsupervised learning which consists of a 1D CNN-based auto-encoder with reconstruction loss and penalized MMD loss and domain classification loss. (b) step II: supervised learning which utilized triplet loss for label-distinguishable representation.

represents ‘negative’ sample data with different class labels against the anchor, and the margin is a positive scalar. From the Eq-6, we can see that the objective of the triplet loss is to pull samples with the same class labels closer together in embedding space while pushing dissimilar ones apart.

C. Latent Space Similarity

Similarity of participants’ embeddings in the latent space is often helpful in demonstrating the relationships among participants and how the representations perform in a reduced-dimensional space. Since cosine and manhattan distance are two commonly used approaches to depict and quantify the latent space similarities, we used these two types of distance measures to evaluate the effects of MMD loss. We evaluated the pairwise distance of participants’ representations in the test set to measure the latent space similarities. The formula of the cosine and manhattan distance is given as follows:

$$D_{\text{Cosine}}(x, y) = 1 - \frac{<x, y>}{||x||_2 \cdot ||y||_2}$$

$$D_{\text{Manhattan}}(x, y) = \sum_{i=1}^{n} |x_i - y_i|$$

where $x$ and $y$ are both given vectors with the same dimension. In order to obtain the pairwise distance, we first used t-distributed stochastic neighbor embedding (t-SNE) to transform the representations into a 2-dimensional vector in this study as it is commonly used to realize the dimension reduction for high-dimensional data [23]. As cosine and manhattan distance requires input vectors to be in the same dimensionality, we flattened the latent space of each participant from the baseline and MMD models to the $1 \times 32$ vector space which is used for the computation.

III. EXPERIMENTAL SETTING

Two public datasets, including CLAS [24] (stress detection) and Apnea-ECG [25] (sleep apnea detection), are used to evaluate the proposed framework.

A. Dataset

1) CLAS Dataset: The CLAS [24] dataset was collected for the automatic assessment of certain states of mind and emotional conditions using physiological data. The dataset consists of recordings of electrocardiogram (ECG), photoplethysmogram (PPG), electrodermal activity (EDA), and acceleration (ACC) signals. The dataset was collected from 62 healthy subjects who participated in three interactive tasks and two perceptive tasks. Labels for arousal, valence, and stress were assigned based on the stimuli tags for the interactive tasks. In this paper, we conducted experiments for stress detection (class 0: non-stressed, class 1: stressed) using EDA signals. The raw EDA data were first prepossessed with a low pass Butterworth filter (a cutoff frequency of 0.2 Hz), and then split into 10-second segments. The train/test set was split in a subject-independent manner. Table I summarizes the EDA training and testing sets used for the experiments.

|                  | #Participants | #Non-stressed | #Stressed |
|------------------|---------------|---------------|-----------|
| Training Set     | 45            | 746           | 247       |
| Testing Set      | 13            | 269           | 90        |
| Total            | 58            | 1015          | 337       |

2) Apnea-ECG Dataset: The recordings of Apnea-ECG [25] dataset include continuous ECG signals and sets of annotations for apnea (respiratory signals) from 70 participants. The recordings of 35 participants were used as a training set, and the rest was used as a testing set. The length of the recordings ranged from slightly less than 7 hours to about 10 hours each, and the labels were provided as indicators of the presence (class 1) or absence (class 0) of sleep apnea in each minute of each recording. Thus, we split the ECG recordings into individual one-minute segments. Table II summarizes
the details about the training and the test sets used in our experiment.

| #Participants | #Normal breath | #Disordered breath |
|---------------|----------------|--------------------|
| Training Set  | 35             | 10496              | 6514               |
| Testing Set   | 35             | 10685              | 6548               |
| Total         | 70             | 21181              | 13062              |

For the CLAS and Apnea-ECG datasets, we used instance-wise min-max normalization for each time series segment on the training set and testing set separately and independently to ensure that each normalized segment had a similar scale within (0,1).

