AdaptEEG: A Deep Subdomain Adaptation Network with Class Confusion Loss for Cross-Subject Mental Workload Classification

Wu Sun, and Junhua Li*, *Senior Member, IEEE*

***Abstract*—EEG signals exhibit non-stationary characteristics, particularly across different subjects, which presents significant challenges in the precise classification of mental workload levels when applying a trained model to new subjects. Domain adaptation techniques have shown effectiveness in enhancing the accuracy of cross-subject classification. However, current state-of-the-art methods for cross-subject mental workload classification primarily focus on global domain adaptation, which may lack fine-grained information and result in ambiguous classification boundaries. We proposed a novel approach called deep subdomain adaptation network with class confusion loss (DSAN-CCL) to enhance the performance of cross-subject mental workload classification. DSAN-CCL utilizes the local maximum mean discrepancy to align the feature distributions between the source domain and the target domain for each mental workload category. Moreover, the class confusion matrix was constructed by the product of the weighted class probabilities (class probabilities predicted by the label classifier) and the transpose of the class probabilities. The loss for maximizing diagonal elements and minimizing non-diagonal elements of the class confusion matrix was added to increase the credibility of pseudo-labels, thus improving the transfer performance. The proposed DSAN-CCL method was validated on two datasets, and the results indicate a significant improvement of 3~10 percentage points compared to state-of-the-art domain adaptation methods. In addition, our proposed method is not dependent on a specific feature extractor. It can be replaced by any other feature extractor to fit new applications. This makes our approach universal to cross-domain classification problems.**

***Index Terms*— Deep Subdomain Adaptation Network, EEG, Brain Computer Interface, Mental Workload, Deep Learning, Cross-Subject Classification.**

I. Introduction

MENTAL workload (MW) describes the amount of psychological and cognitive effort required to perform a specific task or activity [1], [2]. Too low MW might cause boredom, lack of desire, inattention, and blunders. Too high MW might impair a person's decision-making capability and work efficiency, as well as lead to physical and mental health issues [3]. Therefore, it is crucial to recognize MW accurately so as to enhance work efficiency, lower error rates, and improve personal health [4].

When assessing the MW state, subjective and objective measurements are commonly employed. Subjective measurements commonly involve the use of questionnaires [5], or direct communication between researchers and subjects [6]. On the other hand, objective measurements rely on physiological signals such as electroencephalography (EEG) [7], electrocardiography (ECG) [8], and functional near-infrared spectroscopy (fNIRS) [9]. Objective measurements are considered more reliable and accurate as they are based on biological indicators rather than subjective perceptions. Among these physiological signals, EEG is particularly favored by researchers due to its excellent temporal resolution, safety, and cost-effectiveness.

Recent research [10], [11], [12] shows that deep learning models outperform traditional machine learning methods, such as k-nearest neighbors [13], random forest [14], and support vector machine [15], in the within-subject MW classification. In the studies involving cross-subject MW recognition, inter-subject variability in EEG data poses a challenge to accurate recognition. The application of subject-specific models to new subjects could potentially result in a reduction in recognition accuracy for the model. To improve the generalizability of the model in real-world situations, it is advisable to create MW recognition models that are suitable for diverse subjects.

Domain adaptation (DA) approaches may be an effective way to address the challenge of the low generalization to new subjects, which adapts the learned knowledge (source domain) to fit new subjects (target domain). Combining adaptation methods and deep learning networks has grown in popularity due to their outstanding feature extraction capabilities. It can be divided into two categories. The first category is based on adversarial learning methods [16], [17], [18], with the core idea being to achieve feature transformation between source and target domains by training feature extractors and domain

W. Sun is with the Laboratory for Brain-Bionic Intelligence and Computational Neuroscience, Wuyi University, Jiangmen 529020, China.

J. Li is with the School of Computer Science and Electronic Engineering, University of Essex, Colchester CO4 3SQ, UK, and also with the Laboratory for Brain-Bionic Intelligence and Computational Neuroscience, Wuyi University, Jiangmen 529020, China. He was with the Singapore Institute for Neurotechnology, National University of Singapore, Singapore.

* Indicates the corresponding author (Junhua Li, junhua.li@essex.ac.uk; juhalee.bcmi@gmail.com).

classifiers in adversarial competition. Thereby, the model's generalization is improved in the target domain. The second category is the metric-based technique [19], [20], [21], which assesses the distance or similarity between the source and target domains.

Recently, the employment of DA methods to handle the cross-task or cross-subject problem with EEG data has contributed to an increase in the performance and reliability of MW recognition. For example, Zhou et al. [22] and Guan et al. [23] attempted to assess cross-task MW by combining manually extracted features with traditional transfer learning methods, such as transfer component analysis (TCA) and joint distribution analysis (JDA) and achieved higher recognition accuracy than non-transfer learning methods. Many researchers have combined DA methods and deep learning networks to solve the MW cross-subject problem. For example, Guarneros et al. [24] proposed the custom domain adaptation (CDA) model, which embeds the maximum mean discrepancy (MMD) loss in the deep learning model to reduce the distributional differences between subjects. On this basis, Zhou et al. [25] proposed combining adversarial learning with MMD loss after feature extraction to increase the network's capacity to extract domain-invariant features. Yin et al. [26] used an adaptive stack denoising autoencoder (SDAE) to handle a cross-subject classification task, allowing for continuous monitoring of an operator's mental strain level in a collaborative human-computer context while warning of transient performance decline. The above methods can effectively classify samples from both domains. However, these approaches consider the data distribution as a whole, ignoring specific distributions that occur within each category. This results in an insufficient and erroneous transfer from the source domain to the target domain.

The transfer learning method mentioned above can effectively align overall feature distributions across different domains but lacks fine-grained information, which results in blurry classification boundaries. If each category is considered as an independent subdomain and their feature distributions are aligned in respective subdomains, it might result in distinct classification boundaries for enhancing the performance of cross-subject MW classification. We drew Fig. 1 to illustrate the differences of feature alignments achieved by global DA and subdomain adaptation (SA).

We proposed a novel deep SA network that integrates CCL

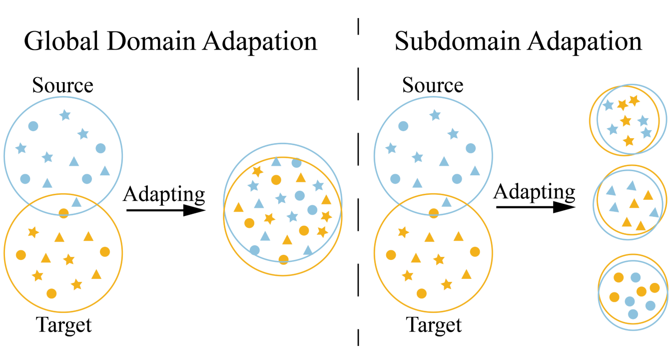

Fig. 1. Illustration of global DA and SA. Global DA ignores fine-grained information during feature alignment, whereas SA aligns features separately for each category.

to enhance the precise and effective classification of cross-subject MW. The DSAN-CCL model initially employs temporal and spatial convolutions to extract MW features from both the source and target domains. The alignment of subdomain feature distributions between the source and target domains is achieved using local maximum mean discrepancy (LMMD). CCL aims to mitigate the ambiguity of the pseudo-labels for the data in the target domain so that the credibility of pseudo-labels is enhanced. The contributions of this paper are delineated in the following three points.

