Deep Subdomain Adaptation Network Improves Cross-Subject Mental Workload Classification

Wu Sun
Laboratory for Brain-Bionic Intelligence and Computational Neuroscience
Wuyi University
China
sw450236419@gmail.com

Junhua Li [1,2]
[1]School of Computer Science and Electronic Engineering
University of Essex
UK
junhua.li@essex.ac.uk

[2]Laboratory for Brain-Bionic Intelligence and Computational Neuroscience
Wuyi University
China
juhalee.bcmi@gmail.com

***Abstract*—Cross-subject classification is of great practical value in the mental monitoring. The trained model on a person can be transferred to another person without retraining. To date, it is achieved by global domain adaptation without considering differences in subdomain distributions. In this case, there is a lack of sensitivity to specific information associated with each category. To solve this problem, we proposed a deep subdomain adaptation network (DSAN) to estimate mental workload levels across different persons. In the proposed DSAN, the first temporal and spatial layers were designed as a feature extractor. The features extracted by the feature extractor were aligned between the source samples and target samples in each subdomain separately. The alignment loss calculated by local maximum mean difference (LMMD) was back-propagated to update the weights of the feature extractor to enhance the feature extraction performance. Subdomain adaptation was achieved over iterations during the model training. The proposed subdomain adaptation is not specialized for a particular feature extractor, as shown in this paper. It is universal and can be applied after any feature extractors. Two datasets (Dataset MATB and Dataset SFE) were used to evaluate the proposed DSAN. The results showed that the proposed DSAN outperformed the compared methods in terms of classification accuracy, showing an elevation of 3%~7%. This study provides an effective solution for the cross-subject mental workload classification and will promote practical applications of mental workload monitoring.**

Keywords*—*Deep Subdomain Adaptation Network, EEG, Cross-Subject Classification, Mental Workload, Local Maximum Mean Difference, Brain Computer Interface

I. Introduction

In the field of brain-computer interfaces, researchers have shown increasing interest in the mental workload (MW) paradigm [1]. The evaluation of MW helps to optimize task execution and decision-making processes in various environments, thus increasing efficiency and reducing error rates. It can be expressed as the proportion of mental resources utilized for a specific task compared to the overall mental capacity of a person [2]. Either extremely low or too high a workload can lead to a negative impact on human performance in task implementation [3]. Therefore, it is paramount to accurately determine the level of MW so that the appropriate MW can be maintained to maximize productivity.

Up to this point, MW has been assessed by subjective and objective metrics [4]. The subjective measurement relies on the human's perception and self-evaluation using a predefined questionnaire, such as the NASA task load index [5] and the primary task performance method [6]. Though subjective measurements can be easily carried out, they cannot provide a real-time and objective assessment. In contrast, objective measurements depend on physiological signals, including electroencephalogram (EEG), electrocardiogram (ECG) [7], and functional near-infrared spectroscopy (fNIRS) [8]. Among these signals, EEG is one of the most frequently utilized signals because of its high temporal resolution, portability, safety, and cost-effectiveness. Hence, we utilized EEG in this study to classify different levels of mental workload.

Various methodologies have been devised to categorize mental workload levels using EEG data, such as k-nearest neighbors (k-NN) [9], random forest (RF) [10], and support vector machine (SVM) [11]. In addition, the studies [12], [13], [14] have shown that deep learning models have surpassed traditional machine learning methods on the tasks of within-subject MW classification. However, the presence of individual variability in EEG data presents a difficulty in cross-subject workload studies. Applying subject-specific models to new subjects may result in a decline in recognition accuracy for the model. It is essential to create mental workload recognition models that can be applied across different subjects to improve generalization in real-world scenarios.