B. Model Description

1) Step 1: Unsupervised Learning: For the CLAS dataset, the encoder consists of 9 CNN layers and a fully connected layer (named ‘Embedding’), whereas the decoder consists of corresponding 8 up-sampling layers. For the Apnea-ECG dataset, the encoder consists of 12 CNN layers and an Embedding layer, whereas the decoder consists of corresponding 11 up-sampling layers. The detailed structures of the framework including the shape of input/output, kernel size, normalization method, and activation function are shown in Table III and IV for the CLAS and Apnea-ECG dataset respectively.

2) Step 2: Supervised Learning: The same encoder structure was used in supervised learning (step2), but with two extra fully connected layers (fc-1 and fc-2 shown in Table III and IV) as a classifier.

3) Person-specific Models: The performance of the person-specific models was calculated to compare against the baseline and the proposed PiRL frameworks. We used the original training set of CLAS dataset to obtain the performance of person-specific models. Specifically, as shown in Table IV for each participant, the original training set was divided into a new training set and testing set with a ratio of 70:30%. The person-specific models were trained on the training sets (Tr1) and tested on the testing sets (Te1) to obtain the results. The final prediction accuracy of the person-specific models was reported as the average and standard deviation of accuracy of all participants.

C. Training Process

In step 1, the unsupervised learning phase, we used an Adam optimizer with an initial learning rate of 0.001, and the learning rate was decayed by 0.9 after every 5 epochs. The batch size was 32 for the CLAS dataset and 256 for the Apnea-ECG dataset. The length of sequence for the CLAS dataset was 960 and 6000 for the Apnea-ECG dataset. We trained the model for 100 epochs. The optimal weight for MMD loss was selected to be 0.2 from 0.1 to 0.5 with an increase of 0.1 in cross validation.

In step 2, the supervised learning phase, the parameters such as batch size and learning rate remained the same, and four supervised models: 1) baseline model, 2) MMD loss only, 3) triplet loss only, 4) MMD and triplet loss, were evaluated. The optimal weight for the triplet loss was selected to be 0.2 from 0.1 to 0.5 with an increase of 0.1 in cross validation. To evaluate the performance, accuracies and standard deviations of 10 runs were reported.

Our model was implemented in the Pytorch deep learning framework, and was trained and tested on the NVIDIA GeGorce 3090Ti GPU.

IV. RESULTS AND DISCUSSION

We tested the proposed PiRL frameworks on two datasets, including CLAS and Apnea-ECG datasets for applications in mental health and physical health. Experimental Setting III already included detailed information on two datasets and experimental settings such as hyper-parameters. For each supervised prediction model, we pre-trained and fine-tuned the encoder at the beginning of each epoch. We compared the prediction accuracies of the PiRL models against the ones of the baseline (only an auto-encoder and a supervised learning model without any additional constraints) and person-specific models.
**TABLE IV**