1) We proposed the model DSAN-CCL, which processes EEG data without handcrafted feature engineering. DSAN-CCL is an integration of deep learning and transfer learning, which benefits from excellent feature representation due to deep neural network and high capability of knowledge transfer from the source domain to the target domain due to transfer learning (in our case, transfer across subjects for MW classification).

2) We utilized predicted probabilities for each category to construct class confusion matrix and minimized it to improve the credibility of pseudo-labels, thereby enhancing the effectiveness of transfer learning.

3) The proposed method is generic and does not depend on a specific feature extractor. In other words, the feature extractor can be replaced to fit new classification scenarios.

The rest of the paper is organized into the following sections. Section II covers the theoretical background, including the definition of the cross-subject MW recognition problem and the discrepancy metric. Section III details the proposed model DSAN-CCL. Section IV presents comparison results, model ablation, and discussions. Finally, a summary and future work are given in Section V.

II. Theoretical Background

A. Problem Definition

To articulate our study clearly, relevant concepts associated with DA are initially presented. In our study, we utilized the leave-one-subject-out cross-validation approach to assess the cross-subject MW classification for excluding potential biases due to the subject selection for the testing data. In each fold, the data from a subject are testing data, which are without category labels, represented as $D_t = \{(x_j^t)\}_{j=1}^{N_t}$. This is considered as the target domain. The data from the remaining subjects and their labels are denoted as $D_s = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$, which is considered as the source domain. N_t represents the number of unlabeled samples in the target domain, and N_s represents the number of labelled samples in the source domain. We assume that the source and target domains share an identical label space and feature space, where $y^s \in \{1,2,\cdots,K\}$ and $x^s, x^t \in R^{C\times T}$. Here, C represents the number of EEG channels per sample, T stands for the number of data points per channel, and K indicates the number of MW categories. However, there are distinct marginal distributions $P_s(x^s) \neq P_t(x^t)$ and conditional distributions $P_s(y^s|x^s) \neq P_t(y^t|x^t)$. The objective is to mitigate inter-domain bias by transferring source domain knowledge to

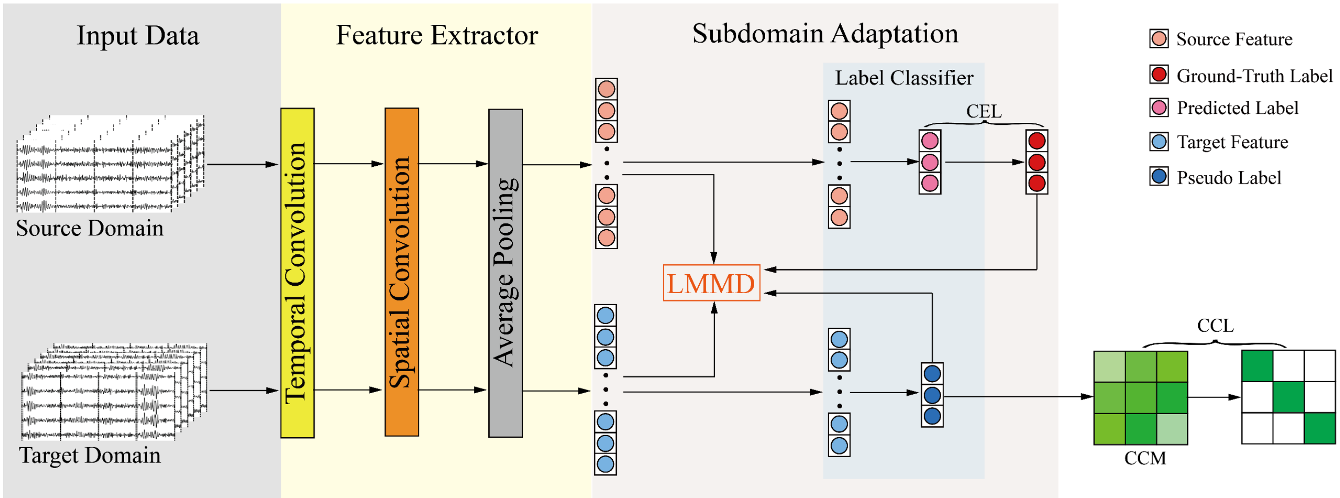

Fig. 2. Model architecture of the proposed DSAN-CCL. The LMMD aligns the feature distributions between the source domain and the target domain for each category. The process of minimizing CCL aims to enhance the credibility of the pseudo-labels.

enhance the generalization capabilities of the model on the target domain.

B. Discrepancy Metric

DA tries to align the distributions between domains. MMD [27] is often used to quantify the difference between the source and target domain distributions p and q, measured by calculating the L2 of the expectation difference between the two distributions in a high-dimensional space. It is defined as follows:

$$\mathcal{L}_{\mathrm{MMD}}(p,q) \triangleq \left\|E_p[\phi(x^s)] - E_q[\phi(x^t)]\right\|_H^2 \tag{1}$$

The function $\phi(\cdot)$ is a high-dimensional mapping function that maps the extracted features into the reproducing kernel Hilbert space H [28], utilizing the kernel function h to measure the similarity between the two domains in the feature space, where $h(x^s, x^t) = \langle\phi(x^s), \phi(x^t)\rangle$, $\langle\cdot,\cdot\rangle$ denotes inner product of vectors. the MMD metric is redefined as follows:

$$\begin{aligned}\mathcal{L}_{\mathrm{MMD}}(p,q) &= \left\|\frac{1}{N_s}\sum_{i=1}^{N_s}\phi(x_i) - \frac{1}{N_t}\sum_{j=1}^{N_t}\phi(x_j)\right\|_H^2 \\ &= \frac{1}{N_s^2}\sum_{i=1}^{N_s}\sum_{j=1}^{N_s} h(x_i^s, x_j^s) + \frac{1}{N_t^2}\sum_{i=1}^{N_t}\sum_{j=1}^{N_t} h(x_i^t, x_j^t) \\ &\quad - \frac{2}{N_s N_t}\sum_{i=1}^{N_s}\sum_{j=1}^{N_t} h(x_i^s, x_j^t)\end{aligned} \tag{2}$$

III. Proposed Model

We employed a domain-sharing convolutional network as a feature extractor in the proposed model (see the model architecture in Fig. 2). The feature extractor consists of both temporal convolution and spatial convolution, as well as average pooling. It extracts high-level features from the preprocessed EEG data. Subsequently, the extracted features are fed into a SA module for two objectives. The primary objective is to predict the sample labels of either the source domain or target domain based on the input features. The other objective is to quantify the differences in feature distributions of the source and target domains for each category (subdomain) using LMMD and to minimize the inter-domain variances. Finally, the pseudo-labels generated by the label classifier are fed into the CCL module to enhance the credibility of the pseudo-labels for the improvement of transfer effectiveness.