In response to this concern, a variety of domain adaptable approaches have been developed [15], [16], [17]. Domain adaptation (DA) considers the data from some subjects as the source domain and the data from a new subject as the target domain, which then transfers knowledge from the source domain to the target domain. A typical DA algorithm [18] seeks to acquire domain-invariant features that can be applied to cross-subject classifications by minimizing the disparities in data distribution between domains. Thereby, models that have been trained using source data can make predictions for the data in the target domain. Notable instances are deep adaptation neural networks (DANN) [18], transfer joint matching (TJM) [19], minimal mean difference (MMD) [20], and deep domain adaptation (DDA) [21]. Following global domain adjustment, classifiers have the potential to effectively categorize samples from both domains [20], [21]. Nevertheless, these methods consider the data distribution as a whole and disregard individual distributions existing in each category, which results in inadequate and inaccurate transferring from the source domain to the target domain.

In order to address the aforementioned issue, we proposed to consider each category as a distinct subdomain. In this

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

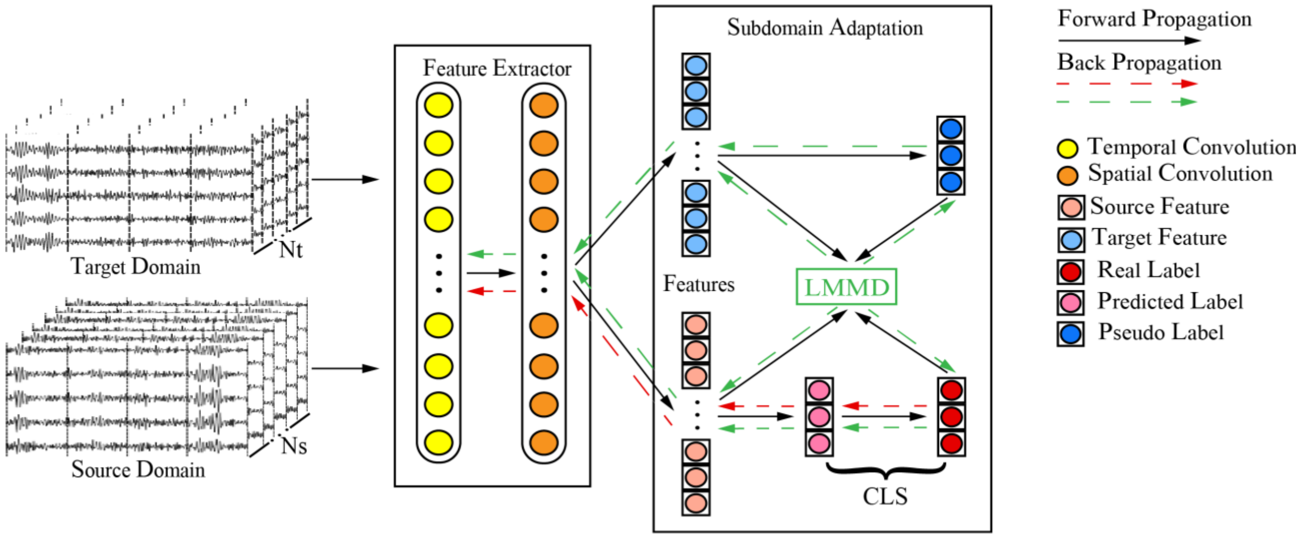

Fig. 1. DSAN model's architectural design.

study, the categories are mental workload levels (i.e., low, medium, and high). The objective is to minimize variations in subdomains across subjects in the estimation of mental workload. Aligning data distributions between the source and target domains within each subdomain can enhance the performance of mental workload classification. To this end, we proposed a deep subdomain adaptation network (DSAN). In the proposed DSAN, any feature extractors that extract mental workload-related features can be embedded. These features' distributions are aligned between the source domain and the target domain for each category, respectively. Because the alignment is done separately for each category, data labels are required. It is straightforward for source domain data as both data and labels are available. However, the target domain data lacks available labels. Paired data from the source domain was used to train a label classifier. Then, this label classifier was used to generate pseudo-labels for the data in the target domain. Local maximum mean difference (LMMD) was adopted to measure the alignment. LMMD loss was back-propagated to calculate gradients for updating network weights towards the minimization of LMMD loss. During the back-propagation, the weights of the feature extractor embedded in the DSAN were also updated to optimize the extracted features. Two different datasets were used to evaluate the proposed DSAN.