**Structure of the Framework for Apnea-ECG Dataset.** B: batch size; L: length of sequence.

| Layer Name | Input Shape | Output Shape | Kernel |
|------------|-------------|--------------|--------|
| Conv1d-1   | [B, 1, L]   | [B, 16, L/2] | [1x4]  |
|            |             | BatchNorm, ReLU |        |
| Conv1d-2   | [B, 16, L/2]| [B, 32, L/4] | [1x4]  |
|            |             | BatchNorm, ReLU |        |
| Conv1d-3   | [B, 32, L/4]| [B, 64, L/8] | [1x4]  |
|            |             | BatchNorm, ReLU |        |
| Conv1d-4   | [B, 64, L/8]| [B, 64, L/16]| [1x4]  |
|            |             | BatchNorm, ReLU |        |
| Conv1d-5   | [B, 64, L/16]| [B, 128, L/32]| [1x4]  |
|            |             | BatchNorm, ReLU |        |
| Conv1d-6   | [B, 128, L/32]| [B, 128, L/64]| [1x4]  |
|            |             | BatchNorm, ReLU |        |
| Conv1d-7   | [B, 128, L/64]| [B, 256, L/128]| [1x4]  |
|            |             | BatchNorm, ReLU |        |
| Conv1d-8   | [B, 256, L/128]| [B, 256, L/256]| [1x4]  |
|            |             | BatchNorm, ReLU |        |
| Conv1d-9   | [B, 256, L/256]| [B, 512, L/512]| [1x4]  |
|            |             | BatchNorm, ReLU |        |
| Conv1d-10  | [B, 512, L/512]| [B, 512, L/1024]| [1x4]  |
|            |             | BatchNorm, ReLU |        |
| Conv1d-11  | [B, 512, L/1024]| [B, 1024, L/2048]| [1x4]  |
|            |             | BatchNorm, ReLU |        |
| Conv1d-12  | [B, 1024, L/2048]| [B, 1024, L/4096]| [1x4]  |
|            |             | BatchNorm, ReLU |        |
| Embedding (fc) | [B, 1024] | [B, 8] | - |
|            |             | BatchNorm, ReLU |        |
| fc-1       | [B, 1024]   | [B, 32]     | -      |
|            |             | BatchNorm, ReLU |        |
| fc-2       | [B, 32]     | [B, 2]      | -      |

A. Stress Detection using Electrodermal Activity (EDA) with CLAS Dataset

Table [VI] shows the prediction accuracy in all types of supervised learning models. To examine the statistical significance of the accuracy, we conducted an ANOVA (post-hoc: Tukey) test, and the corresponding results are also included in Table [VI]. The prediction accuracy of the baseline framework was treated as the reference group. The domain classification loss-based model did not show a significant increase in accuracy. The MMD loss-based model showed a slight numerical increase but no statistically significant difference. The accuracy of the remaining three PiRL frameworks reached over 79% and they were all statistically significantly higher than the baseline results with smaller standard deviations. The best framework for detecting apnea was the combination of MMD and triplet loss since it achieved the highest prediction accuracy. The prediction accuracy of the person-specific model was highest which exceeded 95% and statistically higher than the baseline.

B. Sleep Apnea Detection using ECG with Apnea-ECG Dataset

Table [VII] shows the prediction accuracy of apnea detection using supervised learning models. The baseline model obtained a prediction accuracy of 75.2% with a standard deviation of 0.014. The accuracy of the domain classification loss-based model showed a slight numerical increase but no statistically significant difference. The accuracy of the remaining three PiRL frameworks reached over 79% and they were all statistically significantly higher than the baseline results with smaller standard deviations. The best framework for detecting apnea was the combination of MMD and triplet loss since it achieved the highest prediction accuracy. The prediction accuracy of the person-specific model was highest which exceeded 95% and statistically higher than the baseline.

C. Embedding Evaluation

The effects of our proposed PiRL framework on latent space similarities among participants are shown in this section. As full subject ID information on the testing set was missing in the apnea dataset, we evaluated the latent space similarities only using the CLAS dataset. Table [VIII] illustrates the latent space similarities of invariant representations by cosine and manhattan distance among pair-wise participants in the baseline and MMD model to further explore the influence of MMD loss. Although the cosine distance of the MMD model (mean: 0.9981) was numerically less than the baseline model (mean: 1.0075), there was no statistical difference under the ANOVA test regarding the baseline as the reference group. Meanwhile, the average value of the pair-wise manhattan distance is 380.3 and 306.9 for the baseline and MMD models respectively. The MMD model showed statistically shorter distance than the baseline model with the p-value equals to $2.2 \times 10^{-16}$ under the ANOVA test.

| TABLE VIII |
| Cosine and Manhattan Distance in the Baseline and MMD Model on the CLAS Dataset by ANOVA |
|----------|----------|-----------|
| Cosine Distance | 1.0075 | 0.9981 | × |
| Manhattan Distance | 380.27 | 306.94 | ✓ (2.2·10^{-16}) |

The embeddings in the latent space of each participant turned got closer to each other using our PiRL framework. Consequently, by quantifying the embedding similarities in the latent space among participants using the manhattan distance, the results successfully reveal that MMD loss is capable of generating more participant-invariant representations on the generic models and thus, improving the performance. The ineffectiveness of cosine distance in the evaluation might be due to the loss of the coordinate information in the high dimensional space. The cosine similarity, which is based on
cosine distance, measures the angle between vectors, therefore, if two embeddings have a small angle but are far apart in the original high dimensional space, they would not be considered similar.