A. Domain-Shared Feature Extractor

Feature extraction module is utilized for the extraction of EEG features. It integrates temporal and spatial convolutions to effectively capture spatiotemporal information in EEG signals. The initial layer of the module consists of a one-dimensional convolutional layer with n kernels in the size of $(1, 13)$ and the stride of $(1, 1)$. This layer aids in capturing temporal-dependent characteristics along the temporal dimension through convolutional operation. The following one-dimensional convolutional layer functions in the spatial dimension using n kernels in the size of $(C, 1)$ and the stride of $(1, 1)$ to capture the interaction information among the various electrode channels. To expedite the training process and improve the generalization of the model, a batch normalization layer was incorporated, along with Gaussian error linear units (GELU) as the activation function. Subsequently, an average pooling layer is employed to enhance feature smoothing and decrease computational complexity. The layer is characterized by a kernel size of $(1, 35)$ and a stride of $(1, 7)$. Moreover, the inclusion of a dropout layer ($d = 0.5$) serves to mitigate the risk of overfitting. Finally, the high-dimensional features obtained are flattened, and the extracted features are then projected into a 1024-dimensional feature space using a linear mapping layer. In summary, the input data (x^s, x^t) are mapped to the feature f through the feature extractor G_f with parameter θ_f, denoted as $[f^s = G_f(x^s; \theta_f)$, $f^t = G_f(x^t; \theta_f)$]. This module is intended to provide a more informative feature representation for subsequent processes. The parameter settings are summarized in Table I.

TABLE I
PARAMETER SETTINGS

	Layer	Parameter	Output
	Input	Shape: (1, C^{a}, T^{b})	/
Temporal Conv	Conv_1	Kernel: (1, 13) Stride: 1 Padding: 0	(n, C, T_1^{c})
Spatial Conv	Conv_2	Kernel: (n, 1) Stride: 1 Padding: 0 BatchNorm2d GELU	(n, 1, T_1)
	Average Pooling	Kernel: (1, 35) Stride (1, 7) Padding: 0 Dropout: 0.5	(n, 1, T_2^{d})
Projection	Dense_1	Flatten Units: 1024	(1024, 1)
Label Classifier	Dense_2	Units: 512 GELU Dropout: 0.5	(512, 1)
	Dense_3	Units: 3	(3, 1)

a: C represents the number of EEG channels per sample. b: T stands for the number of data points per channel. c: $T_1 = \frac{T-13}{1} + 1$. d: $T_2 = \frac{T_1-35}{7} + 1$.

B. Label Classifier

We used two fully connected layers (see the parameter settings in Table I) as the label classifier in the proposed model. The first layer utilizes the GELU activation function, whereas the output of the following layer is subjected to the SoftMax function to obtain the predicted label $\hat{y}$. The process can be summarized as follows: the intermediate features f, generated by the feature extractor, are mapped to the predicted label $\hat{y}$ through the label classifier G_c utilizing the parameter θ_c, formulated as $\hat{y} = G_c(f;\theta_c)$. The training process entails the utilization of labelled samples from the source domain and applying a cross-entropy loss function to ensure the prediction precision of the label classifier. The loss function of the label classifier can be formally defined as:

$$\mathcal{L}_{\text{CEL}} = -\frac{1}{N_s}\left[\sum_{i=1}^{N_s}\sum_{k=1}^{K} I[\mathrm{y}_i^s = k]\log(G_c(f;\theta_c) = k)\right] \quad (3)$$

If $\mathrm{y}_i^s = k$, then the function $I[\mathrm{y}_i^s = k]$ takes the value of 1; otherwise, it takes the value of 0.

C. LMMD-Based Subdomain Adaptation

In cross-subject MW classification, current DA techniques treat all data as a whole, thus overlooking the nuanced distribution of data within individual categories. To tackle this issue, we employ the LMMD [29], [30] methodology, which considers each category data as a subdomain. LMMD offers more precise subdomain alignment by evaluating the kernel-mean embedding correlation of the subdomains in Hilbert space. This approach enables the feature extractor to acquire domain-independent features and accomplish the alignment of subdomain feature distributions. The formula for LMMD is presented as:

$$\mathcal{L}_{\text{LMMD}}(p,q) \triangleq E_k \left\| E_{p^{(k)}}[\phi(f^s)] - E_{q^{(k)}}[\phi(f^t)] \right\|_H^2 \quad (4)$$

This LMMD can be considered a weighted MMD, where weights $\{w^{sk}, w^{tk}\}$ are assigned to the samples from each category. The weights $\{w^{sk}, w^{tk}\}$ can be mathematically represented as:

$$w_i^{sk} = \frac{\mathrm{y}_i^{sk}}{\sum_{(x_j,y_j)\in D_s}\mathrm{y}_j^{sk}} \quad or \quad w_i^{tk} = \frac{\mathrm{y}_i^{tk}}{\sum_{(x_j,y_j)\in D_t}\mathrm{y}_j^{tk}} \quad (5)$$

For the samples in the source domain, weights w^{sk} can be determined according to their actual labels. For the samples in the target domain, we used the label classifier to generate pseudo-labels and then weights w^{tk} are computed based on these pseudo-labels. The optimization of (4) can be achieved as follows:

$$\mathcal{L}_{\text{LMMD}}(p,q) = \frac{1}{K}\sum_{k=1}^{K}\left\| \sum_{x_i^s\in D^s} w_i^{sk}\phi(f_i^s) - \sum_{x_j^t\in D^t} w_j^{tk}\phi(f_j^t) \right\|_H^2 \quad (6)$$

The inputs of LMMD include the features f^s, f^t from the source and target domains, the actual labels y^s from the source domain, and the pseudo-labels $\hat{\mathrm{y}}^t$ generated by the label classifier for the samples in the target domain. Inputting this information into (6) yields the complete LMMD loss as follows:

$$\begin{aligned}\mathcal{L}_{\text{LMMD}}(p,q) = \frac{1}{K}\sum_{k=1}^{K}\Bigg[&\sum_{i=1}^{N_s}\sum_{j=1}^{N_s} w_i^{sk}w_j^{sk}h(f_i^s,f_j^s) \\ &+\sum_{i=1}^{N_t}\sum_{j=1}^{N_t} w_i^{tk}w_j^{tk}h(f_i^t,f_j^t) \\ &-2\sum_{i=1}^{N_s}\sum_{j=1}^{N_t} w_i^{sk}w_j^{tk}h(f_i^s,f_j^t)\Bigg]\end{aligned} \quad (7)$$

D. Class Confusion Loss

In the computation of the LMMD loss, as indicated in (7), the pseudo-labels generated by label classifiers are incorporated. If these pseudo-labels are not reliable, it leads to a decline in the model's performance. The quality of the pseudo-labels plays a critical role in determining the final model's performance. Hence, we proposed to utilize CCL [31] to enhance the credibility of the pseudo-labels and improve the overall performance of the transformation process. The primary concept posits that the confusion arising from distinct classes can be inherently manifested in the product of the class probabilities predicted by the label classifier and its transpose. As depicted in Fig. 3, within this procedure, the information entropy of samples is initially computed based on the class probabilities generated by the label classifier, represented by the following:

$$E(\hat{y}_r) = -\sum_{k=1}^{K}\hat{y}_{rk}\log\hat{y}_{rk} \quad (8)$$

$\hat{y}_{rk}$ denotes the probability that the r-th sample belongs to the category k. According to entropy theory, a lower entropy value indicates a higher certainty in category prediction. To achieve this, the sign of the entropy will be inverted and fed into the