II. METHOD

In this study, we aimed to classify mental workload across human subjects. This is more meaningful from the viewpoint of practical application because the trained model can be directly applied to a new subject without retraining. Leave-one-subject-out cross validation was employed in model evaluation. The data from a subject is the target domain data $D_t = \{(x_j^t)\}_{j=1}^{N_t}$, and the data from the remaining subjects are the source domain data $D_s = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$, where N_t denotes the quantity of unlabeled samples and N_s denotes the quantity of labeled samples. It is expected that the source and target domains share the same feature space and label space, i.e., $x^s, x^t \in R^{C\times T}$ and $y^s \in \{1,2,\cdots,K\}$. Here, C denotes the quantity of channels in EEG signal for each sample, T denotes the quantity of data points in each channel, and K denotes the quantity of categories of MW. In usual, the marginal probability distributions of these two domains exhibit dissimilarities, namely $P_s(x^s) \neq P_t(x^t)$. Therefore, we extracted MW-related features and aligned the features' distribution to solve this problem. As depicted in Fig. 1, we used a domain-sharing convolutional network as a feature extractor. This network extracts high-level feature representations from preprocessed EEG data. Subsequently, the extracted features were fed into the subdomain adaptation module to achieve two objectives. The primary objective is to predict the labels of the data from either the source or target domain based on the inputted features. The second objective is to quantify the disparities between features' distributions between the source domain and the target domain for each category (subdomain) using LMMD and minimize the differences between the domains.

A. *Feature Extractor and Label Classifier*

Feature Extractor: The purpose of the feature extractor is to encode preprocessed EEG data and generate features that facilitate the following distribution alignment of the features.

TABLE I. ARCHITECTURE PARAMETERS OF DSAN.

	Layer	Kernel Size	Input Size	Output Size	Activation
Temporal Conv	Conv_1[a]	(1, 13)	(1, n, T)	(P1[b], n, T)	GELU
Spatial Conv	Conv_2	(n, 1)	(P1, n, T)	(P2, 1, T)	GELU
	Pooling	(1, 35)	(P2[b], 1, T)	(P2, 1, T//[c]35)	
Projection	Linear		(P2*T//35, 1)	(1024, 1)	
Classifier	Fully-Connected		(1024, 1)	(512, 1)	GELU
	Fully-Connected		(512, 1)	K	

[a] The conv_1 layer contains the padding operation.

[b] P1 and P2 are the numbers of filters, and n denotes the number of channels.

[c] // represents the floor division.

One-dimensional temporal convolution was used to extract temporal characteristics contained in the EEG signal, and the other one-dimensional spatial convolution was used to extract spatial characteristics. The Gaussian error linear unit (GELU) activation function was utilized to enhance the nonlinearity of the model and prevent issues such as gradient explosion and vanishing. Then, the features outputted from convolution layers were fed into a pooling layer to down-sample features. In this paper, the input EEG data (x^s, x^t) were mapped to the features $[f^s = G_f(x^s; \theta_f), f^t = G_f(x^t; \theta_f)]$ through feature extractor G_f, using the parameters θ_f.

Label Classifier: To guarantee the classifier's prediction accuracy, the training process utilizes data from the source domain. The extracted features representing temporal and spatial characteristics of EEG were inputted into the classifier G_c with the parameters θ_c to obtain the predicted label $\widehat{y^s}$. In other words, $\widehat{y^s} = G_c(G_f(x^s; \theta_f); \theta_c)$. It comprises of two fully-connected layers, with the first layer utilizing the GELU activation function and the output values from the second layer are directly utilized to compute the cross-entropy loss. which is mathematically described as follows:

$$\mathcal{L}_{cls} = -\frac{1}{N_s}\left[\sum_{i=1}^{N_s}\sum_{k=1}^{K} I[y_i^s = k]\log\left(G_c\left(G_f\left(x_i^s; \theta_f\right); \theta_c\right)\right)\right] \quad (1)$$

If $y_i^s = k$, then the value of $I[y_i^s = k]$ is 1; Otherwise, it is 0. Table I displays the specific structural characteristics of the feature extractor and label classifier.