V. CONCLUSIONS AND FUTURE WORK

In this work, we proposed PiRL, which utilizes MMD and triplet loss for learning participant-invariant representations, to improve the performance of generic health detection models. We evaluated the performance and effectiveness of our framework using two public datasets for mental and physical health. As preliminary results, we demonstrated that our proposed PiRL outperformed the baseline models and helped generic models achieve better performances. Performance improvement was not observed using DANN technique. In future work, we will investigate other approaches to further optimize and interpret the representations in health applications.

ACKNOWLEDGMENT

This work is supported by NSF #2047296 and #1840167.

REFERENCES

[1] Oh, S. L., Ng, E. Y., San Tan, R., Acharya, U. R. (2018). Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats. Computers in biology and medicine, 102, 278-287.

[2] Erdenebayar, U., Kim, Y. J., Park, J. U., Joo, E. Y., Lee, K. J. (2019). Deep learning approaches for automatic detection of sleep apnea events from an electrocardiogram. Computer methods and programs in biomedicine, 180, 105001.

[3] Han Yu and Akane Sano. Semi-supervised learning and data augmentation in wearable-based momentary stress detection in the wild. arXiv preprint arXiv:2202.12935

[4] Radhika, K., Oruganti, V. R. M. (2020, December). Transfer learning for subject-independent stress detection using physiological signals. In 2020 IEEE 17th India Council International Conference (INDICON) (pp. 1-6). IEEE.

[5] Radhika, K., Oruganti, V. R. M. (2021, January). Deep multimodal fusion for subject-independent stress detection. In 2021 11th International Conference on Cloud Computing, Data Science Engineering (Confluence) (pp. 105-109). IEEE.

[6] Gjoreski, M., Gjoreski, H., Luštrek, M., Gams, M. (2016, September). Continuous stress detection using a wrist device: in laboratory and real life. In proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing: Adjunct (pp. 1185-1193).

[7] Kogler, L., Müller, V. I., Chang, A., Eichhoff, S. B., Fox, P. T., Gur, R. C., Derntl, B. (2015). Psychosocial versus physiological stress—Meta-analyses on deactivations and activations of the neural correlates of stress reactions. Neuroimage, 119, 235-251.

[8] Nakashima, Y., Kim, J., Flihura, S., Seiderer, A., André, E. (2015, September). Stress recognition in daily work. In International Symposium on Pervasive Computing Paradigms for Mental Health (pp. 23-33). Springer, Cham.

[9] Nkuriyiyezu, K., Yokokubo, A., Lopez, G. (2019). The effect of person-specific biometrics in improving generic stress predictive models. arXiv preprint arXiv:1910.01770.

[10] Schmidt, P., Reiss, A., Dürrichen, R., Marberger, C., Van Laerhoven, K. (2018, October). Introducing wesad, a multimodal dataset for wearable stress and affect detection. In Proceedings of the 20th ACM international conference on multimodal interaction (pp. 400-408).

[11] Zenonos, A., Khan, A., Kalogridis, G., Vatsikas, S., Lewis, T., Sooriyabandara, M. (2016, March). HealthyOffice: Mood recognition at work using smartphones and wearable sensors. In 2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops) (pp. 1-6). IEEE.

[12] Bsoul, M., Minn, H., Tanil, L. (2010). Apnea MedAssist: real-time sleep apnea monitor using single-lead ECG. IEEE transactions on information technology in biomedicine, 15(3), 416-427.