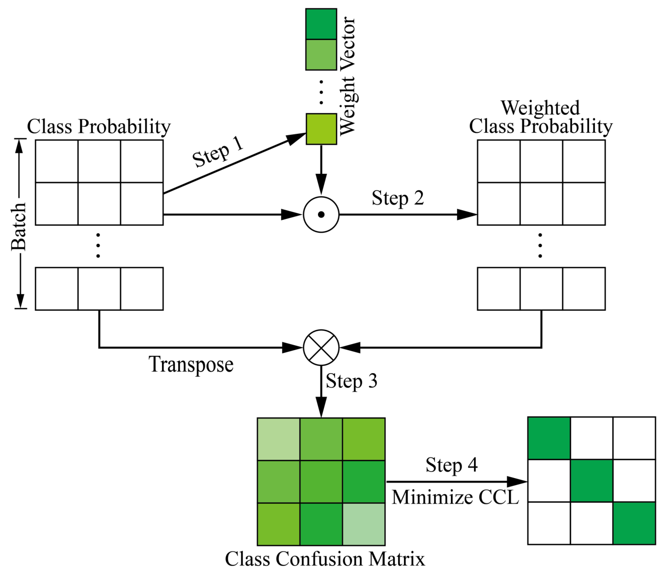

Fig. 3. Illustration of the computation of CCL. In Step 1, the entropy weights of the samples are calculated. In Step 2, the weighted class probabilities are obtained by multiplying the class probability by the weight. The CCM is generated by multiplying the weighted class probabilities by the transpose of the class probabilities in Step 3. Finally, the CCL is minimized in Step 4 by minimizing the non-diagonal elements of the CCM while maximizing the diagonal elements.

exponential function. Ultimately, the weight of each sample can be expressed as:

$$a_r = \frac{B \cdot (1 + \exp(-E(\hat{y}_r)))}{\sum_{j=1}^{B} (1 + \exp(-E(\hat{y}_j)))} \tag{9}$$

a is a weight vector. B denotes the batch size, where the sample weights are multiplied by a factor of B to ensure that the total weight of samples in a batch equals B, an average weight of 1 for each sample. The predicted class probabilities are multiplied by their corresponding weights to calculate weighted class probabilities. These weighted class probabilities are then multiplied by the transpose of the class probabilities to obtain the class confusion matrix (CCM), denoted as:

$$CCM = \hat{Y}^T(\hat{Y} \odot a) \tag{10}$$

$\hat{Y}$ is a class probability matrix. $\odot$ denotes the element-wise multiplication of the vector a with each column of $\hat{Y}$. CCM is further normalized as follows:

$$CCM'_{uv} = \frac{CCM_{uv}}{\sum_{o=1}^{K} CCM_{uo}} \tag{11}$$

The CCM' illustrates the degree of confusion for each class in the target domain. The values on the diagonal represent the probabilities of accurate classification. We aim to maximize the value of the elements located on the diagonal while minimizing the value of the remaining elements. Therefore, the CCL can be formulated as:

$$\mathcal{L}_{\mathrm{CCL}} = \frac{1}{K}(\sum_{u=1}^{K}\sum_{v=1}^{K} CCM'_{uv} - \sum_{u=1}^{K} CCM'_{uu}) \tag{12}$$

The cost function of the proposed DSAN-CCL model comprises three components: label classifier loss, LMMD of the SA and CCL. The final loss function of the proposed model is

Algorithm 1 DSAN-CCL Model Training

Require: the source and target samples (x^s, x^t); learning rate $l = 0.0001$; batchsize $B = 128$; max epoch $e = 100$.

For $(epoch;\ epoch \le e;\ epoch \leftarrow epoch + 1)$ **do**
  While $batch \le\ batches$:
    A batch of source data $\{(x_i^s, y_i^s)\}_{i=1}^{B}$;
    A batch of target examples $\{x_j^t\}_{j=1}^{B}$;
    Capturing feature f^t and f^s;
    Computing $\mathcal{L}_{\mathrm{CEL}}$ by Eq. (3);
    Obtain the pseudo-labels $\{\hat{y}_j^t\}_{j=1}^{B}$;
    Computing $\mathcal{L}_{\mathrm{CCL}}$ by Eq. (12);
    Computing $\mathcal{L}_{\mathrm{LMMD}}$ by Eq. (7);
    Computing $\mathcal{L}_{\mathrm{total}}(\theta_f, \theta_c)$ by Eq. (13);
    Update $\theta_f,\ \theta_c$;
    $batch {+}{=} 1$;
  End
End

the sum of the losses of the above three components (see Eqs. (3), (7), and (12)) and is defined as:

$$\mathcal{L}_{\mathrm{total}}(\theta_f, \theta_c) = \mathcal{L}_{\mathrm{CEL}} + \lambda_l \mathcal{L}_{\mathrm{LMMD}} + \lambda_c \mathcal{L}_{\mathrm{CCL}} \tag{13}$$

Where λ_l and λ_c are denoted as the weights of $\mathcal{L}_{\mathrm{LMMD}}$ and $\mathcal{L}_{\mathrm{CCL}}$, respectively. The algorithm for the model training is presented in Algorithm I.

IV. Results and Discussions

A. Datasets and Methods Used in Comparisons

To validate the superiority of DSAN-CCL in the cross-subject MW classification, two datasets were employed for performance evaluation: a private dataset known as the simulated flight experiment (SFE) [32] and a publicly available dataset named the multi-attribute task battery (MATB) [33] (see the information in Table II). For the MATB dataset, we performed cross-subject classification separately for each session (Session 1 and Session 2). Detailed descriptions of the dataset and preprocessing steps can be found in [14], [32], [33]. The experiment protocol for the private dataset was reviewed and approved by the Institutional Review Board of the National University of Singapore. A consent form was given by each subject before the start of the experiment.

We selected six state-of-the-art DA methods. To ensure the fairness of the comparison, all methods used the same feature extractor (i.e., S-Net). The following is a brief summary of the comparative methods.

TABLE II
Information About Datasets

	SFE	MATB
Number of Subjects	7	15
Number of Classes	3	3
Sampling Rate	256 Hz	250 Hz
Number of Channels	62	61
Number of Samples	540	447

1) Shallow-Net (S-Net) [34]: S-Net trained using only source domain data, serving as a benchmark in our comparisons.

2) Custom domain adaptation (CDA) [24]: CDA integrates MMD with a deep learning network to reduce the differences between the source and target domains.

3) Deep domain adaptation (DDA) [25]: DDA integrates adversarial learning with MMD to tackle the cross-subject MW issue.

4) Correlation alignment (CORAL) [35]: CORAL employs correlation alignment to address distributional variations across multiple domains.

5) Domain adversarial neural network (DANN) [36]: DANN employs adversarial learning to facilitate the adversarial training of feature extractors and domain classifiers, thereby extracting domain-invariant features.

6) Conditional adversarial domain adaptation (CDAN) [37]: CDAN enhances DANN by incorporating multilinear conditional conditioning and entropy conditional conditioning to improve classification performance.

7) Siamese deep domain adaptation (SDDA) [20]: SDDA utilizes MMD and cosine-based center loss to minimize the disparity between embedded features in the source and target domains.

B. Comparison Results

We aim to transfer between subjects (cross-subject MW classification). In this case, we adopted leave-one-subject-out cross-validation to evaluate the models. In each round, one subject is designated as the target domain, while the rest of the subjects are designated as the source domain. In the below results, an accuracy for subject x was obtained when this subject x was designated as the target domain and the rest of the subjects were designated as the source domain. The average accuracy was calculated by averaging across all subjects.