B. *LMMD-Based Subdomain Adaptation*

In the domain adaption, all data are treated as a whole. In this case, detailed information about the data distributions for each category cannot be captured. In this study, we treated each MW category separately as a subdomain and adapted each subdomain to retain complete information using local maximum mean difference (LMMD) [22]. LMMD provides more refined subdomain alignment by measuring the discrepancy between kernel-mean-embedding-related subdomains in the Hilbert space.

$$\mathcal{L}_{\text{LMMD}}\left(p^{(k)}, q^{(k)}\right) \triangleq E_k \left\| E_{p^{(k)}}[\phi(f^s)] - E_{q^{(k)}}[\phi(f^t)] \right\|_H^2 \quad (2)$$

Here, $p^{(k)}$ and $q^{(k)}$ denote the distributions of $D_s^{(k)}$ and $D_t^{(k)}$, respectively. H represents a reproducing kernel Hilbert space (RKHS) with a kernel h. Where $h(f^s, f^t) = \langle \phi(f^s), \phi(f^t) \rangle$. $\phi(\cdot)$ represents the feature mapping that projects the feature (f) to the RKHS. $\langle \cdot,\cdot \rangle$ denotes inner product of vectors. We add a weight parameter w^k to weight samples of different categories. Therefore, optimizing formula (2) can be expressed as

$$\mathcal{L}_{\text{LMMD}} = \frac{1}{K}\sum_{k=1}^{K}\left\| \sum_{x_i^s \in D^s} w_i^{sk}\phi(f_i^s) - \sum_{x_j^t \in D^t} w_j^{tk}\phi(f_j^t) \right\|_H^2 \quad (3)$$

Among them, w_i^{sk} and w_i^{tk} represent the weights of f_i^s and f_i^t for category K, respectively. The weights were calculated by

$$w_i^{sk} = \frac{y_i^{sk}}{\sum_{(x_j, y_j) \in D_s} y_j^{sk}} \qquad or \qquad w_i^{tk} = \frac{y_i^{tk}}{\sum_{(x_j, y_j) \in D_t} y_j^{tk}} \quad (4)$$

where y_i^{sk} and y_i^{tk} are the kth class entries of vector y_i^s and y_i^t, respectively. For the source domain data, we can calculate w_i^{sc} using the available label y^s according to the formula (4). However, there are no available labels for the target domain data to compute weights. Here, we used pseudo labels $\hat{y}^t$ outputted by the label classifier to replace y^t for calculating the weights w^t of the target domain data.

Embed the LMMD algorithm into the model. In each batch, f^s, f^t and the real label y^s of the source domain and the pseudo label $\hat{y}^t$ obtained by the classifier of the target domain are used together as inputs to LMMD. Therefore, the calculation of subdomain distribution differences is as follows:

$$\begin{aligned}\mathcal{L}_{LMMD} = \frac{1}{K}\sum_{k=1}^{K}\Bigg[&\sum_{i=1}^{n_s}\sum_{j=1}^{n_s} w_i^{sk} w_j^{sk} h\left(f_i^s, f_j^s\right) \\ &+ \sum_{i=1}^{n_t}\sum_{j=1}^{n_t} w_i^{tk} w_j^{tk} h\left(f_i^t, f_j^t\right) \\ &- 2\sum_{i=1}^{n_s}\sum_{j=1}^{n_t} w_i^{sk} w_j^{tk} h\left(f_i^s, f_j^t\right)\Bigg]\end{aligned} \quad (5)$$

The cost function of the proposed DSAN model comprises two components: the error of the label classifier and LMMD loss of the subdomain adaptation, which can be expressed by the addition of formulas (1) and (5).

$$\mathcal{L}_{\text{total}}\left(\theta_f, \theta_c\right) = \mathcal{L}_{\text{cls}} + \lambda \mathcal{L}_{\text{LMMD}} \quad (6)$$

Where $\lambda = 1$.