[13] Nath, D., Singh, M., Sethia, D., Kalra, D., Indu, S. (2020, March). A comparative study of subject-dependent and subject-independent strategies for EEG-based emotion recognition using LSTM network. In Proceedings of the 2020 the 4th International Conference on Computer

---

**TABLE V**

| Person-specific models’ data separation | #Samples in EDA dataset | #Samples in Apnea-ECG dataset |
|----------------------------------------|-------------------------|------------------------------|
| Training Set (Tr1)                     | 695                     | 11907                        |
| Testing Set (Te1)                      | 298                     | 5103                         |
| Total                                  | 993                     | 17010                        |

**TABLE VI**

| Performance in stress detection on CLAS dataset. P-values are calculated by ANOVA (post-hoc: Tukey) | Baseline | DANN | MMD | Triplet | MMD+Triplet | Person-Specific |
|-----------------------------------------------|----------|------|-----|---------|-------------|------------------|
| Accuracy                                     | 64.3%    | 64.5%| 66.5%| 70.1%   | 70.6%       | 86.8%            |
| SD                                           | 0.012    | 0.010| 0.014| 0.011   | 0.010       | 0.189            |
| P-value < 0.01                               | ✓        | ✓    | ✓   | ✓       | ✓           | ✓                |

**TABLE VII**

| Performance in sleep apnea detection on Apnea-ECG dataset. P-values are calculated by ANOVA (post-hoc: Tukey) | Baseline | DANN | MMD | Triplet | MMD+Triplet | Person-Specific |
|-----------------------------------------------|----------|------|-----|---------|-------------|------------------|
| Accuracy                                     | 75.2%    | 75.7%| 79.5%| 79.1%   | 79.9%       | 95.7%            |
| SD                                           | 0.014    | 0.013| 0.010| 0.013   | 0.009       | 0.011            |
| P-value < 0.01                               | ✓        | ✓    | ✓   | ✓       | ✓           | ✓                |
[14] Healey, J. A., Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. IEEE Transactions on intelligent transportation systems, 6(2), 156-166.

[15] Koldijk, S., Neerincx, M. A., Kraaij, W. (2016). Detecting work stress in offices by combining unobtrusive sensors. IEEE Transactions on affective computing, 9(2), 227-239.

[16] Nakashima, Y., Kim, J., Flutura, S., Seiderer, A., André, E. (2015, September). Stress recognition in daily work. In International Symposium on Pervasive Computing Paradigms for Mental Health (pp. 23-33). Springer, Cham.

[17] Wu, M. H., Chang, E. J., Chu, T. H. (2018, April). Personalizing a generic ECG heartbeat classification for arrhythmia detection: a deep learning approach. In 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPPR) (pp. 92-99). IEEE.

[18] Li, Y., Pang, Y., Wang, J., Li, X. (2018). Patient-specific ECG classification by deeper CNN from generic to dedicated. Neurocomputing, 314, 336-346.

[19] Bethge, D., Hallgarten, P., Grosse-Puppendahl, T., Kari, M., Mikut, R., Schmidt, A., Özdenizci, O. (2022, May). Domain-Invariant Representation Learning from EEG with Private Encoders. In ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1236-1240). IEEE.

[20] Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., Smola, A. (2012). A kernel two-sample test. The Journal of Machine Learning Research, 13(1), 723-773.

[21] Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., ... Lempitsky, V. (2016). Domain-adversarial training of neural networks. The journal of machine learning research, 17(1), 2096-2030.

[22] Schroff, F., Kalenichenko, D., Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 815-823).

[23] Van der Maaten, L., Hinton, G. (2008). Visualizing data using t-SNE. Journal of machine learning research, 9(11).

[24] Markova, V., Ganchev, T., Kalinkov, K. (2019, November). Clas: A database for cognitive load, affect and stress recognition. In 2019 International Conference on Biomedical Innovations and Applications (BIA) (pp. 1-4). IEEE.

[25] Thomas Penzel, George B Moody, Roger G Mark, Ary L Goldberger, and J Hermann Peter. The apnea-ecg database. In Computers in Cardiology 2000. Vol. 27 (Cat. 00CH37163), pages 255–258. IEEE, 2000.