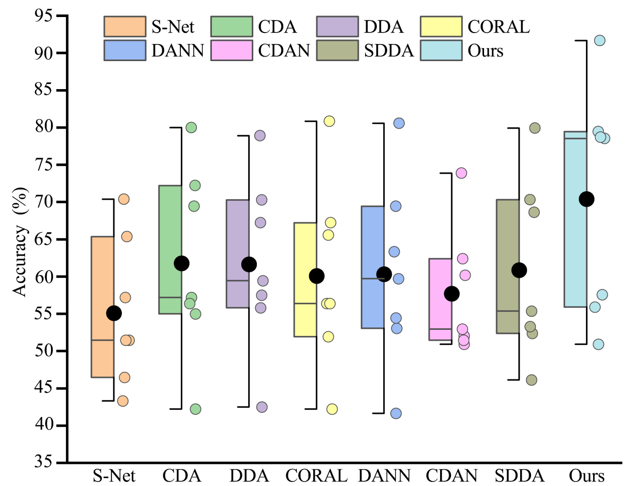

Fig. 4. Comparison results for the dataset SFE. The colored dots indicate the accuracies of each subject. The black dots present average accuracies of each method.

1) Results on the SFE: Fig. 4 shows the classification results of the proposed method and the other compared methods evaluated using the SFE dataset. S-Net lacks DA and is not good for the cross-subject classification. It is obvious that the other methods with DA surpass S-Net in the classification performance. This indicates that DA has a positive effect on the cross-subject MW classification. Compared to other DA methods, the proposed model is considerably better, with an average accuracy of 70.4%, which is elevated by 8~10 percentage points. The performance enhancement is due to the fact that DA is conducted separately in each subdomain in our

TABLE III

COMPARISON RESULTS FOR THE DATASET MATB_SESSION 1 (CLASSIFICATION ACCURACY IN %)

Method \ Subject	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Mean	STD
S-Net	69.35	58.39	44.52	61.30	60.63	62.42	57.97	53.24	62.19	43.18	43.85	69.80	41.39	46.98	55.26	55.36	9.12
CDA	**73.38**	65.32	58.84	65.55	60.85	60.29	62.27	61.74	63.59	**48.55**	49.68	75.39	42.95	47.65	59.17	59.68	8.86
DDA	72.81	61.52	65.10	**66.37**	59.73	61.07	63.98	57.72	65.62	42.95	47.45	74.05	48.99	45.86	55.68	59.26	9.18
CORAL	71.59	64.65	57.94	63.98	60.89	61.52	61.74	60.18	63.53	47.43	48.77	74.27	42.73	47.87	**59.51**	59.11	8.63
DANN	71.67	62.64	60.28	63.09	59.96	63.79	59.73	59.06	68.00	47.76	53.91	72.04	49.20	47.67	54.36	59.54	7.58
CDAN	70.68	62.56	59.51	64.83	60.54	64.65	58.61	59.51	66.22	48.19	**55.89**	70.47	46.58	49.34	53.69	59.42	7.31
SDDA	72.25	66.67	56.38	64.88	**61.07**	60.28	59.79	62.37	61.30	46.98	46.76	74.50	44.52	48.52	58.27	58.97	8.77
Ours	71.36	**74.94**	**77.40**	65.77	60.18	**72.93**	**77.85**	**72.26**	**69.57**	46.53	49.66	**78.97**	**62.19**	**50.56**	57.05	**65.82**	10.52

TABLE IV

COMPARISON RESULTS FOR THE DATASET MATB_SESSION 2 (CLASSIFICATION ACCURACY IN %)

Method \ Subject	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Mean	STD
S-Net	56.38	55.26	59.28	58.84	49.89	50.34	60.40	50.56	59.96	48.32	50.11	57.94	52.57	45.86	62.42	54.54	5.00
CDA	56.82	55.93	58.36	62.67	50.78	63.31	62.48	60.70	64.88	64.65	57.42	58.61	51.60	**52.13**	66.00	59.09	4.82
DDA	57.82	56.78	55.48	63.09	53.47	57.40	62.54	57.46	63.76	62.68	56.29	**59.90**	51.23	48.77	65.10	58.12	4.61
CORAL	57.27	57.91	56.24	61.74	52.69	59.15	61.67	54.25	62.46	57.49	57.24	58.21	51.90	49.34	65.55	57.54	4.18
DANN	55.60	55.27	57.40	60.40	53.24	57.67	59.06	54.36	61.07	54.59	57.66	57.86	48.32	48.55	64.47	56.37	4.18
CDAN	55.93	55.70	58.83	63.82	52.57	61.47	59.89	56.46	61.48	55.03	58.59	59.34	50.64	47.34	61.56	57.24	4.37
SDDA	57.49	59.66	57.27	61.52	51.45	59.67	60.85	53.91	62.47	56.38	56.15	57.72	**52.80**	47.67	**66.67**	57.45	4.60
Ours	**60.18**	**62.19**	**91.50**	**73.83**	**62.86**	**77.85**	**68.01**	**69.57**	**66.67**	**80.54**	**59.51**	58.39	49.22	49.66	61.97	**66.13**	10.96

proposed model to better adapt to local properties and variations.

2) Results on the MATB: Table III and Table IV illustrate the MW classification results on session 1 and session 2 of the MATB dataset. Unsurprisingly, all DA methods achieve higher results than S-Net. This result reiterates the advantage of DA in the cross-subject MW classification. The proposed method DSAN-CCL obtained the highest average accuracies of 65.82% and 66.13% for session 1 and session 2, respectively. The average accuracy in session 1 is 10.5 percentage points higher than that of S-Net and 6~7 percentage points higher than that of the compared methods. For session 2, it is 12 percentage points higher than S-Net and 7~9 percentage points higher than others. These results further demonstrate the superiority of DSAN-CCL.

3) Statistical Analysis: Wilcoxon paired signed rank test [38] was used to assess whether the performance difference between the proposed DSAN-CCL model and the compared models is statistically significant. The statistical analysis results show that DSAN-CCL is significantly better than S-Net without DA on the SFE dataset ($p<0.01$). DSAN-CCL is also better than the other compared models at the significance level of 0.05. When compared to all other methods on both sessions of the MATB dataset, DSAN-CCL achieved significantly better performance ($p<0.01$).

C. Exploring Methodological Versatility

To comprehensively investigate the generalizability of our proposed model and its association with feature extractors, we used two typical networks in the EEG signal classification, namely EEGNET [39] and Conformer [40], to substitute the feature extractor (S-Net) in our method. According to the results shown in Fig. 5, it is evident that the performance was considerably enhanced when SA was adopted compared to the case without (W/O) SA. These results suggest that our proposed approach is independent of particular feature extractors and can be seamlessly transferred to various feature extraction methods. This is useful for adapting to different tasks in which features need to be extracted by a different feature extractor, making our approach universal and scalable.

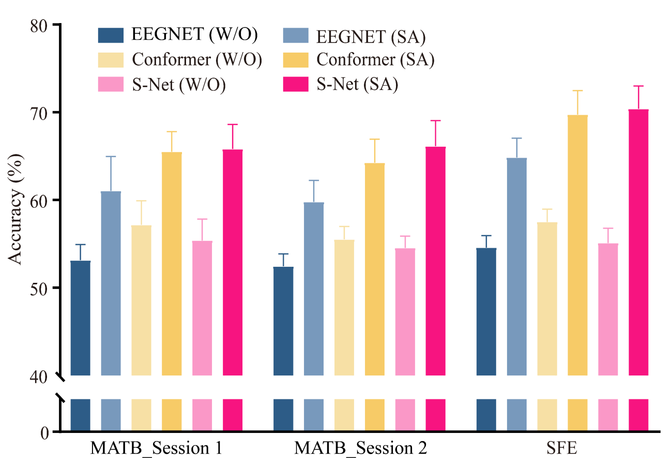

Fig. 5. Classification performance when different feature extractors are adopted, with or without SA.