III. Datasets and Results

A. *Description of Datasets*

To comprehensively evaluate the proposed DSAN for classifying levels of mental workload across subjects, we used two different datasets: simulated flight experiment (SFE) [14], [23] and multi-attribute task battery (MATB) [24].

The SFE dataset includes 7 human subjects who operated a simulated aircraft using a joystick and a keyboard. Three operation difficulties for piloting the aircraft were imposed to induce three levels of mental workload, respectively. Each subject performed three identical sessions, each consisting of three levels of mental workload tasks (i.e., low, medium, and high), lasting 6 minutes (2 minutes for each task).

The MATB dataset includes 15 human subjects who performed NASA MATB tasks. Each of the subjects completed three experimental sessions. Each experiment records a one-minute resting state and a fifteen-minute task state. The workload level was adjusted by giving different numbers of subtasks and different complexities of each subtask, resulting in low, medium, and high levels of workload.

B. *Methods Comparison*

We compared the proposed DASN with other methods in terms of classification accuracy of mental workload. To ensure a fair comparison, the same feature extractor was used in this study. The compared methods are briefly listed below. Shallow-Net: A traditional deep learning model designed for encoding features of EEG signals, which does not include domain adaptation [25]. Custom domain adaptation (CDA):

TABLE II. COMPARISON RESULTS FOR THE DATASET SFE (CLASSIFICATION ACCURACY IN %)

Method / Subject	Shallow-Net	CDA	DDA	CORAL	DANN	DSAN
1	70.37	80.00	78.89	80.83	80.56	80.19
2	57.22	69.44	70.28	67.22	69.44	72.59
3	46.48	42.22	42.50	42.22	41.67	39.07
4	51.50	57.22	57.50	51.94	53.06	57.22
5	43.33	56.39	59.44	56.39	59.72	56.30
6	65.37	72.22	67.22	65.56	63.33	75.93
7	51.48	55.00	55.83	56.39	54.44	56.67
Mean	55.11	61.79	61.67	60.08	60.32	62.57
STD	9.12	11.82	10.86	11.49	11.57	13.33

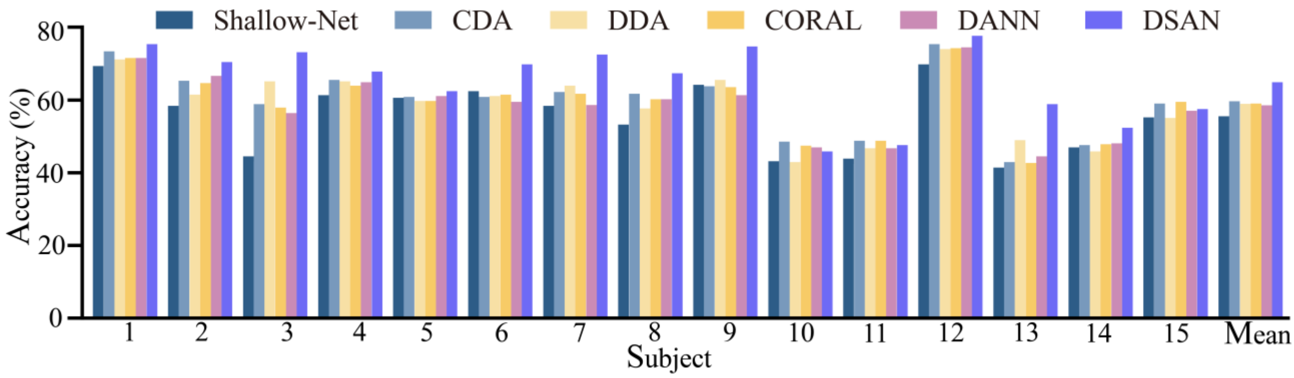

Fig. 2. Comparison results for the dataset MATB.