D. Model Analysis

In order to comprehensively understand the proposed DSAN-CCL, we conducted the following explorations using the public dataset MATB. First, an ablation study was performed to assess the influence of weighted losses on the model's performance. Second, a comparison among different pseudo-label losses was performed to indicate the effectiveness of the CCL. Third, the intermediate features extracted by the feature extractor were visualized to show their distributions and centers, showing the differences between the source domain and the target domain. Finally, quantitative distances were used to measure the DA performance for each method. The method comparison was given.

1) Ablation Study: An ablation study was performed on both sessions of the MATB dataset to find out the best parameters (i.e., weights λ_l and λ_c) of LMMD loss and CCL in the DSAN-CCL model. When a weight of 0 is assigned to a loss, it means that this loss is completely removed and does not contribute to the training of the DSAN-CCL model. We attempted different combinations of weights for these two losses and presented the results in Fig. 6. For example, [1, 0] stands for only adopting the LMMD loss. The cases [0, 0.5] and [0, 1] are excluded because the presence of a CCL presupposes the existence of a LMMD loss. The results show that the classification accuracy is lowest when both losses are removed from the model (see the case [0, 0] shown in Fig. 6). The accuracy is gradually increased with the increase of the weight of the LMMD loss (λ_l). This trend is detrimentally influenced when the weight of the CCL (λ_c) exceeds that of the LMMD loss. As the weight λ_l is increased, the model elevates the emphasis on the distribution alignment between the source domain and the target domain, resulting in a reduction of cross-domain discrepancy. This positively improves the model's generalization to the target domain. The increase of the weight λ_c strengthens the effect to maximize one of the predicted class probabilities while minimizing the rest of them, which enhances the certainty of the model prediction.

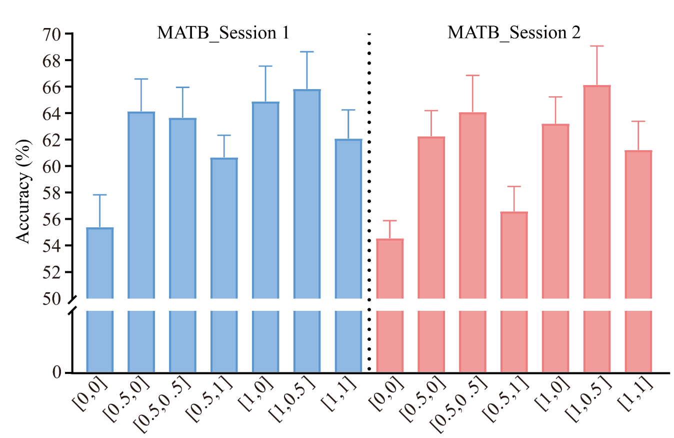

Fig. 6. The performance of the proposed method was evaluated with different weights assigned to the LMMD loss λ_l and CCL λ_c.

TABLE V

PERFORMANCE COMPARISON AMONG PSEUDO-LABEL LOSSES (IN %)

	Session 1			Session 2		
	CCL	PEL	CML	CCL	PEL	CML
1	71.36	66.67	70.47	60.18	61.74	60.63
2	74.94	68.01	69.80	62.19	62.86	61.74
3	77.40	75.84	78.08	91.50	78.30	86.35
4	65.77	57.27	58.17	73.83	70.69	72.48
5	60.18	61.07	61.52	62.86	58.84	59.28
6	72.93	68.68	71.59	77.85	62.42	69.35
7	77.85	61.97	67.79	68.01	63.31	63.76
8	72.26	74.94	76.06	69.57	68.46	68.23
9	69.57	74.72	75.62	66.67	65.77	66.00
10	46.53	51.01	51.23	80.54	79.19	78.75
11	49.66	51.90	54.36	59.51	57.05	56.82
12	78.97	73.15	76.96	58.39	61.52	61.30
13	62.19	61.52	61.07	49.22	49.44	51.01
14	50.56	47.87	47.87	49.66	52.35	52.35
15	57.05	55.48	53.91	61.97	63.98	64.65
Mean	65.82	63.34	64.97	66.13	63.73	64.85
STD	10.52	8.98	9.82	10.96	7.92	9.06

However, a large value set for λ_c can reduce the model performance. This is because CCL forces one class to be dominant in the prediction, but it cannot guarantee that the prediction is correct. If the prediction is not correct, it poses a negative effect on the model training. Therefore, the model performance is decreased. According to the results, the model achieves the best performance for both sessions when $[\lambda_l, \lambda_c] = [1, 0.5]$.

2) Pseudo-Label Loss: In our proposed method, CCL is used to evaluate the quality of the generated pseudo-labels. In order to show the superiority of CCL, two methods, pseudo-label entropy loss (PEL) [41] and class-margin loss (CML) [42], are included in the comparison. According to the results of the ablation study, a weight of 0.5 is optimal. Therefore, the weight is kept at 0.5 in the comparisons. Table V shows the comparison results based on the MATB dataset. CCL is the best method among the methods in the comparison, which is around 1~2 percentage points higher than the other methods. The higher performance achieved with CCL might be due to the fact that CCL also takes between-class penalties into consideration. For the methods of PEL and CML, only within-class penalties are included.

3) Feature Visualization: DA facilitates the feature extractor to capture distinctive domain-invariant features. We visualized the features to inspect the effectiveness of DA methods. The extracted features are highly dimensional and cannot be visualized directly. Therefore, we employed t-distributed stochastic neighborhood embedding (t-SNE) [43] to reduce the dimensionality of the features. We took Subject 8 from session 1 of the dataset MATB as an example for illustration. As depicted in Fig. 7, the centers of each subdomain are indicated by symbols, and the extent of each subdomain is visualized in colors. When looking at the distributions of the source domain only, we found that there is no overlap between the categories for all methods. It means that all methods are able to distinguish each category perfectly in the source domain, showing a good outcome of the model learning with available labels in the source domain. When transferring to the target domain, all other methods have a relatively large overlap between the categories in the target domain except the proposed DSAN-CCL method. The proposed method has distinct boundaries among categories except for a certain overlap between the medium MW category and the high MW category. When taking both source and target domains into consideration, the proposed method exhibits the closer centers between the source domain and the target domain for each category. It also exhibits a better overlap between the domains for the same category. In contrast, other methods showed much less overlap and farther centers. This result demonstrates that the proposed method is able to better align features between the source domain and the target domain and is advantageous for the cross-subject MW classification.

4) Distribution Discrepancy: To further illustrate the validity of the proposed methodology. We utilized A-distance [44], [45] to quantify the difference in the feature distributions between the source domain and the target domain. The A-distance is defined as $d_A = 2(1 - 2\epsilon)$. Here, ϵ represents the generalization error of a simple classifier (e.g., SVM with linear

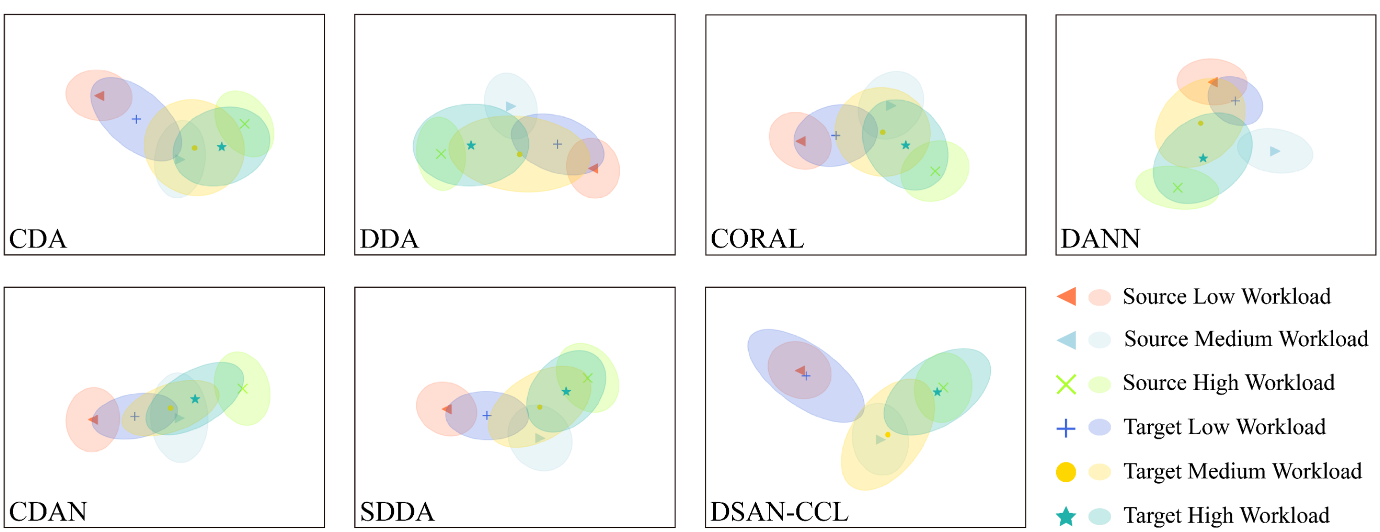

Fig. 7. Visualization of the features extracted by the feature extractor with different DA methods. The symbols indicate the centers of each subdomain. The color-coded areas delineate the extent of each subdomain.

TABLE VI
A-Distance of the Feature Distributions between the Source Domain and the Target Domain for Each Method from the Global Perspective

	Session 1	Session 2
	d_A	d_A
CDA	1.890	1.883
DDA	1.803	1.800
CORAL	1.873	1.889
DANN	1.836	1.865
CDAN	1.837	1.866
SDDA	1.891	1.900
Ours	**1.746**	**1.759**

TABLE VII
A-Distance of the Feature Distributions between the Source Subdomain and the Target Subdomain for Each Method

	Session 1			Session 2		
	d_A			d_A		
	Low	Medium	High	Low	Medium	High
CDA	1.872	1.894	1.879	1.893	1.904	1.899
DDA	1.860	1.830	1.835	1.842	1.876	1.868
CORAL	1.869	1.883	1.862	1.886	1.901	1.895
DANN	1.859	1.853	1.851	1.861	1.888	1.875
CDAN	1.860	1.850	1.856	1.855	1.896	1.878
SDDA	1.873	1.885	1.875	1.885	1.904	1.918
Ours	**1.759**	**1.814**	**1.813**	**1.774**	**1.820**	**1.818**

kernel) in the binary classification problem of distinguishing between source and target domains. We applied A-distance to the feature distributions between the source domain and the target domain from a global perspective, and between the source subdomain and the target subdomain from a subdomain perspective for each method, respectively. The features from the samples of both source and target domains are used to train the classifier to obtain the generalization error. For this quantitative metric, a smaller A-distance indicates a better alignment between the source domain and the target domain. The results are shown in Table VI and Table VII. DSAN-CCL obtains the smallest A-distance in both sessions, which is consistent with the observations of feature visualization in Fig. 7. Once again, it demonstrates that the proposed model not only makes closer between the source domain and the target domain from the global perspective but also aligns them better for each subdomain.

E. Limitations

In this study, we treated the samples from all subjects as a sample pool for the source domain during the training. Each training sample was utilized equally for the training without considering the sample variation across the samples from different subjects. If the sample variation is considered during the training, the performance could be further improved. In addition, the number of subjects is not large in both of the datasets used in this study. It would be better to have more subjects used in the performance evaluation. However, to some extent, the evaluation of leave-one-subject-out cross-validation releases this limitation. Our dataset, as well as the publicly available datasets, are of three levels of MW, each of which has an equivalent number of samples. The equivalent sample size might not always be the case. The imbalanced situation has not been tested in this study. Moreover, the MW state might be more complex in the real world. The experiments for collecting both datasets used in this study do not completely simulate the complex real-world situation. This inevitable factor could introduce a confounding contribution to the assessment of the proposed method. We will investigate these aspects in our future work.

V. Conclusion and Future Work

This study proposed a SA-based transfer learning model (i.e., DSAN-CCL) to classify MW levels, which achieved better performance in cross-subject classification. DSAN-CCL 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 MW classification. CCL is used to improve the credibility level of pseudo-labels generated by the label classifier and to further improve the transfer performance of subdomain adaptation. In addition, our study demonstrated that it does not rely on a particular feature extractor and is applicable to any feature extractor. The feature visualization further confirmed the advantage of DSAN-CCL, showing more separable features among categories but more aligned features between the source domain and the target domain.

As mentioned in the section on limitations, we will consider sample variation to improve the performance in the future. Moreover, we will utilize complementary information from different source domains to enhance target DA performance. One of the potential approaches is to employ a feature-weighting strategy for multiple source domains. Alternatively, adversarial training can be used to construct a shared feature space that retains distinct characteristics of each source domain, thus preventing information loss. This might be helpful to accomplish good adaptation in diverse and complex situations.

References

[1] C. D. Wickens, “Multiple resources and performance prediction,” Theor Issues Ergon Sci, vol. 3, no. 2, pp. 159–177, 2002.

[2] Y. Zhou, S. Huang, Z. Xu, P. Wang, X. Wu, and D. Zhang, “Cognitive workload recognition using EEG signals and machine learning: A review,” IEEE Trans Cogn Dev Syst, vol. 14, no. 3, pp. 799–818, 2021.

[3] P. Zhang, X. Wang, J. Chen, W. You, and W. Zhang, “Spectral and temporal feature learning with two-stream neural networks for mental workload assessment,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 27, no. 6, pp. 1149–1159, 2019.

[4] J. Li, “Thoughts on neurophysiological signal analysis and classification,” Brain Science Advances, vol. 6, no. 3, pp. 210–223, Sep. 2020, doi: 10.26599/bsa.2020.9050020.

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

[6] D. Tao, H. Tan, H. Wang, X. Zhang, X. Qu, and T. Zhang, “A systematic review of physiological measures of mental workload,” Int J Environ Res Public Health, vol. 16, no. 15, p. 2716, 2019.

[7] W. L. Lim, O. Sourina, Y. Liu, and L. Wang, “EEG-based mental workload recognition related to multitasking,” in 2015 10th International Conference on Information, Communications and Signal Processing (ICICS), 2015, pp. 1–4.

[8] H. Qu, X. Gao, and L. Pang, “Classification of mental workload based on multiple features of ECG signals,” Inform Med Unlocked, vol. 24, p. 100575, 2021.