Incorporating maximum mean difference (MMD) into deep neural networks to align distribution disparities across domains [20]. Deep domain adaptation (DDA): Employing adversarial learning and MMD techniques to mitigate distribution disparities among diverse domains [21]. Correlation alignment (CORAL): It utilizes correlation alignment to tackle the issue of distribution divergence across several domains [25]. Domain adversarial neural network (DANN): Utilizing adversarial learning techniques to address feature distribution challenges across diverse domains [18].

The comparison results of methods on the SFE dataset are listed in Table II. Fig. 2 shows the comparison results obtained based on the MATB dataset. The comparison results derived from both datasets showed that the lowest classification accuracy was achieved by the Shallow-Net, which did not include domain adaptation. The classification accuracy was greatly increased due to domain adaptation. This reflects that domain adaption was useful for cross-subject classification because shared features across subjects could be extracted with the help of domain adaptation. The results show an improvement of 3%-5%. Our method aligns features from a refined perspective, achieving clearer classification boundaries. The results have improved by 3%-7%.

C. *Does It Rely on Feature Extractor?*

In order to investigate whether the proposed model is universal and does not rely on a particular feature extractor, we employed two popular networks (i.e., EEGNET and Conformer [27], [28]) to replace the used feature extractor. The results are displayed in Table III. It is clear that the classification accuracy of mental workload was elevated from the condition without (W/O) subdomain adaptation to the

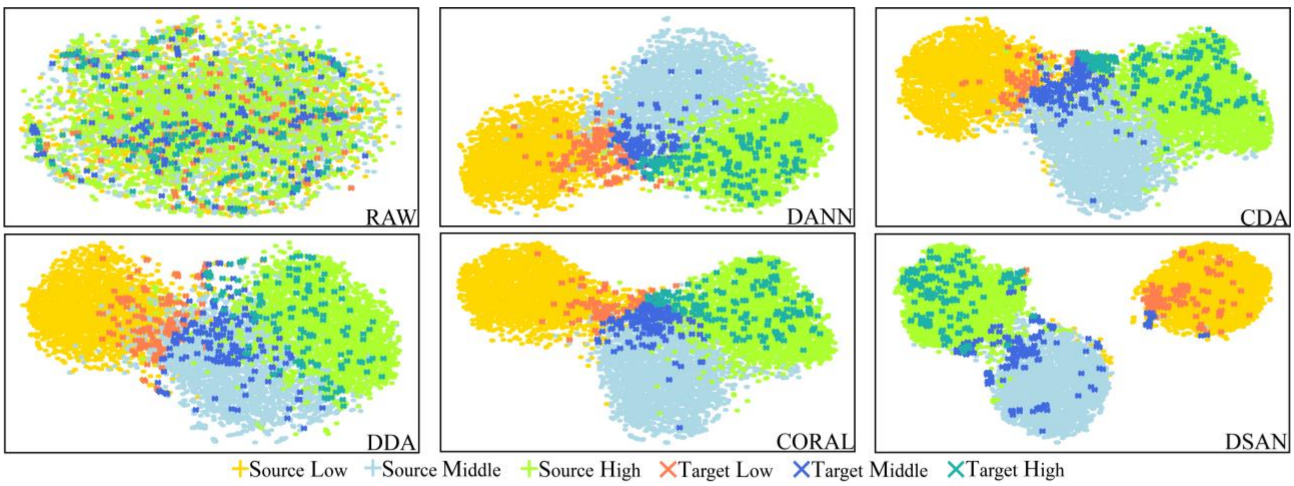

Fig. 3. Visualization of features extracted by the feature extractor with the help of different domain adaptation methods.