[9] T. Shimizu et al., “Measurement of frontal cortex brain activity attributable to the driving workload and increased attention,” SAE International Journal of Passenger Cars-Mechanical Systems, vol. 2, no. 2009-01–0545, pp. 736–744, 2009.

[10] U. Asgher, K. Khalil, Y. Ayaz, R. Ahmad, and M. J. Khan, “Classification of mental workload (MWL) using support vector machines (SVM) and convolutional neural networks (CNN),” in 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), 2020, pp. 1–6.

[11] D. Das Chakladar, S. Dey, P. P. Roy, and D. P. Dogra, “EEG-based mental workload estimation using deep BLSTM-LSTM network and evolutionary algorithm,” Biomed Signal Process Control, vol. 60, p. 101989, 2020.

[12] S. Yang, Z. Yin, Y. Wang, W. Zhang, Y. Wang, and J. Zhang, “Assessing cognitive mental workload via EEG signals and an ensemble deep learning classifier based on denoising autoencoders,” Comput Biol Med, vol. 109, pp. 159–170, 2019.

[13] L.-W. Ko, R. K. Chikara, Y.-C. Lee, and W.-C. Lin, “Exploration of user’s mental state changes during performing brain–computer interface,” Sensors, vol. 20, no. 11, p. 3169, 2020.

[14] Z. Pei, H. Wang, A. Bezerianos, and J. Li, “EEG-based multiclass workload identification using feature fusion and selection,” IEEE Trans Instrum Meas, vol. 70, pp. 1–8, 2020.

[15] I. Guler and E. D. Ubeyli, “Multiclass support vector machines for EEG-signals classification,” IEEE transactions on information technology in biomedicine, vol. 11, no. 2, pp. 117–126, 2007.

[16] Q. She, T. Chen, F. Fang, J. Zhang, Y. Gao, and Y. Zhang, “Improved domain adaptation network based on wasserstein distance for motor imagery EEG classification,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 31, pp. 1137–1148, 2023.

[17] C. Gao, H. Uchitomi, and Y. Miyake, “Cross-Sensory EEG Emotion Recognition with Filter Bank Riemannian Feature and Adversarial Domain Adaptation,” Brain Sci, vol. 13, no. 9, p. 1326, 2023.

[18] Z. He, Y. Zhong, and J. Pan, “An adversarial discriminative temporal convolutional network for EEG-based cross-domain emotion recognition,” Comput Biol Med, vol. 141, p. 105048, 2022.

[19] W. Liu, C. Guo, and C. Gao, “A cross-session motor imagery classification method based on Riemannian geometry and deep domain adaptation,” Expert Syst Appl, vol. 237, p. 121612, 2024.

[20] Z. Miao, X. Zhang, C. Menon, Y. Zheng, M. Zhao, and D. Ming, “Priming cross-session motor imagery classification with a universal deep domain adaptation framework,” arXiv preprint arXiv:2202.09559, 2022.

[21] X. Cui, T. Wang, X. Lai, T. Jiang, F. Gao, and J. Cao, “Cross-Subject Seizure Detection by Joint-Probability-Discrepancy-Based Domain Adaptation,” IEEE Trans Instrum Meas, vol. 72, pp. 1–13, 2023.

[22] Y. Zhou et al., “Cross-task cognitive workload recognition based on EEG and domain adaptation,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 30, pp. 50–60, 2022.

[23] K. Guan, Z. Zhang, T. Liu, and H. Niu, “Cross-Task Mental Workload Recognition Based on EEG Tensor Representation and Transfer Learning,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2023.

[24] M. Jiménez-Guarneros and P. Gómez-Gil, “Custom Domain Adaptation: A new method for cross-subject, EEG-based cognitive load recognition,” IEEE Signal Process Lett, vol. 27, pp. 750–754, 2020.

[25] Y. Zhou et al., “Cross-subject cognitive workload recognition based on EEG and deep domain adaptation,” IEEE Trans Instrum Meas, 2023.

[26] Z. Yin and J. Zhang, “Cross-session classification of mental workload levels using EEG and an adaptive deep learning model,” Biomed Signal Process Control, vol. 33, pp. 30–47, 2017.

[27] K. M. Borgwardt, A. Gretton, M. J. Rasch, H.-P. Kriegel, B. Schölkopf, and A. J. Smola, “Integrating structured biological data by kernel maximum mean discrepancy,” Bioinformatics, vol. 22, no. 14, pp. e49–e57, 2006.

[28] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, “A kernel two-sample test,” The Journal of Machine Learning Research, vol. 13, no. 1, pp. 723–773, 2012.

[29] Y. Zhu et al., “Deep subdomain adaptation network for image classification,” IEEE Trans Neural Netw Learn Syst, vol. 32, no. 4, pp. 1713–1722, 2020.

[30] Wu Sun and Junhua Li, “Deep Subdomain Adaptation Network Improves Cross-Subject Mental Workload Classification,” International Conference on Cyberworlds, 2024, pp. 138–143.

[31] Y. Jin, X. Wang, M. Long, and J. Wang, “Minimum class confusion for versatile domain adaptation,” in Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXI 16, 2020, pp. 464–480.

[32] Y. Yu, A. Bezerianos, A. Cichocki, and J. Li, “Latent Space Coding Capsule Network for Mental Workload Classification,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2023.

[33] Passive BCI Hackathon—Neuroergonomics, “[Online]. Available: https://www.neuroergonomicsconference.um.ifi.lmu.de/pbci/.

[34] R. T. Schirrmeister et al., “Deep learning with convolutional neural networks for EEG decoding and visualization,” Hum Brain Mapp, vol. 38, no. 11, pp. 5391–5420, 2017.

[35] X.-C. Zhong et al., “A deep domain adaptation framework with correlation alignment for EEG-based motor imagery classification,” Comput Biol Med, vol. 163, p. 107235, 2023.

[36] H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, and M. Marchand, “Domain-adversarial neural networks,” arXiv preprint arXiv:1412.4446, 2014.

[37] M. Long, Z. Cao, J. Wang, and M. I. Jordan, “Conditional adversarial domain adaptation,” Adv Neural Inf Process Syst, vol. 31, 2018.

[38] R. F. Woolson, “Wilcoxon signed-rank test,” Wiley encyclopedia of clinical trials, pp. 1–3, 2007.

[39] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, “EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces,” J Neural Eng, vol. 15, no. 5, p. 56013, 2018.

[40] Y. Song, Q. Zheng, B. Liu, and X. Gao, “EEG conformer: Convolutional transformer for EEG decoding and visualization,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 31, pp. 710–719, 2022.

[41] D.-H. Lee and others, “Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks,” in Workshop on challenges in representation learning, ICML, 2013, p. 896.

[42] L. Zhang, H. Li, J. Cui, W. Li, and X. Wang, “Class subdomain adaptation network for bearing fault diagnosis under variable working conditions,” IEEE Trans Instrum Meas, vol. 72, pp. 1–17, 2022.

[43] L. Van Der Maaten, “Accelerating t-SNE using tree-based algorithms,” The journal of machine learning research, vol. 15, no. 1, pp. 3221–3245, 2014.

[44] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, “A theory of learning from different domains,” Mach Learn, vol. 79, pp. 151–175, 2010.

[45] Y. Mansour, M. Mohri, and A. Rostamizadeh, “Domain adaptation with multiple sources,” Adv Neural Inf Process Syst, vol. 21, 2008.