TABLE III. CLASSIFICATION ACCURACY COMPARISONS (MEAN ± STD %) ON SFE AND MATB.

Method / Dataset	EEGNET (W/O)	EEGNET (SA)	Conformer (W/O)	Conformer (SA)
SFE	54.57±8.1	63.33±10.7	57.51±7.9	63.13±10.5
MATB	53.14±6.9	61.04±15.1	57.17±10.6	65.03±10.2

condition with subdomain adaptation (SA) for both feature extractors (i.e., EEGNET and Conformer). These results demonstrated that it does not rely on a particular feature extractor. It should work for other feature extractors.

D. Domain Adaptation to Feature Extraction

In the backpropagation, the loss from the domain adaptation contributes to weight updating of the feature extractor. We, therefore, attempt to explore the effect of domain adaptation methods on feature extraction in the feature extractor. In order to visualize high-dimensional features derived from the feature extractor in a two-dimensional plane, t-distributed stochastic neighbor embedding (t-SNE) [29] was utilized to reduce dimensionality. The data from subject 9 in the dataset MATB were used as an illustrative example. The distributions of raw EEG (RAW) and the features extracted with the help of DANN, CDA, DDA, CORAL, and DSAN are depicted in Fig. 3. It can be seen that DSAN is the best among these methods, and the features of the source and target domains are closely clustered when they are from the same category. The clusters of each category are more separable. For the other domain adaptation methods, features from the target domain are interleaved around the boundaries among categories. This issue can be significantly mitigated by DSAN due to the subdomain adaptation. Another advantage of DSAN is that the features of the source and target domain overlap more for each category. These results suggest that separate feature alignment for each subdomain can yield more domain-invariant features, which benefits the cross-subject classification.

IV. CONCLUSIONS

This study proposed a subdomain adaptation-based transfer learning model (DSAN) to classify mental workload levels, which achieved better performance in cross-subject classification. DSAN aligns the feature distributions between the source domain and the target domain separately for each category while forcing the feature extractor to generate more domain-invariant features, collectively resulting in better classification performance in the mental workload classification. In addition, our study demonstrated that it does not rely on a particular feature extractor and is applicable to any feature extractor. The visualization of features further confirmed the advantage of DSAN, showing more separable features among categories and more overlapped features between the source domain and target domain. Our subsequent work is to further enhance DSAN performance by taking into account the confidence of pseudo labels and integrating contrastive learning methods.

REFERENCES

[1] Wickens C D.: Multiple resources and performance prediction. Theoretical issues in ergonomics science 3(2): 159-177 (2002).

[2] Zhou Y, Huang S, Xu Z, et al.: Cognitive workload recognition using EEG signals and machine learning: A review. IEEE Transactions on Cognitive and Developmental Systems, (2021).

[3] Heard J, Harriott C E, Adams J A.: A survey of workload assessment algorithms. IEEE Transactions on Human-Machine Systems, 48(5): 434-451 (2018).

[4] Reid G B, Nygren T E.: The subjective workload assessment technique: A scaling procedure for measuring mental workload. Advances in psychology. North-Holland, 52: 185-218 (1988).

[5] Hart S G, Staveland L E.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Advances in psychology. North-Holland, 52: 139-183 (1988).

[6] Tao D, Tan H, Wang H, et al.: A systematic review of physiological measures of mental workload. International journal of environmental research and public health, 16(15): 2716 (2019).

[7] Qu, H., Gao, X., Pang, L.: Classification of mental workload based on multiple features of ECG signals. Inform. Med. Unlocked (2021).

[8] Shimizu T, Hirose S, Obara H, et al.: Measurement of frontal cortex brain activity attributable to the driving workload and increased attention. SAE International Journal of Passenger Cars-Mechanical Systems, 2(2009-01-0545): 736-744 (2009).

[9] Ko L W, Chikara R K, Lee Y C, et al.: Exploration of user's mental state changes during performing brain–computer interface. Sensors, 20(11): 3169 (2020).

[10] Pei Z, Wang H, Bezerianos A, et al.: EEG-based multiclass workload identification using feature fusion and selection. IEEE Transactions on Instrumentation and Measurement, 70: 1-8 (2020).

[11] Lim W L, Sourina O, Liu Y, et al.: EEG-based mental workload recognition related to multitasking. Communications and Signal Processing (ICICS). IEEE, 1-4 (2015).

[12] Asgher U, Khalil K, Ayaz Y, et al.: Classification of mental workload (MWL) using support vector machines (SVM) and convolutional neural networks (CNN). Mathematics and Engineering Technologies (MET). IEEE, 1-6 (2020).

[13] Shao S, Han G, Wang T, et al.: EEG-Based Mental Workload Classification Method Based on Hybrid Deep Learning Model Under IoT. IEEE Journal of Biomedical and Health Informatics (2023).

[14] Yu Y, Bezerianos A, Cichocki A, et al.: Latent Space Coding Capsule Network for Mental Workload Classification. IEEE Transactions on Neural Systems and Rehabilitation Engineering (2023).

[15] Tang X, Zhang X.: Conditional adversarial domain adaptation neural network for motor imagery EEG decoding. Entropy, 22(1): 96 (2020).

[16] Zhou Y, Xu Z, Niu Y, et al.: Cross-task cognitive workload recognition based on EEG and domain adaptation. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 30: 50-60 (2022).

[17] Jiménez-Guarneros M, Gómez-Gil P.: Custom Domain Adaptation: A new method for cross-subject, EEG-based cognitive load recognition. IEEE Signal Processing Letters, 27: 750-754 (2020).

[18] Ganin Y, Lempitsky V.: Unsupervised domain adaptation by backpropagation. International conference on machine learning. PMWR, 1180-1189 (2015).

[19] Zhou Y, Xu Z, Niu Y, et al. Cross-task cognitive workload recognition based on EEG and domain adaptation. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 30: 50-60 (2022).

[20] Borgwardt K M, Gretton A, Rasch M J, et al.: Integrating structured biological data by kernel maximum mean discrepancy. Bioinformatics, 22(14): 49-57 (2006).

[21] Zhou Y, Wang P, Gong P, et al.: Cross-subject Cognitive Workload Recognition Based on EEG and Deep Domain Adaptation. IEEE Transactions on Instrumentation and Measurement, (2023).

[22] Zhu Y, Zhuang F, Wang J, et al.: Deep subdomain adaptation network for image classification. IEEE transactions on neural networks and learning systems, 32(4): 1713-1722 (2020).

[23] Pei Z, Wang H, Bezerianos A, et al.: EEG-based multiclass workload identification using feature fusion and selection[J]. IEEE Transactions on Instrumentation and Measurement, 70: 1-8 (2020).

[24] [Online].Available:https://www.neuroergonomicsconference.um.ifi.lmu.de/pbci/

[25] Schirrmeister R T, Springenberg J T, Fiederer L D J, et al.: Deep learning with convolutional neural networks for EEG decoding and visualization. Human brain map, 38(11): 5391-5420 (2017).

[26] Zhong X C, Wang Q, Liu D, et al.: A deep domain adaptation framework with correlation alignment for EEG-based motor imagery classification. Computers in Biology and Medicine, (2023).

[27] Lawhern V J, Solon A J, Waytowich N R, et al.: EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. Journal of neural engineering, 15(5): 056013 (2018).

[28] Song Y, Zheng Q, Liu B, et al.: EEG conformer: Convolutional transformer for EEG decoding and visualization. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 31: 710-719 (2022).

[29] Van Der Maaten L.: Accelerating t-SNE using tree-based algorithms. The journal of machine learning research, 15(1): 3221-3245 (2014)